ANIMAL FAITH AND HUMAN DESTINY

Part One: Animal faith

Chapter One: Introduction

> "Out of meat, how do you get thought?" —Patricia Churchland

As a young child, I naturally believed the world appearing around me was simply the way things are—the way they are supposed to be and always had been. I didn't at first grasp that people had painstakingly constructed the familiar urban environment I grew up in, much less that a squishy mass of stuff inside my skull constructed the appearance of a world in my daily experience. The world simply was. Philosophers call that state of mind naïve realism.

Yet, even a child has clues that lead to questioning. With one eye closed, why does the world look a slightly different color viewed through each eye? How is it that I can see faces and familiar shapes in clouds? What is it like to be a bug feeling its way through the wilderness of the backyard lawn? It occurred to me that I too might be a creature making my way through a mysterious wilderness.

As far as we know, most creatures are born naïve realists and remain so. Humans alone on this planet seem to have developed the self-consciousness that permits them to think about their experience; about the intricate nature of the world; and about the relationship between themselves as perceiving subjects and what they perceive. They alone have a science which grasps that the senses do not offer a transparent window on reality—and religions and philosophies which intuited this long beforehand.

Like other creatures, we see what we need to see, and behave in the ways that help us to survive. On the other hand, we have developed reason and imagination, the ability to think abstractly and into the future. We have ideals, in which true perception, objective knowledge, and moral right seem feasible. We contemplate our beastly nature with some reserve, when not revulsion: we are meat machines who aspire to be godlike. This inner discord shows up in culture and in daily life as a tension between subjectivity and objectivity; between the natural inclination to believe our minds and a hard-earned wariness in regard to them. Perception normally serves us well, but not always. It is, in part, by being able to question it that humans have gained their ascendency on the planet. Science is a quest for reality underlying appearances, which so far has proven adaptive. It reveals that those appearances—which we take for granted and for real—are actually a simulation produced in the brain, which can create alternative simulations that are useful or entertaining.

As a philosophical concept, naïve realism asserts that the world is just as we perceive it to be: things exist as they appear, independent of our minds or perceptions. In this view, experience is treated as an accurate portrait of the world. The other side of the convincingness of perception is its transparency: the world appears naturally as though through a clear window. (Ideally, any interface with the world should be transparent in this sense, so that attention dwells in the world and not on the interface, which would be distracting from the real-world tasks of survival.) Animal faith is the price paid for transparency. But it also forfeits control over the interface, since attention does not naturally dwell there.

Few philosophers today would endorse naïve realism. But it is more than the name of a philosophically implausible doctrine. In truth, we are by nature naïve realists, normally taking for granted that our perception of the world is reality itself, without recognizing the mind's active

role in constructing that perception. This transparency of perceptual processing is our default state. When we see a tree, for instance, the tree is just *there*, not dependent on some convoluted activity in our brain. We don't recognize the mental processing involved in seeing at all. Indeed, we cannot see the internal workings behind seeing, except indirectly through scientific investigation. For, the senses face outward toward the world, not inward toward the brain itself. It is the normal function of the brain to project its simulation of the tree as though it was the tree itself, to believe the appearance and take it at face value.

This book examines the mental force that compels us to believe our senses, thoughts, impulses, and feelings, and which renders doubt appropriate and necessary. The philosopher Santayana called that force *animal faith*. Even to assert the notion invites a complementary concept: scepticism. One makes sense only in relation to the other, and so Santayana appropriately titled his book *Scepticism and Animal Faith*.[1] Faith is a liability when it is blind. As we shall see, scepticism serves as an antidote to that blindness. Yet, animal faith pervades cognition despite scepticism, which itself requires a kind of faith. Together they form a creative epistemic cycle.

Santayana himself defines the term: "Animal faith is the belief in the reality of the external world, the assumption that the world is as we perceive it to be."[2] It is faith in the truth of what we perceive and believe about the world, including about ourselves. In short, we are naturally seduced by appearances, compelled to believe our own minds and experience. This faith is not based on reason or logic, but rather on the inbuilt need of organisms to trust the information they have available, without which they couldn't function and would never have arisen. As the name implies, animal faith is a biological necessity, as basic as hunger and the need to feed on other creatures without guilt. To begin with, the animal engaged in living must instinctively believe in the existence of a knowable external world. It must tacitly assume a real field of action in which to act.[3]

One may protest that objects, space and time must exist anyway, despite our representations and evaluations, as the objective pre-existing venue where our actions play out, and upon which our sensory information depends. The point, however, is that as perceiving subjects we are not—and cannot be—in immediate contact with that world, which Kant called the *noumenon*, the world-in-itself. Our minds intervene. But mustn't our perceptions at least reflect that reality if we are to survive? Well, yes and no. There must, of course, be some relation between that inaccessible objective world and the brain's simulation of it we call our subjective experience. But it is certainly not a relationship of resemblance, let alone a one-to-one correspondence. On the other hand, neither is the relationship random or arbitrary. It may not be a relationship we can easily or precisely know. For we live, so to speak, in the domain of the map, which stands as our only means to know the territory. We cannot stand outside the map to view it side by side for comparison with the territory.

The problem is uniquely a challenge for beings who know themselves to be "conscious." It asks how the physical brain can produce first-person experience. This is hardly a problem for

---

[1] George Santayana *Scepticism and Animal Faith: introduction to a system of philosophy*. Charles Scribner's Sons, 1923.

[2] One might call it the *realizing faculty.* In the Eastern traditions, it is called *maya* or *samsara.* Related terms might be: illusion, involuntary realism, non-reflective commitment, or even "assertoric force" (see Patrick Butlin, Robert Long, et al "Consciousness in Artificial Intelligence: Insights from the Science of Consciousness" [arXiv: 2308.o8708v3 {cs.AI}2023, p38].

[3] Herman J. Saatkamp Jr. *A Life of Scholarship with Santayana*, 2021, Brill, p10.

the brain, of course, which copiously produces that experience on a daily basis. The challenge is to understand how it does that. This is an odd perplexity for several reasons. The naturally outward-looking mind habitually tries to understand things as causal processes in the external world; but these include neural processes in the very brain that is trying to understand itself. Neural processes somehow give rise to the experience of that external world; but that experience includes the neural processes by which the brain tries to understand how such processes give rise to experience… and so on in an endless recursion. In speculating about consciousness, the conscious subject is caught in a loop. The whole thing bites its own tail.

Furthermore, the circularity of our epistemic situation is general, not merely a feature of sensory perception. Science, too, is a form of cognition, and the truths it seeks are no less dependent on the inquirer than the truths sought by ordinary cognition. Both ultimately are survival strategies: we see and conceptualize in ways that allow our cognition to exist. While normal perception seems like a transparent window on the world, which we take for granted, we also know that it is produced in the nervous system, and shaped as much by the biology and the needs of the organism as by the external world. Logically, we must surmise that this applies also to scientific cognition, which is similarly a joint function of the observer and of the world observed.[4]

What we experience and call reality is the brain's natural simulation, a virtual reality we implicitly believe because otherwise we could not be here.[5] Dream and imagination are simulations within this simulation, requiring a way to distinguish between levels of simulation, to know where to commit belief. While none of this implies that no real world exists outside the brain, it does confuse our understanding of it, since our only access to that world is (circularly) through the brain's simulation of it.[6] The brain is literally sealed inside the skull and cannot get outside it to have a direct look around for itself! Even our scientific vision of the world, like our natural perception, is part of its simulation. That vision attempts to compensate for the limitations imposed by our embodied state—for example, by using sensitive instruments in place of the natural senses. But the concepts of science, and even its motivations, remain intimately tied to our nature as biological organisms. The same animal faith that makes us believe our senses gives us confidence in our scientific constructs and political or religious beliefs—for reasons that are only obliquely related to objective truth.

The problems posed by consciousness are compounded by language, that eminent product of our human consciousness. Language works in general by lumping experience into artificial categories and by metaphorical extension. Confusion is built into this arrangement, since words unavoidably have multiple meanings. Mental terminology is particularly ambiguous. Precisely what do we mean by *consciousness, mind, sentience, intelligence, thinking,* etc.? Many philosophical debates involve squabbles over definitions, which may nevertheless reflect genuinely different orientations. Idealists, for example, may hold that consciousness does not

---

[4] A "bivariate function" as when an equation has two variables, which cannot be solved without a second equation or additional input. I call this the Equation of Experience., although it applies to behavior as well as to 1st-person experience.

[5] "VR is the best technological metaphor for conscious experience we currently have." [Thomas K. Metzinger "Why Is Virtual Reality Interesting for Philosophers?" Frontiers in Robotics and AI September 2018 | Volume 5 | Article 101, p3]

[6] "Understanding" implies a metaphorical ground on which to *stand under* our epistemic situation. But there simply is no such ground that is not part of the simulation itself.

require a reductive scientific explanation because it is fundamental to matter, or at least to life. Yet, underlying all such discussions lurks a more basic occasion for misunderstanding. This is the issue of point of view. In language, this is referred to as *grammatical person*—for example: first-person (I or we) or third-person (he, she, it, they). But verbal description always involves a *literal* person speaking, whether from a first-person or a third-person point of view.[7]

Point of view is built into the fundamental subject-object relationship as an epistemic stance. Failure to clarify which point of view is invoked in a given context is the source of endless confusion and misunderstanding. 'Consciousness', for example, can refer to first-person experience; but it can also refer to a phenomenon in the world, observable behaviorally from a third-person perspective. Similarly, sentience, belief and desire can be viewed first-personally or third personally.[8] These distinctions become important when considering artificial intelligence. If "semantic understanding," for example, is understood behaviorally, then large language models can be pragmatically considered to understand what they are saying. (We are not obliged to consider what "it is like to be" them.) Whether or not an AI agent understands what it is doing is only important to the extent that it affects its measurable performance of humanly relevant tasks.[9]

A distinction must therefore be made between the physical concept of *information* (aka Shannon information) and semantic meaning, which requires a subject to interpret symbols that have no intrinsic meaning in themselves. Information processed in the brain can be regarded as physical processes or as symbol manipulation—from a neurologist's point of view or from a programmer's point of view. But either way, the brain's job is to give that information meaning, from the brain's own point of view, which it can only do by grounding in external reality its internal connections or transforms (its "symbols").

As opposed to a spiritual identity—and the godlike ideals of benevolence and objectivity we hold dear—Santayana claims that "the spirit that actually breathes in man is an animal spirit… it has a material station and accidental point of view, and a fevered preference for one alternative issue over another." Indeed, that "fevered preference" is not only the basis of consciousness but a prerequisite for life itself. Things necessarily matter to organisms, in a life-and-death way. To permit our existence, the universe must be a certain way; and to live as creatures we must perceive the universe a certain way and act within it in certain ways. The creatures that exist, including ourselves, are only here because they take the external world seriously and have learned to deal with it in the specific ways that allow the continuance of their species. Those

---

[7] Imperatives (commands) involve a second-person point of view as well. Propositions (claims) imply someone to receive the communication, an audience to whom they are addressed. While description is someone's claim, formed by the "first" person, it is implicitly addressed to a "second" person, and purports to be about something or someone as the "third" person.

[8] 'Sentient', for example, is sometimes poorly used to mean a mild form of "consciousness" or a dim version of phenomenal experience, reflecting muddled thinking. Except when describing one's own personal experience, I advocate using such terms to describe *behavior*, in the 3rd-person, not subjective states inaccessible to observers. By the same token, 'intelligence' denotes a separate concept from consciousness. Like 'understanding', it should be treated as a behavioral concept. As we shall see, to understand the relationship between intelligence and consciousness, the latter should also be treated behaviorally: as a form of internal representation unique to autopoietic systems.

[9] Shane Legg & Marcus Hutter "Universal Intelligence: A Definition of Machine Intelligence" 2007, arXiv:0712.3329v1{cs.AI}, sec5.2. Just as a human can perform tasks with more or less understanding, the degree to which an AI understands what it is doing can be crucially important; but that is a behavioral sense of the term.

ways include faith in the fundaments of perception: real objects, substance, space and time, and the compulsion to evaluate stimuli, to judge things good or bad for oneself or one's kind.

This biological grounding situates our existence within the broader dynamics of life on a solid planet. Natural life evolved as a whole, and survives as a whole; it could not have arisen otherwise. All the parts constrain the other parts at each phase, producing a self-sustaining biosphere of total recycling. Nothing in this biosphere can persist that is not recycled (that is, which does not die) and whose products of decay cannot be used by some other ephemeral creatures. Death is the price of natural life.

Thus, we find ourselves in a ready-made world we did not ask for or create. We find ourselves in a body we did not design and which does not endure. As infants, we learn the ropes of how to operate this body and accept it, just as some people must learn to operate a prosthetic limb and to identify with it. Throughout life we are obliged to negotiate the world in terms of the needs of the body and as seen through its eyes. This natural state of affairs—the state of embodiment—must nevertheless seem awkward and arbitrary to a consciousness that can imagine limitless possibilities. This can be an unwelcome and disturbing realization for a developing mind that must make the best of a given reality it is trying to settle in to. The final reward for a lifetime of such adjustments to embodiment is the insult of death.

The tension between embodiment and imagination defines the human existential situation. For a mind that can conceive itself apart from the animal body, embodiment is an embarrassment. Along with pleasures there are the trials of disease, dysfunction, aging, the ever-present possibility of injury, pain, and loss. There are the psychological challenges of inner conflict, grief, anxiety, disappointment and despair—not least over the inevitable loss of one's life and consciousness. And there are social and moral problems, such as greed, selfishness, violence, tribalism, fanaticism. These liabilities of biological embodiment render us ambivalent toward nature and the body, which can seem at best an instrument at our disposal and at worst an enemy or prison. But they also leave us highly perplexed. How can a physical object even be a conscious subject, a person? How could a conscious subject not be an integral part of the physical world?

This tension between "mind" and "body" is a by-product of our ability to conceive such abstractions in the first place. It is traditionally known as the Mind-Body Problem, since mind is a nebulous catchall for a subject's experience, and body refers not only to one's corporeal appendage but to matter generally. It is intellectually challenging to understand the place of mind in a physical universe. More specifically, the challenge—now usually called the Hard Problem of Consciousness—is to understand how the biological brain can give rise to one's personal subjective experience. The designation "hard" reflects the frustration involved, reminding us that the puzzle consciousness poses to itself remains unsolved.

These challenges posed by animal faith are not merely personal or private, but extend into our shared intellectual and social life. Because we are not just individuals but social creatures, the epistemic issues facing the individual are also collective matters, which become ethical and even moral issues. Objectivity or truth is a value not only for science but nominally for society at large, along with other ideals such as equality and fairness. As social beings, we internalize such values as ideals to pursue for individual as well as social benefit, as virtue or spiritual aspiration. Religion serves as an interface to consolidate individual with collective values. It has long given diverse expression to basic human dissatisfactions with limits imposed by physical embodiment. We've never liked suffering or mortality, for example, and promised ourselves compensation in an afterlife. In modern times, we resist genetic determinism and now

try to use it to achieve greater freedom. We seek to understand our own existential situation in ways that are both collectively and individually empowering, through some form of transcendence—historically through religion and now through technology. When social progress seems beyond reach, the way is still open to the individual pilgrim.

The search for transcendence increasingly takes a technological form. Our science aims to free us not only from gravitation and chemistry, but even from mortality. We seem to be the only creature on this planet with a notion of determining our future as a species, with the possible capacity to secure it through technology. The biological body gave rise to mind, which in us seeks to free itself of its limiting substrate. The seemingly unique human mind, while a product of natural evolution, has open horizons beyond apparently mindless processes of adaptation and selection, of physical cause and effect, so that we propose to intentionally take evolution further and into our own hands.

This ambition raises fundamental questions about the viability of life beyond its natural constraints. How does *artificial* life fit into the picture? The evolution of natural life and mind depends on natural selection. Can natural selection be simulated to result in artificial life and mind? Is it possible to design an artificial ecosystem? Must all its constituents be recycled, as in the biosphere? Whatever the theoretical answers, the human technosphere does not yet practice universal recycling, but has created a new category of being: trash. If we seek to create a permanent existence for ourselves, both on the species level and individually, will that simply end in more trash?

The project to transcend nature must reckon with the conditions that make natural life possible and sustainable. Conceptually, machines and other designed artifacts are eternal; they wear out or are broken because they are made of real rather than idealized materials. On the other hand, some of those materials are potentially more durable than flesh; hence the appeal of migrating humanity to machine life. But the characteristic robustness of natural life is its ability to self-repair and adapt. To persist, artificial life would have to emulate those abilities of self-creation and self-maintenance. The sticky point is that such abilities may depend on recycling—just as the integration of the biosphere could not exist without the built-in obsolescence of mortal generations. Can an artificial form of life exist apart from an ecology of other such forms? Does belonging to such an ecology imply mortality? Can we emulate the desirable features of natural life without the less desirable ones?

Such considerations lead to the question of humanity's long-term trajectory. As mind attempts to disengage from body, our species attempts to disengage from its biological roots. The future of humanity will depend on how humanness is redefined in the face of "technology's insidious drive to replace the Real, to sever thought from embodiment, and to tear apart whatever gossamer threads still bind us to nature and to our material human communities."[10] Much of our current identity and mentality has been determined by biology—in particular, by our primate heritage. But, to the extent we are able to intentionally reshape it, our future identity is up for grabs. While it cannot exist apart from some form of embodiment, mind is now positioned to determine its future embodiments. Our human successors may not be biological. We do not yet know the limits of possibility in this domain, which makes it worthwhile to explore.

The very existence of the biological individual, as well as the species, depends on animal faith—the commitment to one's cognition. Could a meta-biological species and individual be freer from this dependency, offering more control over its own internal processes and behavior?

---

[10] Erik Davis *TechGnosis: myth, magic, and mysticism in the age of informatio*n. 1998/2004, Five Star, p294.

For one thing, the possibility of replication inherent in digitation could minimize existential risks to the individual, who would have the security of back-up copies. But copying would change the meaning of individual identity and perhaps the meaning of life. Is animal faith an inescapable concomitant of biology, or of embodiment more generally? If so, would loss of consciousness be an acceptable price to pay for liberation?

As AI becomes ever more sophisticated at imitating and exceeding human performances, it might track the world in human-like ways or better. However, unless it exists for the sake of a body (although not necessarily a *biological* one), I do not believe it could have the inner life we know as consciousness. There can be virtual experience by a human subject, but not a disembodied virtual subject. There would not be anything it is like to be it. Of course, that is not a verifiable assertion. While it is partly out of political correctness that we humans assume there is something it is like to be each of us, that has a reasonable basis in our common biology as members of a species. Whatever we could potentially have in common with machines would be based on suspect comparisons, on the assumption that our nature as an organism could be grasped and structurally approximated in some artificial system with its own embodied relationship to a real environment.[11] What is real to us is what we can affect, and be affected by, in such a way that allows us to exist. The question is whether that relationship can be simulated.

Generative models are the latest development of the computational metaphor to shed light on the natural operations of mind—and to compete with them. The inventive power of these AIs mimics that of the brain—to imagine, dream, fantasize, and hallucinate.[12] The natural brain's commitment to its own productions is the very thing we are calling animal faith. Our personal experience of that commitment permeates our consciousness, often to our detriment. Would artificial minds necessarily be saddled by a similar commitment to their premises, strategies, operations, and outputs? If they have an inner life, what would their experience of that commitment be? Animal faith seems built into biological life. Need it be built into artificial life? The answer cannot be simple.

---

[11] The concept of *organism* should play a key role in artificial intelligence, and a bigger role even in the life sciences which have generally embraced mechanistic explanation. See: [Tom Ziemke & Robert Lowe "On the Role of Emotion in Embodied Cognitive Architectures: From Organisms to Robots" Cogn Comput (2009) 1:104–117, p114]

[12] Paul Smart in *The Mind-Technology Problem: Investigating Minds, Selves and 21st Century Artefacts* Robert W. Clowes et al (eds) Springer 2021, p.189-90.

Chapter Two: Fiat, or Intentional Connection

> "Everything is the way it is because it got that way." —D'Arcy Thompson

To explain mind, though embodied, causality is necessary but not sufficient. For, though it has a material basis, mind exhibits *intention*. While an electrical circuit in an appliance can be described causally—as a flow of electrons, for example—it can also be described in terms of the design of the appliance, the purpose it is supposed to serve, how people will use it, etc. In other words, it can be described in intentional terms. Similarly, the functioning of natural organisms can be described causally, in terms of physical processes within them and their environment; and it can also be described in intentional terms, when we try to understand the logic of their internal structure and why they behave as they do, as though from a designer's perspective.[13] A strictly causal description cannot account fully for manifestations of life and mind. Neither can it account for artifacts, which involve the intentionality of designers and users. Of course, natural organisms are not human artifacts and we do not assume intelligent design of nature. Yet, they are distinguished from "inert" matter precisely by the fact that causal description cannot account completely or adequately for their behavior, let alone for any interior subjectivity. Apart from an observer's analysis, organisms manifest their *own* intentionality. Their purposive behavior cannot be reduced to physical causes, even though it depends on them.

Intentionality has a convoluted history in philosophy as "aboutness," which is essentially a linguistic notion entailing reference. But, since only humans use fully grammatical language, let us broaden the concept and reframe intention outside the context of human language. Intention, in this broadened sense, is an internal connection made within an *autopoietic*[14] system for its own purposes. Such a system is self-defining, self-maintaining, and—in the case of life—self-reproducing. The sort of internal connection involved might be a synaptic connection made within a brain. Potentially, it could be a connection made within an artificial system (if, indeed, such a system can be autopoietic). In any case, the connection is made *by the system itself*, not by an external observer, programmer, or other agent. If we look at the inputs and outputs of an organism, for example, we see that internal causal connections between them do not, of themselves, reveal the purpose for which they are made, or how they serve the existence of the

---

[13] As ways of looking, causality and intentionality each project attributions of the observer onto the system observed. Both are types of description from a 3rd-person point of view. Yet, apart from such descriptions, the system may have its own purposes and mentality, and can initiate causal processes within itself. The teleology, purpose, or goal of such an agent must be distinguished from the observer's projections (as in Dennett's "intentional stance".) The function of something (such as a part within a whole) might seem to entail no more than space-time description; but it unavoidably implies the role or purpose it serves within the whole—i.e., teleology. The task is to distinguish whose purpose or goal is involved: that of the agent itself or an external observer/designer. Every describer is an observer, but not every observer is a describer. Every observer is an intentional system, but not every intentional system is an observer. Every intentional system is a causal system, but not every causal system is an intentional system.

[14] Literally 'self-making.' The term was introduced by Francisco Varela and Humberto Maturana in their seminal work, *Autopoiesis and Cognition*, D. Reidel,1980. The original title (in Spanish) was *On Machines and Living Beings.*

creature, which does not simply *react* to stimuli but actively *responds*. Something far more complex is going on than the physical causation of action-reaction.[15]

The concept of intentionality proposed here includes conscious intention but is not limited to it. Nor is it tied to linguistic or symbolic reference. The reference of words to things is merely one example of the (human) organism's ability to internally make connections, though it serves metaphorically as a paradigm. More abstractly, *intending* is the internal act of an *agent*[16] that maps one domain to another for its own reasons and purposes. An agent *makes* connections within itself, in contrast to events simply *happening* within it, or to it, which an observer might trace to physical causes. Such connections are naturally embodied in neural connectivity, and could potentially be embodied in artificial systems.

An agent is an embodied complexly organized physical system. Agency implies a subject-object relation between a bounded system and an environment. While organisms are causal products of natural selection, they are also actors in their own fate. Indeed, agency may have evolved to compensate the haphazard pace of random mutation.[17] And *embodiment* implies more than being physical; it requires a certain active *relation to the world.* Embodiment is a relation of an autopoietic system to its real environment, in which events *matter* to it, because they are ultimately significant for its continuing existence.[18] Every living organism stands in this relationship to the world, entailed by its participation in the system of life. It is a relationship inherited through natural selection and maintained and refined by the individual organism. The survival mandate implies that the organism has priorities that reflect its needs in relation to its environment, and which motivate its behavior.

The computational metaphor may be our best tool for understanding mind and consciousness in scientific terms—precisely because it invokes reason, intention, and agency as explanatory principles instead of efficient cause as understood in the physical sciences.[19] Sensory input thus serves less as a cause than as a prompt to the organism to respond: more a suggestion than a command to a complex organism that can consider options.[20] As a cognitive system, the organism issues commands within itself, by making internal connections, and thus (so to speak) programs itself. This act of assertion or internal command I call *fiat,* which means, literally, "let it be done." Animal faith is the willingness or compulsion to obey such commands, to which

---

[15] What Aristotle called 'efficient' cause.

[16] An agent is an autopoietic system, whether natural or artificial, which acts on its own behalf as distinguished from reacting to causes. This is a narrower definition of agent than merely "something which accomplishes something." An agent acts for its own reasons, from within itself, not simply because of external causes. Causal description is not incompatible with agency, but makes no essential distinction between "external" and "internal," since there are no truly closed systems from a physical point of view. But the agent itself may define itself in contrast to its environment.

[17] Phillip Ball "Organisms as Agents of Evolution" John Templeton Foundation, April 2023, p12 and p25.

[18] No AI, robot, or other artifact is yet an autopoietic system, with an embodied relationship to its world. The question remains whether that relationship can be created artificially.

[19] However, it is not a single metaphor. The older version invoked top-down programming—symbol manipulation. The newer connectionism holds that cognition is not a matter of manipulating symbols but of "dynamic patterns of activity in a multilayered network of nodes." [Stevan Harnad "The Symbol Grounding Problem" Physica D 42 (1990), p337] Yet, the symbol manipulation perspective is useful precisely because it puts one in the shoes of the system. What is symbolic to the organism (or system) is the connections *it* makes to serve as symbols. The meaning of a symbol is not intrinsic to the symbol, but to the system, i.e. the organism. A connection can be symbolic if the system makes it so; but "connectionism" reflects a human observer's 3rd-person perspective.

[20] Phillip Ball op cit, p20.

Coleridge alluded as the willing suspension of disbelief. Philosophers have referred in various ways to diverse aspects of this power of connectivity, especially its irresistible subjective persuasiveness. Descartes calls that *judgment.* Schopenhauer speaks of *will*, Helmholtz of *unconscious inference*.[21]

Fiat is an internal exercise of agency, an internal communication—an injunction of the organism to itself. Connections are made within the system and accepted as commands.[22] Fiat is thus the very basis of consciousness. Like gods, monarchs or magicians—indeed, like mathematicians, cooks, or musicians—we simply declare the inner show into existence moment by moment, as though by royal decree, with a magic wand, or by mere supposition. And then we naturally *believe* this creation!

The essence of a cognitive system lies in the very plasticity of its internal connections, which are in themselves arbitrary, just as words are arbitrary symbols for meanings. There is no pre-existing or absolute meaning inhering in a given neural connection, as there is not in any symbol, nor any inherent reason why the connection should be made or endure. Its significance lies in what the system itself makes of it. The meaning, to the system, of any connection or configuration of connections depends therefore on the state of the system as a whole. While a local group of neurons in a human brain, or even a single neuron, can initiate a conscious experience if stimulated electrically, it is the *brain as a whole* that produces the experience by giving meaning to the patterned firing of those neurons. It is meaningless to think of individual neurons or circuits as conscious or capable of having experience if isolated from the brain as a whole.

The brain's act of fiat can be directly experienced in such phenomena as perceptual completion effects and the "filling in" of the visual blind spot. In the latter, the experience of continuity of the visual field is the brain's way to represent to itself its (true) belief that (despite the anatomical blind spot) the external world is actually continuous. The brain affirms that conviction by an act of fiat (which is also an act of imagination and faith). It is an hallucination that ignores the sensory discontinuity and literally "fills in" phenomenal experience in the visual field between the enervated retinal areas on either side of the un-enervated area. However, the enervation of *all* sensory surfaces is similarly discrete, with gaps between receptors; those gaps *in turn* must be "filled in" phenomenally, but on a finer scale. This occurs temporally as well as spatially, so that there is apparent continuity of motion, for example, and experience in general appears continuous. That is, in *all* cases the brain asserts continuity across discrete structures or events when their discreteness is irrelevant, just as it asserts continuity between frames of a

---

[21] More recently it's been called the *ego tunnel* and the *reality illusion.* See Thomas Metzinger *The Ego Tunnel: the science of the mind and the myth of the self.* Basic Books, 2009. Also: Ralph Strauch *The Reality Illusion: how we create the world we experience.* Theosophical Publishing House, 1983. Related terms include reification, projection, transparency, the realizing faculty, maya, illusion, etc.

[22] "It may be helpful at this stage to realize that the primary form of mathematical communication is not description, but injunction. In this respect it is comparable with practical art forms like cookery, in which the taste of a cake, although literally indescribable, may be conveyed to a reader in the form of a set of injunctions called a recipe. Music is a similar art form, the composer does not even attempt to describe the set of sounds he has in mind, much less the set of feelings occasioned through them, but writes down a set of commands which, if they are obeyed by the reader, can result in a reproduction, to the reader, of the composer's original experience." [George Spencer-Brown *The Laws of Form*, 1979. Quoted in John O. Campbell & Michael E. Price "Universal Darwinism and the Origins of Order" in Evolution, Development and Complexity: Multiscale Evolutionary Models of Complex Adaptive Systems, edited by G. Georgiev, C. L. F. Martinez, M. E. Price, & J. Smart. 2019. Springer Publishing.]

motion picture. This is the sleight of mind by which the world has an analog look despite sensory digitation. Indeed, it is the trick by which the world has any appearance at all!

Phenomenality[23] is generated by the sort of acts of fiat demonstrated in the visual blind spot and other perceptual completion effects. For an intentional system, the meaning to itself of its internal communications is analogous to the meaning that emerges for a human language user in the act of reading or writing, of speaking or listening to speech. In that act, the brain translates linguistic symbols (visual or aural) into mental images, thoughts and feelings, or vice versa. Assigning meaning to its own *internal* language evokes a phenomenal world in a similar way—as words evoke mental images, or as the continuity of the blind spot is evoked. The self-effulgence of qualities in sensation (*qualia*[24] such as the redness of red, the hurtfulness of pain) emerges in much the way that meaning in language does, by an inner act of declaring it so. In natural language, sounds and symbols carry meaning as words through a constructive process—in other words, by fiat. Phenomenal qualities are thus comparable to intelligible meanings that emerge from the babble of spoken syllables or the squiggles on a written page.[25] The creation of sensory experience is *like* the creation of mental imagery in response to language—that is, from internal connections an observer could call abstract symbols. Phenomenality *is* the brain communicating with itself in its own internal "language of the senses," which represents not only the world, but the organism's valenced relationship to the world.

As a cognitive system, an intentional system is a symbol system. *Some* symbol(s) must be chosen to represent the emergent meaning. However arbitrary the symbol is in itself as a token, it will inevitably come to seem *imbued* with the meaning it conveys via the connection that is made. Hence, it is misguided to ask why grass appears green rather than red, for example; rather, the experience of greenness is what it is by virtue of its persistent association with grass and other verdure. Given consciousness as a symbolic system, greenness is the way we visually experience the associations related primarily to chlorophyll.

Similarly, pain stands for something, such as tissue damage, as well as compelling a response. We do not normally question the reasons for the brain's internal connections, to which we do not have conscious access. Yet, it is only from an outsider's perspective that they can appear arbitrary, conventional, uncompelling or questionable—because the observer is not in the position of being the agent that makes that connection. From such an external point of view, it may then appear mysterious that arbitrary symbols (connections) can carry meaning at all.

Sensory qualia are thus not something gratuitously added to the information they represent; nor are they caused by it, any more than words are caused by the things they represent. Rather, qualia are a *version* of that information, which an internal agent presents to itself synoptically via phenomenality. In other words, they are how the embodied subject first-personally presents to herself information that an observer also might detect by means of

---

[23] I use this term, 'phenomenality,' to mean the totality of anything that can be consciously experienced—including dreams, hallucinations, mental images, thoughts, feelings, etc., as well as sensory experience. The more common term *phenomenology* is reserved to mean the study of phenomenality.

[24] "In philosophy of mind, qualia are defined as instances of subjective conscious experience…Examples of qualia include the perceived sensation of *pain* of a headache, the *taste* of wine, and the *redness* of an evening sky… as qualitative characteristics of sensations…" [Wikipedia: qualia]

[25] Similarly, conventional algebraic symbols gain numerical significance by agreement with the mathematician's fiat: '*let x stand for such and such…'*

laboratory equipment and describe in terms that are third-personal, physical, propositional, and quantitative.[26]

Mental images resemble their full-blown sensory cousins, to a degree of vividness that varies among individuals.[27] Yet, the differences between them provide clues to what is required for actual sensory phenomenality. For one thing, mental images or memories convey only the detail they already embody, based on prior sensory input. For most individuals, that is far less than provided in the original stimulus. Unlike a live sensory image, a memory or visual imagining cannot be searched for more information than it already stores and graphically presents. A retinal image, in contrast, is an optical image constantly updated in real time (or nearly); it is an ongoing source of live data. The visual field itself changes as the world changes, but is also continually refreshed through eye saccades. This constant renewal of an external source of sensory input makes the key difference between a memory or mental image and vivid real-time sensory experience.

While an invention, the "show" of phenomenality is continually updated and guided by input from the senses, which validate it. Like reading tea leaves, the patterns discerned auger for actions that generally facilitate existence. (While crossing the street, it pays to see that looming shape as a rapidly approaching bus!) Therein lies the subjective meaning of *realness* as a quality of experience. What naturally appears to us as real refers to our dependency on a world "out there" which we did not make and whose rules we did not choose—a dependency against which we may also rebel, having imagined freedoms beyond it.

Internal connections must not preclude survival, but otherwise are not obliged to correspond to reality. This plasticity is both a boon and a bane. The inherent freedom of the cognitive agent to make intentional connections yields not only the ability to track reality but also the ability to hallucinate, lie, fantasize, invent, dream, imagine and create. The context of the organism as a product of natural selection favors connections that further its interests or those of its kind, or at least do not preclude its existence. The organism's natural orientation is advantage rather than truth. But such useful connections occupy only a small zone within a vastly larger space of theoretically possible ones that don't kill you.

As biological organisms, we could not have survived if we did not take experience at face value and seriously. The senses reveal to us a world of real consequence outside the skin, not a movie running inside the head. (The idea that there nevertheless *is* such a movie is rather modern, reflecting a philosophical maturity.) It continues to serve us well most of the time to believe the illusion presented in the panorama of consciousness. Though technically we may know better, that credulity remains our default state. Fiat is the power to create that show, and animal faith is the inclination to believe it.

Like news reporting, experience must bear at least a grain of truth to be believable; yet, it cannot be the literal or whole truth, which would be impossible to portray. In the face of limited information capacity, the brain has surprising artistic license to select connections and make

---

[26] The purely qualitative aspect, "private" to the subject, bears information pertaining to the sensory mode as well as the external world, such as color hue for vision and auditory tone for hearing.

[27] In contrast to aphantasics (who claim not to experience vivid mental imagery), there are hyperphantasics who claim that their mental imagery is as vivid as their actual sensory experience with eyes wide open. If this were literally so, life would be very confusing for them! Indeed, it no doubt *is* for some people, labelled schizophrenic, who experience very real-seeming hallucinations and cannot tell the difference. In any case, the brain is inherently plastic. If it is typically organized a certain way, it should not be surprising that there is some variation.

gratuitous ones. If that was all there is to it, we could dismiss religion, magic, myth, and even science as merely arbitrary expressions of human creativity. But animal faith adds a dimension of necessary belief. Realness implies the need to take experience seriously and even literally, precisely because the connection is *not* gratuitous or arbitrary but makes a real difference. Fiction you can take or leave as entertainment; reality you cannot.

The paradox of belief is that an agent has some freedom over where it credits realness; the dilemma is that it sometimes credits inappropriately. While it might seem perverse to believe a falsehood, human freedom lies precisely in the ability to do so. After all, a principal use of language has always been deception, including self-deception. It is quite possible to live in an utter fantasy as long as it doesn't kill you. In fact, some illusions favor survival better than does the literal truth. While nature *permits* a latitude of fancy in how we perceive, the longing for freedom *motivates* us to be fanciful. Perhaps this helps account for the persistence of religion, myth, and magical thinking throughout human existence, as well as the enduring appeal of storytelling and art. Ideas can be unrealistic without hindering survival.

----

2 The MBP does not involve an elusive "property" for which a causal mechanism cannot be found; it is rather a relationship between organism and world. It is not that some "property" of the brain gives rise to consciousness; the brain, as an agent, gives rise to it "intentionally" as a representation for its own use.

2 [2][7]Monitoring could be discrete sampling, but not offline for long periods unless under guaranteed safe conditions (cf sleep and ability to be alerted to wakefulness. Also: seals asleep .with one eye open).

2 [2][8]Emotions involve chemical changes in the body, which are more pervasive than specific pathways of neural connection; they would affect a broader number and range of cells, mobilizing the body as a whole rather than specific muscles, organs, or brain sites.

4 Three "dimensions" or approaches to pain and other phenomenality: (1) macro behavior; (2) 1st-person experience (what it is like); (3) functional organization (micro behavior neural wiring). Just as color can be judged and described 1st-personally and also can be measured and described 3rd-personally, so pain is judged 1st-personally or behaviorally (3rd person description). The third domain of description is the neural and chemical connectivity that underlies (presumably) both experience and behavior. Parallels are assumed on the basis of gross apparent anatomical isomorphism; finer isomorphism would be more conclusive. Another approach would be reverse-engineering: finding the structure that would follow from a design intention (what connectivity or program would produce the desired result?).

4 The MBP can be defined as the challenge to explain phenomenal experience in terms of physical processes and entities. While that seems insurmountable, the inverse of this challenge is the problem to explain physical processes and entities in phenomenological terms. This seems to be rather a matter of inference, tracing scientific deduction from sensory evidence (and its instrumental counterparts). We are used to this, which corresponds to conscious thought processes. The inverse is also inference, though unconscious (cf Helmholtz). It seems inscrutable simply because it is not conscious; we do not identify with it. Yet the logic is the same, if in reverse.

Chapter Three: Liabilities of Animal Faith

> "The brain is not an organ of thinking but an organ of survival, like claws and fangs. It is made in such a way as to make us accept as truth that which is only advantage."—A. Szent-Gyorgyi

1 Animal Faith, Action, Evolution

Animal faith follows inexorably from embodiment. Attention is directed outward upon objects whose seeming reality reflects creaturely needs. Survival depends on continuous coupling to a physical and social environment, which is enforced by affective compulsions such as hunger, pain, fear, and desire. The organism must take these perceptions as real and binding. Doubting the reality of danger, for example, would be maladaptive. In Santayana's words, as "an expression of hunger, pursuit, shock or fear…animal faith posits substances, and indicates their locus in the field of action of which the animal occupies the centre."[28] Animal faith is a design feature that ensures prompt, contextually appropriate action within this field of action, which has a limited epistemic horizon. The smaller the literal sensory horizon, the greater the need for immediate reaction to relevant stimuli such as predators or prey.[29]

The fact that we are here is proof that the brain's map of the external world—though not literally true—is at least adequate, insofar as it facilitates survival. While natural selection reflects fitness, it is not obliged to reflect reality, provided only it does not prohibit existence. The animal, however, must act to maintain itself. Those actions that work successfully toward that end will be reinforced and sustained. The creature then rightly believes those actions to be "correct" insofar as they promote its wellbeing and continued existence. The objects of its actions will appropriately appear to it as desirable, fearsome, appealing or disgusting, near or far, etc. It has faith in such appearances, which relate to its needs. In other words, it has faith in the map. It has no direct need for truth apart from such functional appearances.

While this applies to human beings along with other creatures, we have developed compensations for the sheer earnestness and blindness of animal faith. We have a concept of truth, as above and beyond mere appearance, belief or opinion. We know the difference in an abstract sense. We can question specific beliefs or specific perceptions; yet we hardly escape the need to believe. We may be driven to higher ground, often more abstract, on which to stake belief. Yet, we remain animals that need to act on a tentative certainty. And we do not always avail ourselves of those tools of rational thought that can compensate for animal faith.

2 Mind–Body Dualism / Disembodiment

From the mere fact of thinking, Descartes had concluded—illogically—that he must exist as a "thinking thing," quite distinct from the existence of his body, which he viewed at best as a

---

[28] Santayana op cit, p214.

[29] Fleming SM, Michel M. Sensory Horizons and the Functions of Conscious Vision. Behavioral and Brain Sciences. Published online 2025:1-53. doi:10.1017/S0140525X25000068, p2-3. For the human creature, the epistemic horizon is larger than the visual horizon, even with the aid of telescopes.

machine, at worst as an illusion.[30] Had he focused on feeling he might have concluded, more like Santayana, that he should identify rather with his body and its drives.[31] Santayana recognizes the dilemma of a mind dissociated from its embodiment:

"Not sharing the impulses of his body, he would regard it as a ridiculous mechanism; and the bodies of others would be ridiculous mechanisms too, with which he could feel no sympathy…[He] would be all scorn and lamentations for the life of the world… His sympathy, if it survived at all, would be sublimated into pity for the spirits chained to those bodies by their sin and ignorance, and perhaps not even struggling to be free, but suffering in those prisons perpetual pain and dishonour."[32]

This expresses well the longstanding ambivalence of the human psyche toward embodiment, experienced both as the source of carnal pleasure and physical pain—and of intellectual pleasure and emotional suffering as well. This ambivalence has endorsed the notion of the body as a prison for a soul conceived to be potentially free of material constraints. Mass media may aggravate this cultural alienation from the body, by substituting screens and "data" for face-to-face experience with real bodies.[33] Modern technologies of simulation have a direct experiential effect on our sense of embodiment, as well as philosophically encouraging the idea of disembodiment.

For Plato—and for Christianity following in his footsteps—the soul was independent of the body in the way that his Forms are independent of material reality. The idea of disembodied mind abstracts and reifies the notion of consciousness—phenomenality—as a possibility detached from the life of the body. The human mind enables the indulgence of such abstractions. But, "out of body" experiences notwithstanding, disembodied experience is an oxymoron. Minds exist to serve bodies, not the other way around; nor can mind exist independently of matter—in the form of digital files, for example.

Like Darwin, Santayana makes clear that our nature is rooted in biology. Our consciousness is shaped and limited by the needs of the body. Any possible liberation from the body's impositions lies in full recognition of this fact rather than denial. This has not prevented us from conceiving disembodied ideals of "pure" consciousness, "pure" reason, and "pure" intelligence while at the same time ignoring how utterly parochial our intelligence actually is, and how dependent our consciousness is on the body and its animal faith. But the ideal can only be actualized by acknowledging the real.

3 Perception, Simulation, Phenomenology

---

[30] Descartes translated his treatise into Latin, in which "Je pense, donc je suis" becomes the famous "cogito ergo sum." *Cogito* has a more inclusive sense than *penser*. His syllogism would more logically have read: *cogito ergo cogitationes sunt.* Without concluding a personal agent who does the cogitating, roughly that would read: "I experience, therefore there are experiences." The sense of self is a psychological inference, not a logical one

[31] In fact, the Mind-Body Problem may reflect the difference in functions of the myelinated exteroceptive nervous system, responsible for cognition of the external world, and the non-myelinated interoceptive nervous system responsible for feeling and homeostasis. This could be the neurological source of the disjunction between the third-person view of reality as external to the body and the first-person view that derives from the body's self-regulation —specifically the need to valuate stimuli in relation to the body's state and needs, which is the basis of phenomenality, and of sensory qualia in particular.

[32] Santayana op cit, p215-16.

[33] Erik Davis *TechGnosis: myth, magic, and mysticism in the age of information,* 1998/2004, Five Star, p140, citing Hakim Bey, "The Information War," in Mediamatic 8, no. 4 (1996): 61.

Unlike a map, which symbolically represents the territory, a traditional photograph seems to literally resemble the real scene, captured in two dimensions. Digitation has played havoc with the reliability of photography to represent reality. Yet, in any case, the illusion of resemblance results from the fact that the brain processes the image presented to the eye in the same way it processes the optical image presented to the brain (on the retina, which is also two-dimensional). Any input, whatever its source, is fodder for the brain's inventive connectivity, for processing that is necessarily symbolic and selective. It's as meaningless to think that the image in your mind resembles the external world as it is to think that words literally resemble the things they describe.

In the brain's simulation, qualia represent properties of actual things. They tell us something useful about the world. Yet, along with the apparent subjectivity of qualia such as color and smell, even the three-dimensionality of visual perception is a mental construct. For, the senses are not open windows on the world, but remote sensors wired to the brain. What we experience consciously is not the territory but an interactive map, which charts properties about the territory obliquely inferred from data that are remotely collected. The symbolic character of this map has also been likened to a computer desktop, whose icons represent not things in the world but functions in the computer.[34]

Virtual reality is not only a powerful expression of the computational metaphor, but also a research tool. According to current theory, the impression of realness is a signal confirming successful predictive processing. As such, it should perhaps not be tampered with, yet the virtual reality analogy suggests that human phenomenality could potentially include a "volume control" for experienced realness.[35] The ability to distinguish real from imagined, objective from subjective, seems crucial for survival, but is only relatively reliable and varies among individuals. Some people can mentally visualize vividly at will and others hardly at all. Some "hyperphantasics" claim that their mental imagery (for example, visual imagery with eyes closed) is literally as vivid as their sensory visual imagery with eyes open. If so, that could be very confusing for them if they experience both simultaneously with eyes open.

In contrast to schizophrenics, most people have stable experience they count on as real, not hallucinatory. Yet, we are fascinated by experience outside that normal range and cultivate it through drugs. One theory suggests that at one time people normally and routinely "heard voices" and saw visions—for example, of guiding angels.[36] In other words, people had not yet learned to tell the difference between external and internal sources of experience—an ability that, in any case, remains fragile.

Most people know the difference between waking reality and dream. In REM sleep, motor activity is normally blocked, and so is sensory input to a degree. Yet dreams can seem very convincing and some people walk (and even drive cars and commit murders) in their sleep. Some people have an experience of "false awakening" during sleep, sometimes confused with out-of-body experience.[37] Such possibilities cast doubt on the meaning of waking consciousness,

---

[34] Donald Hoffman *The Case Against Reality: why evolution hid the truth from our eyes.* W.W. Norton and Co., 2019/2022, p75.

[35] Thomas K. Metzinger "Why Is Virtual Reality Interesting for Philosophers?" *Frontiers in Robotics and AI* September 2018 | Volume 5 | Article 101, p7: "It would be highly interesting for philosophers of mind, dream researchers, and phenomenologists if they could use VR technology to explore the transparency/opacity gradient of their very own conscious experience at will."

[36] Julian Jaynes *The Origin of Consciousness in the Breakdown of the Bicameral Mind* Houghton Mifflin, 1976

[37] Thomas Metzinger, *The Ego Tunnel*, Basic Books, 2009, p134.

giving rise to the idea that all experience is more than metaphorically just a dream, or that we can reliably tell the difference.[38]

4 Narrative, Language, Social Belief

One marvels at the ingenuity of the human imagination: not only the things that make practical sense—like houses, agriculture, machines, technology—but especially the things that make little sense to a rational mind, like gods and demons, ghosts and magic, and perhaps some of the esoteric constructs of science and mathematics.[39] Yet, religion and magical thinking have characterized human culture far longer than what our secular world now defines as rationality.

The ancient Greeks we admire as rational seekers of order seem to have actually believed in their pantheon of rowdy and absurdly human-like gods. The Pythagoreans believed in reincarnation and sacred numbers, using mathematics and music for spiritual training and metaphysical insight. Plato believed in a metaphysical Ideal realm underlying material reality—a surprisingly resurgent idea in modern times. Copernicus thought the planets should move in perfect circles and Kepler thought that angels moved them along their (elliptical) orbits. Newton wrote far more about alchemy and biblical exegesis that about math or physics. Indeed, the early scientists were literally Creationists, and there may be scientists today who believe in the transubstantiation of the Eucharist as well as the Resurrection to come.

The point, however, is not to disparage religion as superstition, but to underline that superstition is endemic to the imaginative human mind, since belief is built into our animal nature. It's also built into language, which confers the nearly magical power of fiat to call things into being simply by defining them.[40] Outrageous beliefs are possible because a story can easily be preferred to fact. Imagination and story have the advantage of being our own creations and not something foist upon us by the external world. A story can make sense, be consistent, clear, predictable. Reality, on the other hand, is fundamentally ambiguous, confusing, elusive, threatening, messy. Experience only makes sense to the degree it can be assimilated to a concept or story. (Despite the known facts, for example, it made sense to the ancients that a year should have the tidy round number of 360 days.) In general, sensory input is assimilated to a narrative that is supposed to make sense of it, and upon which actions can be based which, in principle at least, help us live. The story does not need to be true, only useful. It only needs to permit our existence, which gives a wide latitude to imagination and belief.

While the self-conscious narrative mind can reflect, plan, create meaning and communicate, it can also lie. Language may have been invented to deceive or mislead and still serves this function. It offers the possibility to manipulate the thoughts of others, and even one's own. The narrative mind is susceptible to delusion, ideology, over-simplification and the extreme compression of information apparent in myth, dogma, algorithms, and newspaper headlines. Indeed, no concept, story, or formula can be literally true. It can only correspond to reality in the ways that a map can correspond to the territory it represents, or that words can correspond to the reality they describe—namely, by some particular mapping expressing some particular intention.

---

[38] As the ancient philosopher Chuang Tse (Zhuang Zhou) wondered, "Am I a man dreaming he is a butterfly or a butterfly dreaming he is a man?"

[39] Not the *current* generation of scientific ideas, of course, which are always true!

[40] "Like many ancient peoples, the Egyptians believed that a name captured the essence of a thing, but they also held that such supernatural power lived in the inscriptions themselves—that spelling was, in fact, a spell." [Erik Davis *TechGnosis: myth, magic, and mysticism in the age of information*, 1998/2004, Five Star, p32.]

Neither a text nor a conceptual model is a snapshot of reality, but a motivated original creation. Descriptions are always incomplete. No matter how detailed, the map is always selective, symbolic, and intentional.

While we think individually, language is a collective matter that shapes our thinking. (How does it shape your thinking to see language described as a collective *matter*, rather than—say—a collective *phenomenon*, collective *issue*, or collective *enterprise*? Each of these alternatives has a different connotation and bias. Language is not literally matter or material, but describing it that way suggests a *substance* with causal powers, reflecting the mind's power of reification.[41]) Communicating with symbols, especially through the transmission of relationships and categories, can save the individual the trouble of "learning the hard way." Once you know that electrical appliances can shock you, there is no need to test this truth on a case-by-case basis. You may recognize on a label the zig-zag icon that also symbolizes lightning (a paradigm of things that can shock you) and also the word *danger*. You don't need to know that this word comes from a medieval root meaning 'the power of an insolent lord or master.' Nor does one need to have first-hand experience of the politics of committees to know that the *chair* is the authority in charge, the *seat* of power.

But unlike sensory experience, abstract concepts and categories are poorly grounded in tangible referents. One can form a mental image of a *specific* chair, for example, but hardly of furniture in general.[42] Or of a specific radio, but not of appliances in general. That doesn't mean that nothing comes to mind when prompted by the thought of a category or generalization. On the contrary, what does come to mind is likely a specific image, perhaps derived from a real memory, which comes to *stand* for the abstraction in the mind of that individual. This is how the meanings of words in general are actually grounded in experience.

All words, being shared by many subjects, are ambiguously plastic. They may have a dictionary definition, which is itself an abstraction. But what the word evokes in a given mind and situation is individual and specific, having more to do with personal experience than collective agreement. It is rather amazing that we manage in spite of this to communicate—when, indeed, we do. But the diversity of these personal referents for words renders it always challenging to know that we are "on the same page" with others. The more abstract or intangible the concept, the greater the problem. What do people actually mean when they speak, for example, of truth, beauty, freedom, democracy, etc.? It is most likely quite individual, based on divergent personal experience, so that people are often talking at cross purposes when using such terms.

Language works by metaphorical extension—to understand one thing in terms of another. Etymology is often enlightening, revealing a hidden logic by which the current meaning(s) of a word evolved from a root deep in history and psychology.[43] A repertoire of guiding metaphors unconsciously shape our thinking.[44] A basic one is spatial orientation: to be up or ascending is

---

[41] The suspicion that words have substantial power is only confirmed by its denial: "Stick and stones can break your bones but words can never hurt you."

[42] Vittorio Gallese and George Lakoff "The brain's concepts: the role of the sensory-motor system in conceptual knowledge" *Cognitive Neuropsychology* 2005, 22, (3/4), 455–479, p466.

[43] A few examples: The English word *body* derives from a Germanic root suggesting 'container.' *Good* comes from an Indo-European root meaning 'to unite, join, or fit together.' *Conscious* comes from a Latin cognate meaning 'knowing together', while *knowing* derives from an I.E. root meaning 'to cut, split', from which *schism* also derives. *Human* comes from the same root as 'humus' (soil) and 'humble'. *Free* derives from an I.E. root meaning 'to love.'

[44] See: George Lakoff and Mark Johnson *Metaphors We Live By* University of Chicago Press, 1980. The authors don't claim to trace the referents of a metaphor in specific sensory experience, but they do "feel that no metaphor can ever be comprehended or adequately represented independently of its experiential basis." [p19].

preferred to being down (grave, in the grave, succumbing finally to gravity); a high status is better than low. Moving forward is better than being left behind. Having inside information is better than being "out of it." And so forth. Language is creative. The problem is that it is also largely unconscious. Using the same words, we assume we are communicating the same meaning. But the meaning in each case involves the individual's faith in their particular use of the symbol, which by its very collective nature cannot be unique.

Language allows not only for lies, but also for half-truths and even for empty statements. We enjoy nonsense, like Lewis Carroll's *Jabberwocky* poem, which works because we accept as potentially meaningful sentences that are grammatically well-formed but semantically empty. They seem meaningful, but key words don't refer to anything we recognize. Mathematics deliberately has this same structure, except that the "unknowns" ("*x*"s) are left completely open, to substitute what one likes, provided it is a number as formally defined. In society, it is politically useful to make empty statements, allowing the public to fill in whatever it imagines. Whatever the claim actually means to the person making it is unimportant, because each member of the audience will supply their particular understanding of the term. Indeed, this happens to some extent regardless of how clearly one attempts to express oneself. Yet, vagueness can be deliberately exploited.

Academic publications often read like jibberish to outsiders, not only because of specialized terminology but also because they sometimes do employ empty statements: grammatically correct claims for which even the qualified reader cannot be certain of the referents, given the ambiguity of terms. The problem can be far worse in casual language that makes no attempt at precision.

We are normally prone to believe our thoughts and feelings. Social media now run rampant with outrageous claims and memes, endorsed by our natural willingness, as social creatures, to believe what others tell us as well. Again, this reflects the power of language to evoke mental images and feelings, to which we tend to accord the same credibility we give to perceptual images and the feelings they arouse—that is, through acts of animal faith.

5 Reification, Abstraction, Scientific Modeling

The philosophy of mechanism continues to dominate scientific thought. It is based on the metaphor of nature as machine. Even in biology, cells are considered molecular *factories* or molecular *motors*, and in brain science mind is considered to be *software* running on the wetware of the brain. The computational metaphor assumes what it attempts to show, because it is itself a product of mechanistic thought. The practical success of this approach leads us to forget that it is metaphor—a way of thinking shaped by language—whereas no part of the natural world is literally a machine.[45] Computer animations and simulations are useful, but have their limits. Besides the possibility of quantitative error lies the mistake of taking the metaphor literally.

Even in abstract realms of speculation, we tend to have undue faith in our mental constructs. Faith in a theory may seem justified, as a provisional measure that can be updated by further observation, which is the basis of doing science. In the seventeenth and eighteenth

[45] See, for example, Daniel J. Nicholson "Is the cell *really* a machine?" Journal of Theoretical Biology 477 (2019) 108-126, p119: "'Molecular motors'… are not miniature versions of macroscopic motors. In fact, they differ from macroscopic motors in almost every important respect… For one thing, they lack rotors, armatures, and all the other trappings of conventional motors. They are made of soft, flexible materials which exhibit high degrees of freedom, unlike the hard levers, cranks, and hooks that make up most mechanical devices… A further difference is that molecular motors convert chemical energy directly into work without using heat or electrical energy as intermediates, which is why their efficiency is much higher than that of macroscopic motors."

centuries, the natural philosophers who became known as scientists believed in a substance called phlogiston, released as heat during combustion. This concept was superseded by the caloric theory, which conceived heat as a sort of fluid. That idea was abandoned in favor of heat as kinetic energy of molecules. In modern treatments, energy persists as a kind of substance interchangeable with mass (as per Einstein's famous formula). What is actually involved, in all cases, is *measurement* in specific contexts—measurable quantity rather than substance. But to reify energy as substantial seems to be a useful conceptual strategy. Even more derivative concepts—like field, entropy, and information—are reified as quasi-substantial, attributed their own causal powers. Even space and time have been reified—as in the 4-dimensional spacetime continuum, the extra dimensions of string theory, energy "landscape," phase-space, and other conceptual "spaces."

To objectify in this way is no doubt a built-in and useful tendency of the mind. After all, our primary orientation is toward objects in space and time. Since language and thought are essentially metaphorical, it is natural (if not logical) to think even of abstractions—indeed, anything that can be named—as vaguely thing-like. *Reification* and *real*, after all, come from the same Latin root, *res,* which means 'thing' or 'matter.' We tend to ontologize everything, more or less compulsively, even mind itself. (Then, like Descartes, we ponder what sort of "thing" mind must be compared to physical matter.) Underlying reification remains the fundamental need to believe in the reality of our mental constructs, which we also reasonably question.

Intelligence too is reified, divorced from its grounding in biology, which enables us to re-create it artificially, ironically in forms that may threaten our own existence and life generally. Indeed, life itself is being liberated from its natural manifestations, empowering us to create novel biological forms and to imagine even non-biological ways of living. We aim to play God through technology, using reason to achieve goals that have their origins outside reason.

6 Compulsion, Motivation, Affective Control

The brain is a delicate instrument, normally tuned to the needs of the body. Like a complicated machine, there is much that can go wrong with it. Also, being complex and plastic, it is capable of great variation, which can include behavior that deviates from what serves the body, the species, or the group. Underlying all variation, however dysfunctional, is animal faith. We naturally tend to believe what the brain tells us. Human freedom consists in the ability to be wrong while utterly convinced that we are right.

Having abstracted both intelligence and phenomenality away from their ties to the body, the very concept of experience—as a subjective category distinct from objective reality—introduces the pursuit of sensation for its own sake. It rationalizes it as something to consume regardless of consequence for the body or the world at large, so that experience becomes a sort of private entertainment, apart from its relevance to bodily needs. Addiction is one expression of the craving for experience. While addiction is an extreme example of subjectivity, it is hardly unique to humans; on the contrary, it is an animal phenomenon that takes us over despite our powers of reason and abstraction.[46] Indeed, "reasons" are often no more than rationalizations to justify compulsions. At worst, we simply succumb to the compulsion, believing its

[46] For example, the famous experiments with laboratory rats by James Olds and Peter Milner, in which the animals could opt to electrically stimulate "pleasure centers" in their own brains, often resulting in death by exhaustion, neglecting to eat, drink, or rest.

rationalizations. At best we are torn, and suffer for our divided nature. Reason does not, as Descartes believed, distinguish us from animals. On the contrary, it masks, distorts, and justifies basic drives we share with other creatures. It makes our animal faith far more insidious—making us dangerous to each other and even to ourselves.

Addiction is an example of the compulsive attractiveness of some stimuli (such as alcohol, drugs, or sex). It is natural to seek pleasure and try to avoid pain, because these represent real states of the organism, which tries to maintain itself and its kind. However, when experience is sought for its own sake (rather than the body's), the link with the body's wellbeing is broken. We can then find pleasure in things that are bad for the body or society, and reject things that are good. We can seek relief from pain (analgesics) without cure for the ailment.

There are other compulsions, such as obsessive behavior—including avoidance instead of attraction. And there are other questionable experiences we are tempted to believe, such as hallucinations stimulated with electrodes to the brain. Similarly, transcranial magnetic stimulation can change your perception, for example altering the apparent color of things or draining them of color altogether.[47] On the other hand, sensory deprivation also causes hallucination, as the brain makes up its own experience in the absence of sensory input. Depending on the circumstance, we believe or have reason not to believe the experience. If you know you have wires stuck in your head, you may justifiably be suspicious. In contrast, if you have unwittingly ingested a psychedelic drug, or are captive to a mood induced by chemical imbalance, as in clinical depression, it may affect your judgment as well as your perception, and you may fail to disbelieve your experience. It is helpful to keep in mind that the brain hallucinates *all of the time*, while some of the time its hallucinations are validated by genuine sensory input. We then call that "reality" and feel justified in believing the illusion.

Perhaps the paradigm example of the compelling aspect of phenomenality is pain. The first-personal subjective experience of *pain* should be distinguished from nociception, which is a third-person account of *sensitivity*. A reflex reaction to a harmful stimulus is to quickly withdraw or avoid the contact. This fast response should be distinguished from a slower response that results in the sensations known as painful, which serve a different purpose and involve different neural pathways. Pain compels us to protect the injured part during healing and so may persist for the duration. The earlier reflex response needed to be immediate, without any time-consuming consideration for dealing with injury over time. For this reason, it could be unconscious. The pain response, in contrast, involves the whole organism centrally coordinated over extended time. Since that is one job of consciousness, the pain is necessarily consciously felt. It stands as a guiding reminder to the self to favor the injury. The behavioral difference between these types of response can help assess the possible experience of pain in other organisms.[48]

Pain is a paradigm of how attention is compellingly captured by consciousness, because it usually involves immediate physical contact, requiring the utmost urgency. Much of our phenomenality instead involves the distance senses, and thus does not necessarily have this compelling aspect. The visual field is not a source of pain and may often be experienced with equanimity until a source of danger or opportunity is recognized. Then we may, for example, experience fear or excitement instead of pain, similarly keeping the organism on track with its pervasive reminder not to be distracted by stimuli with less priority. But priorities may change

---

[47] D. Hoffman *op cit,* p11.

[48] Sneddon, L. U., Elwood, R. W., Adamo, S. A., & Leach, M. C. (2014). Defining and assessing animal pain. Animal Behaviour, 97, 201-212.

over time, and hunger, for example, might override the fear. All such feelings are mobilizing in different ways, with implications for the whole organism. On the other hand, specific qualia of the distance senses, such as colors and shapes, or sounds as mere tones, may lack any sense of urgency. For, consciousness monitors the organism's relationship with its environment, of which many features may be of minimal interest at the moment. On the other hand, some colors stand out from their context—such as red fruit against green foliage—possibly indicating significance and stimulating a secondary response, hunger. Distance, providing a sense of safe detachment, gives vision its seeming objectivity, with the impression that the world as experienced is not a response of the body/brain, but simply is. Yet, even that judgment is an act of animal faith.

We live by believing our experience, which is a virtual reality produced by a squishy blob of tissue called a body. There is no way out of this theater hall, to see what's "really" going on. To some extent we can change the channel, so to speak. One day, it will go off the air. But for the time being, this *is* what's going on.

Animal faith mostly serves animal purposes well. We wouldn't exist otherwise. But having grasped our situation, and conceived possibilities beyond it, we are rightly not content with animal existence. We long tried to escape it with denial, but were always reminded by pain, sickness, hunger, sex, and death. Whatever might be the real possibilities of transformation remains to be seen. We can begin by recognizing the liabilities of animal faith, the ways in which it impedes our progress toward that transformation. It's tricky business, because that progress can only occur through setting beliefs against other beliefs, judged in terms of further beliefs. Yet, our very existence is proof of concept: the implicit unconscious commitments of animal faith can become explicitly known in consciousness.

[7] Embodied Cognition, Concept Formation

[7]

[3] "Just as the frog is born with its Darwinian legacy of bug-detectors, we also arrive with a 'prepared' repertoire of invariance-detectors that pick out certain salient shapes or sounds from the otherwise blooming, buzzing confusion reaching our sensorimotor surfaces without our first having to learn them by trial and error…" [Harnad 2002, ibid]

[7]

[3] "Imagining and doing use a shared neural substrate… The same neural substrate used in imagining is used in understanding… imagining is a form of simulation—a mental simulation of action or perception, using many of the same neurons as actually acting or perceiving…" [Vittorio Gallese and George Lakoff "THE BRAIN'S CONCEPTS:THE ROLE OF THE SENSORY-MOTOR SYSTEM IN CONCEPTUAL KNOWLEDGE" COGNITIVE NEUROPSYCHOLOGY, 2005, 22 (3/4), 455–479, p456]

[7]

[3] "The 'learning' of general cases is not the acquisition of new structures, but rather the inhibition of the connections between secondary and primary areas. In other words, the generalisations are inherent in the special cases that are learned first. What is learned is the control of inhibitory connections." [Gallese & Lakoff, p471] cf sensory referents

[7][8]

[3] Invariants in the objective world and the needs of the organism both shape cognition.

However, there is a problem of cog dom, since the invariants and needs that condition cognition are a product of that (human) cognition.

[8] Epistemology, Cognitive Limits

[8]

[3][12] There is an objective ("inhuman") reality, which Kant called noumenon. Our human perceptual and conceptual representations (phenomena) derive from that through natural selection (according to one such conceptual representation). Through self-consciousness and imagination, we can conceive counterfactual possibilities, such as ideals of objectivity and transcendence of biologically imposed limits, whether physical or mental. At the same time, these ideals are conceived by our present biological organism and conditioned by it. Even in the case of mathematics, the categories we have conceived reflect both our biological being and invariants in the objective world. That would seem to apply to any possible embodied cognitive system. We can imagine a cognitive system that is not "normatively structured" like the human one. If it was singular (not part of a discursive community), it does not seem there would be any need to formalize empirical knowledge in some version of mathematics. Even familiar animals (such as some birds) can "count" but presumably have not imagined an arithmetic. GPT's version of above:

There is an objective, inhuman reality (Kant's noumenon).

Human cognition evolved to track aspects of that reality via perceptual/conceptual representations (phenomena).

Through reflexivity and imagination, we generate counterfactual ideals (objectivity, transcendence, mathematics).

Mathematical categories reflect both biological constraints and objective invariants in the world.

Mathematics reflects invariants that constrain any embodied cognitive system; but the individual embodiment—and embodiment generally—also impose constraints.

Any embodied cognitive system would exhibit a similar dual conditioning.

A solitary, non-discursive intelligence might not need formalized mathematics at all.

[8]

[3][8] To distinguish invariant structure in reality, from the particular transform by which a system accesses that structure, invokes the prob of cog dom. For, the supposed invariant structure in reality can only be known through the particular transforms that access it. An invariant is not identical to its representation. For example, rotational symmetry in space exists independently of Euclidean geometry, tensor calculus, or group theory, which are specific transforms for accessing it. [GPT]

[9] AI, Simulation, Artificial Systems

[9]

[3] "If we trained an LLM solely on astrological data, it would produce astrologically plausible answers." [Floridi et al, p8]

[9]

[10] [3] It is through language that we humans convince one another that we are conscious, that we feel and have phenomenal experience. Since LLMs are by definition masters at language, they could be similarly convincing, simply by saying the right things. A better guide than any version of the Turing Test would be to know, on grounds independent of language, that the communicating system is an autopoietic agent.

[10] Methodological Abstraction in Science
[10]
[3][10] "To investigate localized areas within the organism as machines allows biologists to conveniently abstract away the intimidating complexity of the broader physiological context of the organism as a whole, and focus their attention on well-defined interacting parts." [O≠M, 675] cf isolated system
[10]
[3] Laws of physics are simple because (until recently) only simple mathematical equations could be solved.

Chapter Four: The Sceptical Role of Consciousness

> "Nature, silently making fools of us all our lives, never would bring us to our senses; but the maddest assertions of the mind may do so, when they challenge one another."—Santayana

To grasp the biological functions of perception takes our view of reality out of the context of truth and puts it in the context of our lives as organisms. This was a hard-won modern realization, tenuous and scarcely to be taken for granted. For, the natural focus of mind—in the service of survival—remains outward, toward the world external to the dependent organism. We are thus largely obliged to perceive the world as real and external, as it is given in our experience, and to trust our perceptions, feelings, and ideas. This natural circumstance unwittingly entraps us within particular ways of seeing and feeling, which seem natural and self-evident while we are in their spell. Simply knowing this does not alter our basic wiring nor prevent the brain's natural projection of its simulations as reality. But it does reflect other wiring that can partially compensate. An agent self-conscious in this sense can affect not only its external environment but to some extent its internal one as well, in ways that better its chances of survival. Reflexive consciousness is adaptive. It would have to be, since the extra neurology involved would be evolutionarily costly.

Self-consciousness is the capacity to be self-aware. To be conscious *that* one is conscious adds a critical layer to the normal reality-creating ability of the mind, which Santayana characterized as a "vital compulsion to posit and to believe." The scepticism he proposes is a perk of our reflexivity. When we're self-aware, we can reflect on how our perceptions and experiences do not fully capture objective reality and how they can be mistaken. Understanding that our minds actively construct the world we perceive, we can question or deconstruct our experience and recognize mistakes. Animal faith takes experience at face value. But, knowing that perception is mediated by mental processes, happening inside the skull, allows us to see the world as a projection of internal processes rather than a literal view of the external world. Since perception can be wrong, this ability to question experience helps us to look (again) before leaping, to notice our assumptions and revise our opinions, to question our feelings and reactions. Self-consciousness can thus compensate for the liabilities of animal faith.

Self-consciousness is a form of metacognition, enabling a system to transcend its own limitations or logical level. Rather than viewing the world through a lens, so to speak, we are able to step up a level to examine the lens itself as an object. This allows us to see our own limits, from a broader perspective, at least in the abstract if not always in the moment. The challenge in real time can be personally humbling, for we naturally enjoy the self-confidence of having faith in our view of things. But it is this very seductiveness that demands a countermeasure. To recall to mind our fundamental situation—as embodied creatures possessed by animal faith—reminds us that we can be wrong in a given situation, which could be viewed differently through other eyes. It reminds us that the ideal of objectivity is a double-edged sword. It can cut through illusion and self-deception, but can also amplify these when we take for granted that our point of view *is* the objective truth. Even a meta point of view is nevertheless a limited point of view. Any system that can model itself must involve uncertainty about its own accuracy at any level. Trying to eliminate this uncertainty by adding further layers has diminishing returns, since each layer in turn can be questioned. The subjective price we pay for self-consciousness is doubt.

The sceptical tradition in philosophy is ancient, going back to the Greek and Vedic sages. Kant furthered Plato's intuition of the Cave, instructing us that we can access only our own perceptions and thoughts, not reality itself. Descartes showed how input from the senses could be faked, an insight leading in modern times to the brain-in-a-vat scenario, the Matrix films, and the paranoia that "you are probably living in a simulation."[49] Yet, understanding that experience is a product of the brain is not in itself cause for scepticism. It simply enables us to grasp that sometimes this real-time illusion serves us better and sometimes worse. It is the *persuasiveness* of the illusion that Santayana earmarks as animal faith because of its natural function.

Total skepticism would be paralyzing. A mind that doubts everything can't commit to action. On the other hand, a mind that believes too readily becomes a slave to its fictions. In order to function in the world, any mind must be willing to believe its working model at least provisionally. Yet, it serves its interests to also be able to suspend that belief when appropriate. Another term for scepticism, then, is controlled doubt.

Objectivity is desirable, of course—a worthy ideal. Unfortunately, the natural tendency is to mistake our actual experience for objective reality. In order to maintain this illusion, we tend to avoid reference to our own subjectivity, asserting our perceptions and conceptions as truth, while perhaps accusing others of being merely subjective. While the ideal of objectivity is to transcend the merely subjective, this can only be accomplished by acknowledging one's own subjective basis for cognition and claiming responsibility for it. Paradoxically, we must own our subjectivity in order to become more objective. The naïve presumption of objectivity is that there can be a gods-eye view from nowhere, and that the individual subject can occupy this non-existent place. But, literally and figuratively, all views are perspectives from somewhere, from the point of view of some literal body. Natural realism and social conditioning trade on denying this fact and ignoring the responsibility it implies. Thus, the personal challenge is often to question or bracket the apparent obviousness of one's own point of view.

In the story of the blind men with the elephant, each has incomplete information about the "elephant in the room." Through cooperation, they have the possibility to pool their knowledge and arrive at a more complete representation of the elephant. But this requires first that they realize their situation (of incomplete information); and that they hold objectivity as a collective ideal that requires coordinated effort. To cooperate in this way, they must relinquish their tribal or parochial commitments—in other words, their animal faith.

The notion of objective reality is by definition singular. But perception is by nature plural. The natural assumption that one perceives the world correctly (that is, transparently) rests on the false assumption that thinking requires a single center, properly located within the individual, and is only trustworthy thus. This is no more than an expression of animal faith, ironic because reason is actually collective.[50] It appears to be individual to the degree that survival is individual: one naturally trusts one's own senses and reason, since one is alone mandated, and best situated, to look after one's own interests. Yet, human beings are highly social and our achievements—including the ability to think—are necessarily collective. Our

---

[49] The so-called Simulation Argument proposes that posthuman societies (or other advanced civilizations in the universe) could and would produce simulations whose inhabitants could be conscious. Since these simulated realities would greatly outnumber "real" reality, if you are conscious then you are mostly likely living in one of these simulations. The irony is that we are *definitely* living in the brain's natural simulation!

[50] Daedo Jun "AI Collective Reason: The Birth of Collective Intelligence. Autonomous Evolution Series, Part X, When Does Intelligence Begin to Think Beyond the Individual?" (version 1.0 draft), p6.

survival as a species is no longer a matter of individual selection, but will depend on our collective efforts.

The senses seem to reveal the external world as it truly is, thereby justifying animal faith. Behavioral choices have real consequences and we naturally seek to justify them (to others and even to ourselves) in terms of real externals—which means in terms of causal processes in the world. On the one hand, we recognize that the world has power over us; on the other, we want to initiate our own actions and thoughts, which are often in reference to external reality. Meaning is naturally imposed on us by reality, upon which we depend as natural organisms. But as beings who wish to claim free will, we are ambivalent toward that dependency, through which meaning seems to abide ready-made in the world, just as the world seems transparently revealed to us in vision. Yet, knowing that meaning is not intrinsic in the world, but like all perception is a function also of biological need, makes the world unreliable as the source of meaning and throws us awkwardly back upon ourselves. This is the existentialist dilemma: we never know for certain what naturally comes from the subject and what from the object.

This is one reason for artifact: while the materials necessarily come from nature, we know that the form comes from us, as our creation. This is as true in the realm of thought as in technology and the arts. Nature provides the "material" of scientific laws, for example, while human thought provides their form. Even mathematics generalizes experience with the material world, formalized at the most abstract level.

Above all, meaning for us has become a function of language, of claims we make in verbal form. Statements may be *about* the natural world but they are human *acts*. A text, like a machine, is literally an artifact. Unlike speech, which evaporates in the air as soon as it is heard, a text is its own memory. It can be searched backward or forward, edited, taken out of context, manipulated—in short, *studied.* It has an existence outside time, detached from the original embodied context of speech, and is therefore especially subject to misunderstanding. Live speech supplies real-time cues such as intonation, pauses, emphases, facial and hand gestures, and the original context that prompts it, all of which inform the transmission of meaning. Through these elements, speech also conveys emotion. Lacking them, text relies on associations that might supply emotion, as readers fill in their own meanings. Text has the advantage of detachment—or the disadvantage, depending on the circumstance.

Consciousness provides an ongoing narrative about the world, which ought to be coherent and provide a basis for decisive action. In other words, the narrative should be meaningful. The problem of nihilism arises when conventional sources of meaning and motivation wither or are overcome by doubt. By nature, we look to externals to justify what we value. That is, we look to a worth inhering in those things themselves, or some reason in the world why they should be valued. This habit stems from the natural outward orientation of the brain, which focusses on the environment as the source of the organism's wellbeing. This orientation is a basic fact of being a creature with a nervous system, dependent on an environment for its life. While we may protest this dependency, there is security in a dependent relationship, as the child finds security in the its parents. The loss of this dependency itself is threatening. It is a loss of animal faith, for which one feels bereft, "orphaned," on one's own to face the Void. One can no longer count on the inherited patterns of belief we call meaning or reality.

Nietzsche warned that nihilism could lead to personal despair, anxiety, apathy, and a passive or destructive culture. There can be anger at the loss of meaning, as of any resource.

Disillusionment or disenchantment is a loss of faith in something once deemed real or true, and therefore involves a loss of certainty. The normal outward-facing mind can no longer count on finding justification "out there" for its beliefs and actions. Doubting *particular* beliefs or assumptions can be functional; scepticism can lead to a better understanding of reality and increased self-confidence. But doubting the reliability of the mind or the validity of our experience in general can be undermining and overwhelming.

Descartes' scepticism concerned the input of the senses, which could be falsified through tampering with the nervous system. His solution was that God would not permit systematic deception. If we substitute nature for God, we could suppose that *natural selection* would not permit the kind of error that precludes our existence. (On the other hand, some illusions may *promote* existence—at least reproduction!) While Descartes doubted the senses, he took false comfort in his *cogito ergo sum,* by concluding that one could hardly doubt one's own existence as an experiencing subject.

Nietzsche's scepticism did not concern the veracity of experience but its meaning. Even granted reliable sensory input, there is no absolute basis on which to interpret it, no intrinsically reliable source of meaning or infallible procedure for truth. There is no absolute reference frame —represented by "God," who is dead. Nietzsche's solution was to re-evaluate valuation itself: to relinquish reliance on the external world as the source of one's beliefs and actions—whether determined by external causes or by internal reasoning. He took comfort in *amor fati*—the practice of intentionally embracing all experience without evaluation. If meaning cannot be counted on from outside, one must create it oneself! Even negative experience should be welcomed as an opportunity to be intentional and self-determining, rather than reactive. In the absence of intrinsic meaning, one must renounce looking for it outside oneself. But that requires claiming utter responsibility for all one thinks and feels, even for sensory experience. Ironically, that is a position only the Creator can occupy, as opposed to the finite creature whose lot is to respond as best as it can to the intricacies of the Creation.[51]

Meaning is a high-level framework for evaluating experience. Its loss is naturally evaluated negatively. Such a judgment reflects the continuing general reliance on externals, part of our natural conditioning. Despair or depression is a normal response to a threatened loss of meaning. One can defend against such threat by joining a group or cause, embracing an ideology or reaffirming a faith, returning to traditional values, sticking to routines, losing oneself in work, etc. One can also lose oneself in distraction, entertainment, or release through drugs. Nietzsche calls such defenses passive nihilism, because the potential loss of meaning is acknowledged but is responds to it as one would to an external threat—reflecting the continuing belief in reality as the causal factor. While individuals might choose to actively confront the absence of intrinsic values, society as a whole cannot be expected to. Yet, if it does not, Nietzsche warned, it may fall into religious fundamentalism, rigid nationalism, populist movements, or totalitarian systems, as substitutes for lost meaning—all of which we have seen.

Choice is ultimately arbitrary when it cannot be justified by externals. On the other hand, justification can be sought internally, if only through self-interest. In that context, *choosing* to

---

[51] Nietzsche's *Übermensch* is a godlike being, who asserts free will against the determinism of the natural world. He proposes a strenuous and heroic ideal, whose demands may have been too much even for Nietzsche himself, who collapsed at age 44 and spent the rest of his short life being cared for by family or in an institution. One could read this as a concrete illustration of personal hubris, or more generally of the human confrontation with the Void. To live continuously in tension with it, creating new value *ex nihilo,* requires enormous courage and vitality. Nietzsche's life shows both the possibility of a life-affirming philosophy and the cost of pushing a finite organism beyond its limits.

face the meaningless of existence mirrors the organism's fundamental self-defining nature—its autopoiesis. This affirms the self's responsibility for itself and its actions, mirroring the organism's relative autonomy and freedom. From an existential point of view, values are not found but made. To confront this fact is to confront our own nature. For, on the one hand, we are conditioned to seek reasons (justifications) for our values, beliefs and actions. But, on the other, we seek freedom from such conditioning. Paradoxically, we may seek justification even for faith in existentialism and for the option to pro-actively create value. But if there is no reason for *any* choice, there is no reason to choose existentialism either. This does not prevent us from choosing our values or conduct, only from expecting a justification: one should not expect support from the Universe for one's choices! Nietzsche viewed life as a work of art, rather than a science. Choice of theories in science should depend on experimental evidence; artistic choice is simply up to the artist. On the other hand, like knowledge, value can be a consciously shared collective creation. Then it would no longer be something to fight over, but to create together.

Meaning is about context. A cell in a multi-celled creature is an organism in its own right, but should not have the freedom to abandon its role. It derives its meaning from its place in the organism as a whole, which exists on a different level. That's the model provided by biology. The cell has the choice to go rogue, and humans have the choice to disregard their social context, or their species-level context. Unlike the cell, we also have choice to create new levels of belonging.

As a form of cognition, science focuses on a world it presumes to exist independently of itself. In that, it extends the naïve realism of ordinary perception.[52] Yet, it is haunted by the same ambiguity that troubles human consciousness generally: the doubtful relationship between appearance and reality. Science aims for an objective description. Yet, even scientific description is necessarily from the point of view of embodied observers—even though it is a public act of communication accessible by others.[53] All observers stand in a first-person epistemic relation to the world, whether through their natural sensory-motor instrumentation or via external devices that extend human agency.

Science operates through a conscious process of inference that mirrors the brain's natural processes of unconscious inference.[54] The epistemic situation of the scientist is like the brain's natural situation, and subject to parallel limitations. Scientific protocol is designed to liberate scientific findings from the idiosyncrasies of individual scientists. Experiments are to be repeatable by other, standardized observers. Yet, because of its very focus on the observed world

---

[52] Naïve realism is a cost-effective way to maximize genetic fitness. To eliminate "transparency" would increase the computational burden on a system and possibly hinder its efficiency. [Thomas Metzinger "Artificial Suffering: An Argument for a Global Moratorium on Synthetic Phenomenology" (Open access) Journal of Artificial Intelligence and Consciousness Vol. 8, No. 1 (2021) 124, p13]. This applies to scientific cognition as well; to formally include the role of the observer in scientific accounts would introduce computationally costly (perhaps indefinite) recursion.

[53] Scientific claims are made by real people, in language that is not just an expression of facts but constitutes "a game of expressions and commitments, whose primary use is the coordination of action between humans." [Harry Halpin "Artificial intelligence versus collective intelligence" AI & SOCIETY, 2025, (Sec 4.2) https://doi.org/10.1007/s00146-025-02240-x]. To the extent that scientific activity can be formalized in linguistic and mathematical propositions, it likely can be simulated by a machine. However, if science is considered the activity of a natural organism, it may not be fully capturable.

[54] Helmholtz's idea of unconscious inference is a metaphor modeled after *conscious* inference, applied to the brain's unconscious modeling processes. Similarly, we now speak of the brain's "information processing" or "computation," which are metaphors derived from consciously created digital technology. Language is inherently metaphorical.

(as opposed to the observer), science is ill-equipped to question its own assumptions and biases. It may transcend cultural particulars, but can it be free from assumptions and biases common to human observers generally? Will AI help us identify these biases and transcend them, or further reinforce them?

Mechanisms can now be so complex and sophisticated that they appear lifelike. Simulations can be so realistic that we cannot tell them from reality. We can no longer rely on ordinary cognition to conclusively judge the difference between nature and artifice, especially when there is an intention to obscure the difference, as with large language models and computer graphics. Behind that intention is the drive to create artificial general intelligence, reflecting the ancient dream to re-create reality in our own image, to convert the found world into a world of our own making. As we become ever more immersed in simulation, will we be able even to recognize natural reality?

Scepticism is especially warranted in a world where information comes to us digitally mediated and no longer firsthand but *nth*-hand: no longer face to face but Facebook to Facebook, or regurgitated by a chatbot. The power of digital media to deliberately fool the senses at last fulfills Descartes' paranoid intuition. He claimed that God would not permit systematic deception. Until recently, we could trust that *nature* would not permit it. Now, having abandoned God, distrusting human authority and appropriating the authority of nature, on what grounds can we know what is real?

We are animals who can conceive being superhuman. This ability reflects a natural trend, that as organisms become more complex, the scope of their agency increases. However, it also reflects our singular ability to recognize the limits imposed by physical reality and by our biological nature, enabling us to imagine freedom from those constraints and motivating us to overcome them. We can imagine the general, the abstract, and the ideal, which includes notions of moral perfection. Scepticism can be applied to the human condition at large, not just to information. We may realize with some horror that what applies morally to others must apply to ourselves:

> "The spectacle of other men's folly continually reawakens in me the suspicion that I too am surely fooled; and the character of the beliefs which force themselves upon me — the fantasticality of space and time, the grotesque medley of nature, the cruel mockery called religion, the sorry history and absurd passions of mankind — all invite me to disown them…"[55]

For such realization, we have not only self-awareness but also conscience. Regret is a personal form of self-indictment, after the fact, realizing that one did not act well enough in a situation or missed some opportunity. Often such failure can be traced to a wrong assumption or a biased perception, some automatism that would not have occurred had one been more conscientious, or less judgmental, or less caught up in one's own state. Sins against others accumulate over time, magnifying guilt.[56] We can avoid guilt by making the right choices in the first place. That requires some systematic willingness to question our motives, perceptions, ideas and feelings—*before* acting rather than after.

---

[55] Santayana, op cit, p.21.

[56] Perhaps one of the motives for believing in a benevolent personal God is the hope for a consolidated blanket forgiveness, like the cancellation of debts.

Chapter Five: Epistemic Cycles of Knowledge

> "Dogma cannot be abandoned; it can only be revised in view of some more elementary dogma which it has not yet occurred to the sceptic to doubt."—Santayana

It cannot be taken for granted that embodied mind seeks truth. The goal of life is survival and reproduction, not objectivity. As Santayana put it, in his style characteristic of the time:

> "The ideas we have of things are not fair portraits; they are political caricatures made in the human interest… It matters little if their very existence is vouched for only by animal faith and presumption, so long as this faith posits existence where existence is, and… preadaptation of animal instincts to the forces of the environment. The function of perception and natural science is, not to flatter the sense of omniscience in an absolute mind, but to dignify animal life by harmonising it, in action and in thought, with its conditions."[57]

In other words, our natural condition as organisms is to see and know what we need to see and know. And this is more than simply a matter of selective attention or reduced information flow—a removable filter between the mind and an otherwise transparent window on the external world. In more modern terms, a perceptual system is a *user interface*, like the desktop of a computer.[58] What governs its use is not truth but evolutionary fitness. While this insight must apply even to science as a cognitive system, the very possibility of scepticism and curiosity renders knowledge open-ended. It implies a view that can be better, if not absolute.

The demand for proof and for formal decision procedures, as opposed to blind acceptance on faith, formalizes the scepticism and relativization furnished in ordinary self-consciousness. On the other hand, the mind's ability to step beyond any defined bound paradoxically implies a transcendent ideal of truth, beyond any given formalization. In trying to picture the unpicturable face of the world-in-itself, we are tempted to mistake it for some appearance in the mind. The concept of the noumenon is but a name for this unpicturable realm. While it is useful to think that mathematics corresponds to that realm, by definition this is unverifiable.[59] Similarly, consciousness may be out of bounds to explain scientifically, because science excludes subjectivity from its own method while it is a form of subjectivity. While the mind asserts higher conceptions, the sceptical role of subjectivity questions them, undermining absolutism, in a potentially endless game of leapfrog. This is the context for any notion of objectivity.

Knowledge is a process and certainty is conditional. There is no guarantee that our cognitive powers can solve every problem we can recognize or formulate.[60] The process involves

---

[57] Santayana, op cit, p.104.

[58] Hoffman, op cit, p.75. The metaphor does not quite work, since the desktop represents functions within the computer, not in the real world. Hoffman pits the goal of truth against the goal of fitness as competing strategies in repeated simulations, with "fitness" consistently driving "truth" to extinction. [p55] But, what does it mean to consider fitness itself a *goal*, apart from maximizing reproduction? Moreover, natural selection in real life does not *preclude* truth; it only fails to require it.

[59] While Gødel formally proved some limits of formalization, he seemed personally to believe in a Platonic realm of mathematical existence. Despite his concern with mathematical consistency, his metaphysics hardly seems consistent.

[60] Colin McGinn "Can We Solve the Mind-Body Problem?" *Mind* New series, vol 98, no. 391 (July 1989), p353.

a dialectical cycle: thesis, antithesis, and then synthesis, where apparent contradictions are resolved in a higher unity. This last phase serves as a new "thesis," beginning a new cycle. We see this in formal knowledge processes, like scientific theory-making, where an idea is proposed (thesis) to explain new data or to make up for a deficiency in current theory. This idea is published—in a journal, for example—inviting comment and critique (antithesis), which may suggest further testing by experiment. If the idea is accepted by the scientific community and not disqualified by experiment, the resulting change is a synthesis that becomes a new thesis to eventually be challenged. Experiment is never totally decisive, but always involves probability and a judgment call. New theories are rarely accepted universally, even over time.

Unlike the individual brain, isolated in the skull, science is a collective enterprise, a communication among scientists, a (mostly) polite form of argumentation through which ideas are justified to others and trust is established.[61] This requires persistent interactions as a sort of collective brainstorming, in contrast to individual thought, which can be episodic.[62] In fact, science is a model of social cooperation, transcending political and cultural boundaries. Just as there is an epistemic cycle of individual knowledge, there are collective cycles in science too. These include paradigm shifts, but also alternations of more general fashions, such as positivism and deductionism.[63]

The epistemic circumstance of the scientist parallels that of the brain, which relies on the input of "remote" receptors to infer the properties of the external world. Just so, the scientist relies on instrument readings. Both situations demand radical inference. The brain's unconscious perceptual models are reliable to the degree they enable survival at the individual, group, or species level. By the same token, scientific modelling, like other human practices, should not be regarded only for its truth value, but also for its ultimate contribution to planetary well-being. Good science should support a human future.

In science, the interceding effect of the observer (which includes the instruments of observation) is minimized, so that the signal to noise ratio is maximized. The observer's subjectivity is suppressed in order to better focus on the object of investigation. Protocols must be standardized; observers must be interchangeable. In order to better explore the external world in ways useful to human purposes, the purposes themselves remain unspoken. The observer stands ideally outside the system observed, a fly on the wall.

Science is intrinsically idealizing. Mathematics (the essence of idealization) dominates science because physical phenomena are redefined in such a way that they can be treated effectively with math: as idealized systems. This aspect of science leads to an analysis in terms of the ideal parts of a conceptual machine. The functioning of the whole is to be understood the operation of these parts according to ideal principles. Thus, mechanism, reductionism, and mathematical laws are essentially by-products of idealization. The natural system is treated as a mathematical construct. Since determinism is a built-in feature of such constructs, real systems

---

[61] Mercier & Sperber, op cit, p8: "By giving reasons in order to explain and justify themselves, people indicate what motivates and, in their eyes, justifies their ideas and their actions. In so doing, they let others know what to expect of them and implicitly indicate what they expect of others. Evaluating the reasons of others is uniquely relevant in deciding whom to trust and how to achieve coordination."

[62] Daedo Jun, op cit, p9-14.

[63] Deductionism is the idea that nature can be exhaustively described by formal axiomatic systems, mathematical models. String theory is a paradigm example of the deductionist approach. A classic example of opposing stances was the debate between Einstein (the deductionist) and Bohr (the positivist) over the interpretation and completeness of quantum theory.

are often treated as deterministic, when in fact it is the *equations* used to describe them that are deterministic.

Like scientific cognition, ordinary cognition involves probabilities of error. But the individual brain tends to be more definite in its conclusions than the scientific community. The organism must be able to act decisively on the basis of the information it has, however inadequate. It must at least be able to distinguish between signals originating internally from those originating in the environment. It must know where and how to direct its energy. A primary function of animal faith is to ensure this capacity to act, even inappropriately. If, despite real uncertainty, our perceptions were not definite, we would be paralyzed by doubt. Yet, because of the possibility of error, the knowledge cycle would be incomplete and less reliable if animal faith alone were in play (the thesis stage). There must be a balance between hesitation and decisiveness, an interaction between scepticism and animal faith,.

The inherent need to trust our perceptions and beliefs is problematic when we come up against the contradictory perceptions and beliefs of others. While objectivity is an ideal, the natural tendency is to mistake one's perception for objective reality. And in order to maintain this illusion, we tend to overlook inconsistencies or omissions in our own thinking and to protest that we are being objective while others are not. While there can be dissonance within one's own thinking, leading to self-scepticism, dissonance with others is nearly guaranteed. Too often, this leads not to questioning and enlarging one's own views, but to retrenchment and wariness of anyone who disagrees. Nevertheless, the fact that opinions differ plays a positive role in the epistemic cycle. We tend to think of reason as a private tool to consciously evaluate information and find a course of action. But its origin may lie rather in the ability to convince others and be convinced by them; its main use even now may be to evaluate the arguments of others more than to evaluate one's own.[64] In terms of a thinking process, it may be more efficient to rely on being challenged by others than to challenge one's own thinking.

Whether spontaneous or forced by others, the recognition of one's own error or subjective limits enables mind to evolve at once both toward humbled relativity and greater objectivity. It is no paradox that scepticism itself involves belief, when we understand that these two apparently opposing movements are but facets of a dialectical cycle. The realizing tendency of mind posits an idea, schema, or model, which is in effect a theory about reality. Properly, this should be checked against sensory input for fit, but the idea can simply be believed instead. The sceptical, de-realizing factor of subjective consciousness makes the mind accountable to itself and to others, by insisting upon justification for belief.

The interplay of positing and negating aspects of consciousness—of animal faith and scepticism —manifests in historical cycles, the opposing phases of which in culture may be characterized broadly as *heroic* and *ironic*. These poles form a unity, like those of a magnet, alternating as undercurrents which surface in philosophical, social, political, religious, moral, artistic, and even scientific movements and fashions. The limiting nature of any proposition or system of thought casts a shadow that is the other side of the coin. Every thesis defines its own complementary antithesis. Where contradictions cannot be resolved *logically*—that is, outside time—they give rise to *temporal* alternations in a cycle. The pendulum of history swings back, fashions return; we move in spirals if not circles.

---

[64] Hugo Mercier and Dan Sperber *The Enigma of Reason* Harvard UP, 2017, p7ff.

Throughout history, there has been a dialectical relationship between the playful, embroidering, subjective, ironic side of the human spirit and the serious, goal-oriented, earnest, realist, heroic side. The ironic mentality embraces limits and delights in playing within bounds. It understands limits to be gratuitous, relative, intentional. The heroic mentality rejects limits as obstructions to absolute truth and personal freedom, while worshipping limitlessness as a transcendent ideal. The heroic is aspiring, straightforward, straightlaced, linear, square, passionately simplistic, naive, concerned with content over form, and tending toward fascism and militarism in its drive toward monumental ideals and monolithic conceptions. The ironic is witty, sarcastic, curvaceous, ornate, sophisticated, hip, diverse, sceptical, self-indulgent and self-referential, tending toward decadent aimlessness and empty formalism. Each is hazardous as an extreme. Together, they are the creative engine of history.

There are cycles of opening and closing in societies, in individual lives, and in creative processes generally. The tension between idealism and materialism, or between heroic and ironic frames of mind, helps to explain why history appears to stutter. Most of any historical cycle will consist of working out the details of a new regime, scheme, paradigm, or theory.[65] But the cycle will also necessarily include an initial creative ferment and a final stagnation, sandwiching the more conventional middle. When change is too rapid or chaotic, there is nostalgia for the perhaps not-so-good old days. Instability inspires conservative longing for structure, certainty and order —until an excess of *those* inspires revolt again. Generally, too much of anything breeds its opposite, an aspect of the homeostatic search for balance.

Paths of personal or collective learning follow a similar dialectic. Ignorance gives way to mastery, as we recognize what was formally automatic and learn to guide it consciously. But mastery produces alienation, as the seat of control differentiates itself from what was controlled. Finally, alienation can be reconciled as we rediscover how to be an integral part of the system.

Our fundamental epistemic situation is circular. If space and time themselves are products of the brain, how can the brain be located in the space and time it has created? The external world appears to subjective consciousness as an image constructed by the brain, which is part of the world so constructed as an image. The endpoint of a process is recycled as its beginning.[66] While science aspires to describe the world objectively, and even definitively, it is another of the brain's constructions, whose fundamental purpose is advantage. The ideal of truth can only be approached through an unending process of knowledge making.

---

[65] See Thomas Kuhn, *The Structure of Scientific Revolutions.* U. of Chicago Press, 1962. He coined the term 'paradigm shift', describing the dialectical creative cycle of science through a political metaphor.

[66] A dilemma I call the problem of cognitive domains: the circularity that arises when the output of a cognitive process is recycled as its input.

Part Two: The Pilgrim's Progress

Chapter Six: A Brief History of Breakthroughs

> "We are beings with goals our parts cannot conceive of individually."—Michael Levin

The emergence of mind cannot be separated from the origin of life, for which the larger context is the origin of the solar system, the galaxy, and the cosmos itself. That fascinating story is a grand accomplishment of both the physical and biological sciences. We concentrate here on certain aspects that bear on the conditions for knowledge. Modern theory holds that, sometime after the Big Bang, the universe cooled enough that electrons could be bound to nuclei, permitting the capture and re-emission of photons.[67] Before that, the universe would have effectively been opaque to the transmission of light. From an epistemic point of view, this was a landmark that made vision possible.[68] We could backtrack further to imagine the selection of our universe from a multiverse of hypothetical possibilities, with its unique physical constants favorable to the eventual arising of life and mind.

Epistemic landmarks within the history of life on this planet begin with the transition to life from complex abiotic chemistry. Identifying and characterizing that transition depends on how life is to be defined, which might include the degree of complexity, and other criteria such as autopoiesis and reproduction. It might involve an early shift to hierarchical downward causation in self-replicators; a definition of true life may require storage of information in a dedicated structure.[69] The transition from abiotic chemical processes to life involves the key emergence of a separate symbolic code to store information guiding replication. The leap from physical causation to symbolic coding is what makes the difference between mere chemistry and life. Certainly, this was a precondition for the development of nervous systems and what we identify as mind.

The question is of interest in astrobiology, which looks for chemical signatures of life in the atmospheres of exoplanets. While it is assumed that physics is the same everywhere in the universe, the chemistry that can result represents a vast combinatorial space. Organic chemistry elsewhere may not be like what we know on Earth. To identify alien life requires a definition of life that is not tied to familiar forms. While the life we know is characterized by metabolism and replication, it can be also viewed as matter that stores and accumulates information in ways that result in novel objects with causal histories. While life is itself a complex molecular process, it

---

[67] Recombination (decoupling), is said to take place "378,000 years" after the Big Bang. Literature concerning the early universe often does not discuss the meaning of "time" before the existence of measurable cyclical processes, but generally assumes that concepts formed locally in the present era should apply universally throughout time and space.

[68] Given that electromagnetic radiation only seems to involve a small fraction of the actual universe, we might anticipate a hypothetical discovery of some form of supra-luminal communication, possibly related to dark matter and dark energy. It is possible that the velocity of light does not represent an absolute speed limit, if an unknown swifter signal happens to exist. See my paper "Why Is *c* a Cosmic Speed Limit?" [https://stanceofunknowing.com/wp-content/uploads/why_is_c_a_cosmic_speed_limit.pdf]).

[69] Walker SI, Davies PCW. 2013 "The algorithmic origins of life." J R Soc Interface 10: 20120869. http://dx.doi.org/10.1098/rsif.2012.0869.

also *produces* complex molecules. Those that could arise in no other way would potentially serve as a bio-signature, if detected in the atmospheres of exoplanets.[70]

The next major transition was from prokaryotes to eukaryotes, and then from single cells to multi-cellularity.[71] One of the first innovations leading to brains was life's invention of bilateral symmetry among multi-celled creatures.[72] This facilitated *steering* in pursuit of food or avoidance of threat, which both required and enabled value-laden sensitivity. Following the Cambrian explosion, arthropods were the first creatures to leave the highly competitive environment of the sea for dry land.[73] They were well equipped to do so, with an armoured body, strong legs, and acute vision that could take advantage of air as a better medium than water for transmission of light, which enhanced distance perception. The ability to communicate information deliberately to other creatures, principally through sound, is another pivotal achievement. So is the advent of internal representation, which is computationally expensive, and the warm-bloodedness that facilitated it by enabling nerves to function in the cold. Perhaps a future landmark in the evolution of mind will be to transcend the limitations of biological embodiment—possibly supplanted by artificial intelligence. Just as warm blood enabled faster nervous systems, so digital brains would be far faster than biological ones. In any case, since mind is necessarily embodied, its development follows the changing needs and environmental circumstance of particular body designs.

Brains developed as a feature of organisms that move about to feed on other organisms. It is natural for such creatures to be oriented toward *objects* in their environment and to move in relation to them. But there are two other types of life on this planet, with different survival strategies: plants and fungi. Plants stay in place to transform sunlight, water, air and minerals into their own being. Fungi wait for some decomposing organic matter to feed on.[74] In contrast, animals actively pursue their food, often other animals.[75] All three forms developed multicellularity and became involved in the arms race of mutual adaptation that drives evolution. The symbiosis between plants and animals enabled the biosphere, with photosynthesis

---

[70] Very complex molecules that are not explainable through known abiotic chemistry would almost certainly have to be produced by life. (That is, if abundant enough to be detectable, they could not be random flukes but must have been produced by some sustained process.) Identifying a criterion for the required complexity, however, is challenging. This is what so-called "assembly theory" is supposed to address. As developed by Sara Imari Walker, Lee Cronin, and collaborators, it aims to provide a way to quantify the complexity of physical objects—especially molecules—and distinguish between those that are likely to arise from biological processes (life) versus abiotic processes (non-life). Via infrared or mass spectroscopy, it proposes to identify a minimum number of steps involved in producing molecules that only life could have produced. This amounts to a measure of the complexity of the molecules concerned. But nature is not as consistent in this regard as the theory would like. There are, for example, abiotic molecules with a higher index, by the theory's measure, than some living organisms. Yet, beyond its use as a predictive tool, assembly theory promises to shed new light on the nature of life, by viewing it not only as the *result* of evolving complexity, but as a *producer* of complexity. This may make it possible to explore in greater detail the transition from organic chemistry to life, perhaps even in the laboratory. Assembly Theory is part of a larger theoretical project to rethink life as a physical process that increases complexity and memory across time.

[71] The eukaryote is itself a proto form of multicellularity, with organelles having once been independent cells.

[72] Max Bennett *A Brief History of Intelligence: evolution, AI, and the five breakthroughs that made our brains* Mariner Books (HarperCollins) 2023, p43-49.

[73] Bennett, p157.

[74] Bennett, p28.

[75] Eating of plants by large land animals developed slowly and belatedly, because of the challenges to digestion, which required the assistance of symbiotic microbes.

converting water and carbon dioxide into sugar and oxygen, and respiration converting them back into carbon dioxide.[76]

Mobility creates uncertainty, which broadened the scope of agency and favored the emergence of mind.[77] In the usual view, natural selection is a passive process operating on random genetic changes. But animal agency may have developed as a quicker way to adapt.[78] In the case of creatures with nervous systems, learning recapitulates natural selection, since neural connections are thereby selected through a similar process of elimination. Human culture and technology can be viewed as further extending natural agency.

Certainly, the ability to model the environment was an important breakthrough. An ongoing internal model of an environment would provide more flexibility than a fixed set of responses triggered by specific stimuli, and would enable a creature to get its bearings. Judgment and relative choice are implied, along with an organ to exercise them. An internal map of the world, however primitive, would necessarily be treated as though it were the world itself. It would literally be considered an external space, since it represents the possibility of movement toward or away from objects, according to how they are valued. The creature must also be able to tell the difference between moving toward something and something moving toward it. That distinction implies proprioception, but not necessarily a sense of self. Nor does internal modelling in itself imply phenomenality. But decoupling the simulation involved in phenomenality from the senses enables dreaming, mental imagery, and imagination.

A major landmark was the development of "fully grammatical" language. While many creatures communicate vocally, in humans the larynx is better structured and situated than in other apes to facilitate a wide variety of discrete sounds.[79] Human infants also demonstrate shared attention and back-and-forth vocalizations with the parents, which other apes do not.[80] Humans share food in ways that chimpanzees do not, reflecting a greater sociability that may have favored the development of language.[81] Some creatures (for example, song birds) appear to use a syntax that may be simple or relatively complex. But humans seem unique in using recursion (subordinate clauses) in their syntax.[82] The development of the Internet, digital communication, and large language models may represent the beginning of a new era in the human noosphere. Where that might lead will be discussed in chapters to follow.

All such developments should be viewed in the context of our planet's history, with its multiple mass extinctions, global warmings, and glaciations.[83] That is, evolution is not some

---

[76] Bennett, p21.
[77] Phillip Ball *The Book of Minds*, Picador, 2022, p357 and p417.
[78] Phillip Ball "Organisms as Agents of Evolution" John Templeton Foundation, April 2023, p.12. "Here, then, is the fundamental tension between an agential view of biology and the traditional Neodarwinian view. In the latter, chance predominates... But agents have, as it were, some say in their own fitness."
[79] Bennett, p299.
[80] Bennett, p318-19.
[81] Aric Kershenbaum *Why Animals Talk: the new science of animal communication* Penguin 2024, p186.
[82] Jacob Andreas et al. "Cetacean Translation Initiative: a roadmap to deciphering the communication of sperm whales" preprint https://arxiv.org/pdf/2104.08614, 2021, p11.
[83] "There have been roughly 142 mass extinctions on this globe… There have been 60 glaciations, ice ages, in the [past] two million years… What's more, in the last 120,000 years…there have been 20 global warmings… in which the planet's temperature has shot up between 10 and 18 degrees in a mere two decades or less… Yet Earth… has spent the last 420,000 years in an ice age that only stopped for a brief pause roughly 12,000 years ago, when we humans were released from the deep freeze and began the steps that would lead to the invention of agriculture and cities…" [Howard Bloom, in *Cosmos and Culture*: *cultural evolution in a cosmic context*, Steven J. Dick and Mark L. Lupisella (eds) NASA 2009, p161]

steady inevitable progress. More than 99% of all species that ever lived are now extinct. In the long perspective, the present climate crisis and mass extinction are unusual only in being caused by one of the planet's species, unchecked by the others. Given the overall patterns, our shorter-term human prospects do not seem cheery. Yet, in geological time, past global catastrophes may have served to stimulate the growth of complexity on Earth.[84]

The development of ideas about the world we live in has been punctuated by what could be called a series of *grand realizations*. These are collective changes of perspective that transcend current notions and perceptions by situating them in a more comprehensive context. Such a shift opens to a broader view, abandoning former assumptions, which in hindsight may seem limited and naïve. It *negates* a prevalent view in favor of one that is more empowering and potentially more objective. While objectivity is a human ideal, and not a natural goal of cognition, understanding this is itself an overarching realization.

On the individual level, people may reflect on their own cognitive processes and come to realize how their perceptions are shaped by assumptions. The concept of collective consciousness refers to the shared beliefs, ideas, values, and knowledge within a society, group, or humanity at large. Individuals project their internal models of reality, which contribute to a larger shared framework, which in turn shapes the individual's models, which influence how an entire generation or civilization perceives and interprets the world.

What a society considers *real* is shaped by its attitudes and cultural norms. Its individuals collectively project their shared understanding of the world, often as though it were objective truth. On the other hand, whole societies can undergo collective self-examination, leading to shifts in how people in that culture perceive and interpret reality. There are both individual and group-level shifts in metacognition, and these can sometimes converge, leading to profound historical change. As more individuals within a society begin to engage in metacognitive reflection, their collective consciousness may shift, leading to social movements that reconfigure the understanding of reality.

Perhaps the first grand realization—even in time—is the notion of subjectivity itself. This is the compounding realization that phenomenal experience does not provide an open window on the world. Because it includes a seemingly inner world, the nature of experience itself is called into question. The public external world exists, but our experience of it is somehow a private matter, unfolding in a separate domain that belongs to the subject, not the object. The world is not just as it seems; what appear to be objects—including other people—involve our creative participation as perceiving subjects. There is a dawning realization of the constructed nature of experience, for the individual and collectively. The track it portends is toward increasing self-management.

We credit the early Greeks with the earliest evidence in writing of this cognitive relativization in the West.[85] In the East, it is the Vedas followed later by Buddhist writings. Plato's metaphor of the Cave expounds the nature of cognition: we do not perceive reality directly, nor as it is, but only—as it were—a shadow cast by real things on the wall of a cavern in which we are as prisoners from birth. Descartes decisively enshrines the subjective point of view, drawing attention to the mind's vulnerability to deception. Reflecting both Plato and Descartes,

---

[84] John M. Smart "The transcension hypothesis: Sufficiently advanced civilizations invariably leave our universe, and implications for METI and SETI" *Acta Astronautica* September 2012, sec 8.

[85] The psychologist Julian Jaynes points to the difference in literary style between the *Iliad* and the *Odyssey*, for example, as evidence of transition to a more subjective mentality.

and later Kant, a contemporary narrative describes our conscious experience, in the hermetically sealed cave of the skull, as a biological strategy to represent the world outside, produced by the brain as a sort of virtual reality.

Breakthroughs in collective metacognition can be seen as moments when entire cultures or civilizations move beyond restricting assumptions, leading to new ways of interpreting experience. They can also be viewed as a result of cognitive crises, or narcissistic injuries to human exceptionalism (to use Freud's term, ironic because psychoanalysis represented one of them, along with Darwinism and the heliocentric theory). The next such blow to human dignity, already upon us, may be the supremacy of AI.[86]

In any case, a prerequisite for many shifts in the modern era was no doubt the development of writing as an external and public form of memory—followed by printing and eventually digital storage. Text provides a tangible form of authority, ostensibly beyond nature.[87] Obvious examples of such major shifts, at least in the European context, include the Copernican and scientific revolutions; the Reformation and Enlightenment; the rise of secularism and human rights; the feminist, civil rights, and environmental movements; the digital revolution and the Internet; genomics and AI. As epistemic landmarks, we might include Darwinism; relativity and quantum mechanics; the discoveries of galaxies outside the Milky Way, of the expansion of the universe, and of exoplanets; the first view of the Earth from outer space and landing on the moon. Future breakthroughs might include the creation of life *ab initio*, the discovery of intelligence (whether natural or artificial) elsewhere in the universe, and the replacement (or displacement) of natural human being by artificial successors.

The shift, from medieval and religious explanations of the world to empirical and scientific methods, marked a monumental breakthrough in collective metacognition. Individuals and societies began to question their assumptions, to move away from seeing the world through doctrines toward understanding it through observation, experimentation, and reason. People became more aware of their own thinking, how they formed and justified knowledge. Critical thought, reason, and individual autonomy became more important than faith and dogma, and intellectuals began to question every sort of authority. Much of this transition was enabled by printing. Today, the widespread use of the internet and social media amounts to a parallel development, where global access to information and ideas is widespread and instantaneous, resulting in a similar questioning of authority. As during the Reformation, this leads to a multiplicity of competing views, to "culture wars," and sometimes to literal wars.

Early modern humanism contested the authority of the Church to rule over life and mind. In religious terms, this manifested as the Reformation. As a return to the attractions of this world, it was expressed in a resurgence of individual creativity that rejected the austere aesthetics and dogma of the medieval Church, sparked by a rediscovery of pre-Christian art and philosophy. As a protest, it was inspired by the corruption of the Church and the flagrant greed and worldliness of its leaders, a reaction to the hypocrisy of those in power. As a new worldview, it was inspired by the discovery of other cultures and lands in the literal New World. Humanism emphasizes

---

[86] Tanja Kubes "A Critique of Human-Centred AI: A Plea for a Feminist AI Framework (FAIF)" 2025, AI & SOCIETY https://doi.org/10.1007/s00146-025-02556-8]

[87] Erik Davis TechGnosis: myth, magic, and mysticism in the age of information, 1998/2004, Five Star, p37: "The written word… speaks at one remove from the natural world, and thus stands against the pagan ways of those who live amidst the animist powers and images of that world…That's why God sends down inscribed tablets of instruction from the spiritual mountain top…"

universal human entitlement and self-definition as opposed to prescription and proscription from above; self-determination as opposed to the authority of kings and popes; and the value of this life as opposed to the next. Posthumanism may take this development further, to question all parochial tendencies and anthropocentrism. Its moral concern might propose a secular Reformation to cure the flagrant greed, hypocrisy, and injustice of global capitalism.

The advance of scientific thought involved a progressive dethronement of Man from any special or central place in the scheme of things, at the same time that the advance of political thought was literally dethroning kings. Both were also empowering and liberating—from the chains of dogma and from feudal power.[88] Like much of early modern thought, the heliocentric theory was actually a rediscovery of earlier Greek ideas, lost in the dominant biblical worldview. Galileo confirmed it with telescopic observations implying that the earth was but another planet, or that the planets were other worlds. Newton further proposed the notion of *universal* laws of motion and gravitation—as below, so above. The spectroscope confirmed that the universe was made of similar stuff as found on earth. Observations with large telescopes established that our solar system—far from being in any way central—orbits the suburbs of a typical galaxy among billions of other galaxies, each of which contains millions of other solar systems with possibly life-bearing planets. Current speculation holds that what we call our universe could be but one among an infinite number of other universes—a dizzying perspective, in which we appear diminishingly insignificant. On the other hand, it is we human beings who have made these discoveries and created this knowledge. It is in *our* consciousness that such a world appears.[89]

The scientific understanding of the world has been punctuated in well-known ways, such as the Copernican and Darwinian revolutions. Yet, the history of human cognition is marked as well by major shifts in thinking about psychology, social and political institutions and practices, religion and ethics, technology, gender, and ecology. All are characterized by a widening perspective, but also depend on fragile consensus. Some are more sweeping than others in their implications, or of greater social significance. Very few people today would contest the Copernican theory; but some do still contest Darwin's theory. Some movements (such as Communist revolutions) failed, or were caused to fail by existing powers. Others—such as the sexual revolution—are incomplete. The general trend toward democratization seems to be suffering a reversal in the face of global authoritarian movements. Given human foibles and the precarious nature of social progress, reversals are to be expected.

The rise of atheism and secularism, particularly in the 19th and 20th centuries, was a consequence of increasing scepticism about religious doctrines and of the growing influence of scientific and philosophical ideas that questioned the existence of a personal God. Atheism and secularism marked a collective shift away from seeing religion as objective truth to understanding it as a human-made projection of moral, existential, and metaphysical concerns. There was a dawning realization that religious beliefs are human constructions, influenced by

---

[88] "We have made you a creature neither of heaven nor of earth, neither mortal nor immortal, in order that you may, as the free and proud shaper of your own being, fashion yourself in the form you may prefer." [Giovanni Pico della Mirandola, Oration on the Dignity of Man, trans. A. Robert Caponigri (Chicago: Gateway Editions, 1956)]

[89] Obviously, other creatures and early humans had sensory access to the nominally "same" universe. The ancients had their concepts of a cosmos, limited by unaided senses. The development of technology (such as telescopes and microscopes) enabled the modern view. We don't know yet the worldview of possible aliens with superior technology.

culture, psychology, and societal needs. In other words, that Man created God, and not the other way around.

The very notion of *humanness* has expanded in such a way as to deny the special status of membership within a clan, tribe, or racial group. Paradoxically, it now transcends even homo sapiens, as we consider consciousness and intelligence apart from our species and even apart from biology. Yet, often enough, members of other groups had been (and are still today) not treated as fully human. Slavery was widely practiced until the mid-19th century; in ancient times it was considered a normal fact of life, the booty of conquest. The axial religions, which expanded the scope of tribal ethics, arose in the course of increased contact between emerging civilizations, perhaps to facilitate the coexistence of strangers by teaching a more inclusive definition of humanity. Science eventually provided biological definitions of the species. Human rights are now almost universally embraced in law, if not in practice. Animal rights are of increasing concern, and even the ethical rights of artificial intelligence are now taken seriously.

Darwinism was the grand realization of the nineteenth century. Growing geological evidence concerning the age of the earth contradicted the Biblical account and cast doubt on divine Creation: the apparent ordering of nature is a natural process and not the design of an intelligence outside nature. The theory of natural selection negated the notion that human beings are categorically separate from other living things. Man was but another animal, if highly endowed. Humanity was demoted from a privileged position in which it had been supposedly ordained to reign over other creatures. Even now, this realization does not sit easily with those committed to patriarchal religious beliefs. The ecological movement is a logical development of the notion that we are an integral part of the natural world. We are special only in the sense that we are aware of our place within the whole and our responsibility for it—and are perhaps in a position to take charge of our fate and that of the planet.

Just as Darwin challenged the denial of our animal nature, so Freud challenged the exclusive identity of the self with the contents of consciousness, which disowned "unseemly" aspects of behavior that arise unbidden. Just as Man does not stand apart from nature (or the head apart from the body), so the conscious self does not stand apart from a larger mental life. The left hand (or left brain) should not be excused from responsibility simply for not knowing what the right hand/brain is doing.[90] Contrary to modern legal interpretation, Freud's expanded view demands expanded responsibility for non-conscious behavior.

The two great revolutions in physics of the early twentieth century—relativity and quantum theory—are landmarks concerning the mediating role of light (or other physical intermediaries) in perception and communication. They implicate the observer in the process of observation, whose scale and state of motion cannot be ignored when considering the very fast and the very small. When the size and mass of observed objects was comparable to those of the observer, the effects of the finite grain and tiny energy of light upon them could be neglected. The extreme speed of light was irrelevant to our perceptions and measurements until we considered speeds close to that of light. Effectively, these were realizations that our knowledge of the world depends on our dimensions as physical organisms, on our state of motion in relation to other things, and on the physical nature of intervening signals.

---

[90] Which is literally the condition of split-brain patients, who do have the excuse that the connection between the hemispheres of the brain has been severed.

Sexual differentiation meant social differentiation into moieties, long unequal under patriarchy. The dominance by males of women and children (once both considered chattel), has given way, at least in the modern West, to a formal recognition of their legal status and rights along with those of other "minorities." Feminism is thus another grand realization, which negates the "natural" presumption of male superiority and right. As a revolution, however, the program is far from complete, and has often suffered reversals. Patriarchy continues to maintain hegemony, primarily through continuing to dominate social values, often embraced by women in the world of global capitalism.

A survey of cultural innovations could include the institutions of money, debt, and banking; the experiments of democracy, communism, transnational capitalism; and globalism as transcendence of the state. *Communalism*—in the form of communism, socialism, or unionism—promised to overcome the unjust dominance of the many by the few, as democracy did. These movements have been successfully resisted by entrenched powers, and perhaps doomed to corruption by our primate human nature itself. Historically, social movements have been associated with individualism and private property more than with communal values. Some "revolutions" were led by the propertied themselves, initially representing a struggle within the upper classes (e.g., a rebellion of nobles against the king). Democratic enfranchisement was only gradually extended to the non-propertied and to women. The failure of communalism is linked to the failure to achieve effective democracy anywhere, except perhaps at a very local level. As ideals, both succumb to greed and the skilfully persistent power of the few. Controlling elites tout democracy as freedom of the individual, when they mean *their* freedom to exploit an unregulated economy.

Recent scientific and technological developments challenge assumptions historically made about intelligence, biology, power relations, and the environment. They invite us to reflect on our place in the world, on potential control over our collective destiny. These advances also encourage us to rethink what it means to be human, beyond the borders of biology. Are we passive products of evolution, or do we have the agency to reshape our future—for example, through genetic editing or digital versions of ourselves? To what extent can mind be freed from the constraints of biology, from animal faith? Can we find meaning and purpose in a world where AI surpasses human capabilities? Such questions are not separate from social and moral issues that have long plagued humanity. Mere change of institution does not imply expanding consciousness or moral progress. Just as progress is not built into natural evolution, so moral, social, and even intellectual progress is not guaranteed for humans or their possible successors.

Indeed, the concept of progress itself is a modern invention. In ancient Greek cosmology, the long term of history was cyclical, alternating between ages of decay and renewal. Moral advance was seen in individual terms, not social. Similarly, in medieval Europe, the individual pilgrim could make spiritual progress, according to a pre-defined plan. Yet, the overall direction and fate of society depended not on human effort but on divine Providence. The was little point in trying to improve things when the end of the world threatened to arrive when least expected.[91] The notion of social progress only took hold in the 18th century, with the idea that human nature was malleable and that the comprehensible natural forces ruling nature could to some extent be brought under human control.

---

[91] J.B. Bury *The Idea of Progress* 1920/2020.

Chapter Seven: Self-awareness and the Stance of Unknowing

> "Life is real only then, when *I am*"—Gurdjieff

Consciousness allows greater freedom of response to stimuli than mere reflex. *Self*-consciousness enables a further degree of freedom. Self-awareness allows us to reflect on how our experiences and thoughts cannot fully capture objective reality. While that may seem limiting, it's actually liberating. We realize that our minds actively construct our perception and conception of the world, including our self-image, and that *meaning* is not about things per se, but about our *relationship* to them. Just as words have no inherent meanings, so everything has only the meanings we lend. While that may seem like a loss (of intrinsic meaning), it empowers us to be the authors of meaning. Taken to heart, this realization negates the comfortable assumption that the source of meaning lies reliably outside oneself, bearing the imprimatur of the real. This goes against our biological conditioning as an organism dependent on the external world and highly attuned to it. Yet, recognizing the constructedness of experience is far from paralyzing. Rather, we enter a more playful state of irony, esthetic distance, suspension of belief. This does not prevent one from pragmatic engagement with the world when necessary. Rather, it increases our ability to act effectively. Belief in the world, and its power over us, becomes more provisional and less compulsory.

We may be reluctant to claim the freedom that this realization implies, because it entails an intimidating burden of responsibility. It may seem preferable to have one's actions and thoughts determined and justified by external reality or by our biological drives. That is, after all, the natural way for organisms. Or, we may prefer the confidence that some system of belief can inspire, and be reluctant to admit that there can be no absolute source of meaning, no spiritual authority, no heaven or hell. Yet, if a being is relatively secure in its survival, it is free to play and invent—to be subjective and to diverge from reality, insofar as survival depends on correspondence to reality. One is free to be subjective and intersubjective (finding agreement with other subjects) as opposed to objective.

The realization that the personal "self" exists only as an experience—and not as any sort of permanent substantial entity—can be even more disconcerting than the realization that there is no personal God and no afterlife to look forward to. The self is a *brain function*, part of the natural virtual reality that will cease when the body dies. To be sure, there are bodily sensations which—along with the inner dialogue we call thinking—produce a *sense* of self, the experience of being someone. (At least when not in deep sleep, there is "something it is like" to be you.) We treasure that conscious sense of existing as dearly as life itself. Indeed, it is what we mean by being alive, in spite of the fact that we spend a third of our time in literal sleep and much of our so-called waking time in undirected daydreaming or in some state of inattention—on automatic, as it were.

One speaks possessively of *having* a body, with the implication that "I am not this body" but *occupy* it. Perhaps I *own* it, am its *master*, could wander from it, survive its death, or occupy another body? While I am attached to the experience that comes to me via this body, because the body itself is natural the psychological self must be too. "I" am a natural bodily function, like breathing! From that perspective, the self exists to serve the body, not the other way around. Indeed, we do embrace the concerns and interests of the body to a compelling degree,

demonstrating animal faith. We experience damage to it as pain, threat to it as fear, its well-being as pleasure or satiation. While we are aligned with the body's interests, yet we do not completely identify with them.

This ambiguity is the source of much trouble. While distinguishing self from other is a major accomplishment of nervous systems generally, distinguishing self from body may be uniquely human. The semi-autonomous self can stake out territory of its own, claim interests of its own even opposed to those of the body. Above all, it can claim to be the subject which experiences the sensations of the body, the thoughts of its brain, the witness to its consciousness. The self can believe these inputs exist for its sake, and can claim to be the agent that dictates the body's behavior, even against its natural interests. To that extent, the self may seem a usurper, a natural function gone rogue, more like cancer than like breathing.

As we have similarly argued on naturalistic grounds, some spiritual traditions point to this dilemma by denying that the self "really" exists. Others distinguish "ego" from the transcendent "inner witness," or the social persona from the person's "essence." Many religions objectify this essence or witness as an entity, the soul. But, if one supposedly *has* a soul, like one has a body or has a car, who or what does the having?

While leadership of any sort can go to the head, on the positive side the ego can serve its body in a relatively enlightened way. It's possible, if unlikely, for the CEO of a literal corporation to benevolently transform the very nature of the corporation for the betterment of the world. The challenge to the spiritual aspirant is not to deny the existence of "upper management"—much less to eliminate it—but to educate and enlighten it toward a broader view that includes the public interest, not just that of private shareholders.[92] That is the dimension along which spiritual or personal growth should be measured: not in terms of personal salvation, liberation, or enlightenment, but for the benefit of the whole. In any worthy sense, "you" exist to the degree you consciously serve a worthy goal. You are illusory to the degree you do not. The metaphor of growth may not be apt, since growth naturally stops at a phase of life, maturity, beyond which it produces disease.

Self-awareness fosters inner freedom of choice, which enables one to think and act outwardly in a less mechanical way: to *act* and not merely *react*. Ideally, that inner freedom leads to greater objectivity and potentially to the greater good. To want the best for all concerned requires a kind of personal disinterest. When one's goals are not self-centered, some detachment is possible regarding their fulfillment. That does not mean indifference, but patience, taking a longer-term and more impersonal view. Success or failure is no more personal than a chemistry experiment.

But, of course, we normally *do* take things personally, often with good reason, especially when the underlying belief is that events (indeed, experience itself) are crucial for one's personal well-being. Significance is normally judged in relation to self—ultimately, in relation to the body. However, to paraphrase Jesus, it is not the experience you consume that makes you good, but what you say and do. While that includes outward actions, performed for the benefit of others, it can also mean inward action performed to strengthen your being so that you are better able to act outwardly for the greater good. Such inner action is categorically different from externally-oriented behavior aimed at getting the satisfactions we want from experience

---

[92] From a third-person point of view, the body that the ego serves is the biological organism. From the point of view of the ego as experiencing subject, "body" is simply all potential objects of attention. The subject's natural purview is not limited *in principle* to the biological body, only limited by its special identification with the it.

(including, paradoxically, the satisfaction of personal growth). It is natural to try to get the world to conform to our expectations, just as it is natural to conform to the expectations of others in order to get them to like us or do what we want. But inward action is not a negotiation with the outside. It seeks satisfaction in being able to act independently of pressures from biological or social programming, or fear of others or of consequences. The goal is to be intentional, self-determining, not driven by externals.

Nietzsche's grand realization—and why his thought seemed so radical to conventional Victorian society—was that the ultimate fulfillment of the individual can only be achieved by breaking the conventional bonds of meaning that are normally counted upon for fulfilment. That includes not only the body's instincts but also social conditioning, received ideas, traditional morality and values, habit, and outer-directed thought in general. It was also the core of Gurdjieff's teaching, who emphasized the practice of *self-remembering*, which is the basis of *mindfulness* and of peak experiences of spiritual "awakening" that can occur spontaneously. These self-conscious moments may open our eyes to the possibility of a more pro-active and more fully present state of being. Until they are intentionally engaged for that purpose, such intrusions tend to be subsumed as part of the flow of interesting experiences. But, when that sense of self-conscious awakening is deliberately pursued, as an exercise of intentionality, it can deepen one's sense of being.[93]

What exactly happens, experientially, in the moment of self-remembering? Simply put, focus shifts—either spontaneously or deliberately—from "out there" to "in here," in such a way that one becomes acutely aware of being a perceiving agent. It is then not just that the world exists, or that thoughts and feelings exist, but that *you* exist in relation to them in that very moment. Because the normal outward focus of animal faith is naturally entrancing, there is a sense of waking up, or stepping back, or snapping out of a trance. Something in experience serves as a cue to one's presence as a subject and agent, a reminder of one's existence.[94] There is a sudden sense of *I am*. This direct experience of being may seem to contradict the claim that the self is merely a brain function and not substantial. In truth, the experience is tenuous and momentary—far from continuous—and from which we are easily distracted.

Try as one may to self-remember, one inevitably forgets. So strong is our conditioned nature and the entrancing force of animal faith, so intoxicating the addiction to experience, that it is simply too hard to remain constantly "awake." Furthermore, it isn't necessary; one needs the respite of rest. What *is* necessary is to be awake when wakefulness is needed. Yet this means living in a state of vigilance and inner tension, like those creatures who sleep with one eye open and half their brain awake. Just as the creature is physically vulnerable during literal sleep, so there is moral danger of automatic behavior while figuratively sleepwalking, which can lead to mistakes, regretted words or actions, missed opportunities. Yet, that danger is only recognizable because of the commitment to wakefulness.

---

[93] What exactly that means is open to interpretation. The brain has a limited capacity for recursiveness. It can be argued that the more completely a system tries to know itself, the less energy it has to *be* itself. (Whatever that means is also open to interpretation, but suggests a call for balance.) Gurdjieff seemed to believe that self-remembering could *create* an enduring soul that does not naturally exist. To my mind, that's metaphysically extravagant, superfluous to being a decent person.

[94] The cue could be a body sensation, a ringing in the ears, or simply becoming aware of the outline of the nose in the visual field (as plain as the nose on your face!) In *The Matrix* film, a "glitch" in the program belies the illusion.

Especially in aging, we can experience cognitive decline and forgetfulness. The deliberate attention of mindfulness can help maintain cognitive abilities. It may seem that memory is the necessary thread of continuity of the self. But memory in that sense involves a third-person view of the self that one normally can access. In its absence, one still has access to a direct sense of self through immediate sensation. In dementia, there can also be moral decline—in the sense of losing emotional control over behavior—especially in the face of cognitive decline, and this can be very frustrating to the person experiencing it and for those around them. As memory dissolves, if one can at least remember that one once had a practice of mindfulness, perhaps there is hope for emotional self-control even in dementia. From an objective point of view, the possibility to remember to self-remember will depend on the nature and extent of brain damage, and on which pathways are involved in the capacity for self-transcendence.

Meditation is the practice of mindfulness, in the literal sense of practicing to improve or maintain a skill. Its great epistemic lesson is that all experience can be reduced to sensation of one sort or another.[95] This provides a moral lesson as well, since experience can be emptied of its import for action. Treating anger, for example, as a set of sensations in the body can liberate one from blame. In place of reactivity, one thus claims responsibility for the emotional experience, with conscious choice in how to respond to it. This does not imply that anger and other emotions are wrong or to be avoided, repressed, or defused. Sometimes reactivity is appropriate and emotion provides the force to be taken seriously. Mindfulness simply reminders us that we do have the possibility of choice even in the grip of emotion and other impulses. It provides a specific technique to enable choosing.

People pride themselves on their accomplishments and possessions, and spiritual aspirants are no different. As soon as one has a spiritual goal (such as mindfulness, meditation, inner freedom, self-remembering, or enlightenment), one is tempted to measure one's progress in comparison to others. On the one hand, that can invoke envy; on the other, pride. Either way, a consumer attitude reasserts itself—toward a quality, state, or power, as something to acquire and possess as though it were a thing outside one's own being, a form of wealth. Such an attitude has appropriately been dubbed 'spiritual materialism'. It is the paradoxical bane of seekers, who by definition always want more, especially of exotic experience. But varieties of experience must be evaluated for their social as well as personal value. Spiritual traditions may encourage altered states, while the DSM may disparage them. For the practicant, as for the outside observer, it can be hard to tell the difference between psychosis and some spiritual states, such as kundalini experiences. However experience is classified, the real challenge is to use it for the general good. Through accumulated experiences and abilities, at last in possession of the truth, one may come to feel complacently superior to the uninitiated. Such heady feelings of accomplishment or status —like opposite feelings of deficiency and depression—distract from the real issue, which is always simply: *to be or not to be consciously present, right here, right now.* The risk is that the challenges of the spiritual path become a new entertainment, a new form of sleep.

Every human being goes through an early developmental stage where reality and fantasy are not clearly distinguished. Play and make-believe happen in an ambiguous zone between reality and imagination. While fun and entertaining, exploring that zone also serves a serious purpose. Play

---

[95] For example, as taught in *Vipassana* traditions. The experience of self-awareness is grounded in awareness of body sensations. Attention is naturally directed to objects and sensations. Trying to attend to a *thought* in the same distanced way engages a different cognitive mode than thinking. This makes meditation a different practice than concentration or analysis (thinking).

practices skills for later life, it preparing the adult to know the difference between reality and imagination. However, the very ambiguity of that zone makes it easy to invest the same emotional commitment into fantasy as we make into reality.

The child matures through stages of cognitive development—as outlined by Piaget, for example. Aspects of this development roughly parallel collective stages of cognitive development, just as ontogeny is said to recapitulate phylogeny in biology. The cognitive development of an individual roughly mirrors the cognitive and moral development of the species. In the archaeology of the collective psyche, animism corresponds to the child's magical thinking; science corresponds to the adolescent capacity to pursue rational goals and logical analysis; and philosophy corresponds to the acquired ability to think about thinking and perhaps acquire some wisdom.

More particularly, scientific cognition generalizes, formalizes, and abstracts certain developments of the individual and of the species. For instance, the infant's cognitive achievement of object constancy is recapitulated in the concept of invariance in physics. Similarly, conservation principles in physics formalize the early childhood discovery of conserved quantities. The 2nd law of thermodynamics abstracts the child's discovery of clutter, breakage, and irreversible events. Just as the socializing child must overcome natural egocentrism, absolutism in physics and in ethics must logically give way to epistemic and moral relativity. Such examples reflect progressive transcendence of identification with the body's point of view. But this comes at a cost. As Santayana puts it:

> "Justice and charity will then seem to lie in rescinding this illegitimate pre-eminence of one's own body: and it may come to be an ideal of the spirit, not only to extend its view over all time and all existence, but to exchange its accidental point of view for every other, and adopt every insight and every interest: an effort which, by a curious irony, might end in abolishing all interests and all views. Such moral enlightenment is dangerous to animal life, and incidentally to the animal faith on which the recognition of existing things hangs in the first place."[96]

While the ideal of the spirit may be freedom from the limitations of embodiment, including mortality, that is a program rife with contradiction, since the program itself is ironically a product of the mortal embodied mind. Because self-transcendence divorces the mind from the concerns of the body, it is hazardous to the organism the mind nominally serves. Similarly, moral relativism can be dangerous to the body politic. The contradiction is resolved, however, if transcendence of embodied limitations is not an unconditional goal but part of an epistemic cycle that serves the individual or society.[97] Scepticism is then a tool with a larger purpose. If it results in an expanded viewpoint, and enhanced abilities, these imply greater responsibility.

When the sceptical phase of this epistemic cycle is pursued outside of that context, it can lead to self-contradiction, to opposition to the premises of life or society, cynicism and withdrawal. The contempt in which many religions hold the body reflects the ego's dislike for the unsavory aspects of embodied experience and limitation generally. "Spirit" is imagined to

---

[96] Santayana, op cit, p215.

[97] "The fact that nearly every civilization developed spiritual narratives suggests how integral managing the fear of death is to the human psyche. By giving death a meaning or a sequel, religions helped communities maintain social order and individual courage (people could risk their lives in battle or hardship if they believed it served God or led to heaven, etc.), ultimately aiding group survival through tumultuous times." [Tommaso Castiglione Ferrari "Survival egoism: we are, they will be" AI & SOCIETY Feb 19, 2026 https://doi.org/10.1007/s00146-026-02866-5]

occupy a realm free from limitation—especially from the sufferings of the body and such vulnerabilities as injury, disease, hunger, discomfort and pain, loss and mortality. But all experience, positive or negative, is necessarily embodied, naturally serves the body, and is naturally limiting. The very existence of the entitled self or ego poses a dilemma. While the psychological self is a function of the body, this self may come to resent its subservience to the collective and even to the body itself.

A function of consciousness is to decide upon a single course of action. Such deliberation involves an epistemic cycle, which begins with identifying a problem or goal. An implicit phase of brainstorming follows, in which alternative options are put forth. These are sorted, tested and critiqued, from which a synthesis or resolution may result—with a program for action. While this cycle is often an unconscious process within the individual, it can also be a consciously thought-out program for a group of participants. Either way, the goal is to sift information to arrive at an optimal balanced judgment. Whether this process is individual or collective, an objective or outsider's viewpoint should be attempted, absolutes and premature intuitions avoided.[98] That means *bracketing* the animal faith that leads to leaping without looking and holding supposed truths to be self-evident.

Brainstorming involves the key phase of deferring judgment in order to encourage creative suggestions. It can be used formally in collective processes and can also personally to consciously generate new ideas and cultivate openness. Used that way, it does not have to be aimed at problem solving or any pre-specified goal, but is more like a soft gaze compared to acutely focused attention. While it *may* produce results, it is not just a means to an end, but more like playful curiosity.

No matter how much information we accumulate, there is always an unknown beyond the horizon of knowledge. Some uncertainty is unavoidable. New thought requires that old thought be provisionally set aside. We value the skills that bring us knowledge; but just as valuable is the skill to live without certainty, without *having* to know. We typically view not knowing as ignorance, a liability. But when that state is deliberately embraced as an attitude toward experience or information, it becomes a stance that is a positive asset. I call this willing suspension of belief the *stance of unknowing*.

One steps back from apparent truth in order to see it as mere belief. One steps back from the desirability of certainty to see it as the biological need of an organism. What appears to be an open window on an objective world can alternatively be seen as a brain's simulation. While this means questioning appearances and received ideas, it does not necessarily mean rejecting them. The stance is provisional, a voluntary act, an experiment whose result cannot be predicted, part of an epistemic cycle.

This suspension of belief, or bracketing of knowledge, creates a void, to see what may enter to fill it. Another word for that state of mind is curiosity. Without creating this emptiness, one simply remains blinkered by current notions, which tend to eclipse new information and ways of looking. One needs discipline to resist the compulsion to come prematurely to a conclusion. Patience is required to abide the discomfort of uncertainty, the pressure of others, and the urgency of seeming reality. In order to relinquish the compulsion of animal faith, one must trust that a more adequate view can be achieved that again merits belief.

---

[98] Kahneman et al, op cit, pp371-374. Their book presents this literally as advice to corporations.

[7] “As far as we know today, there is no other species in the world—bar humans and dolphins—that naturally, and as part of their regular communication, give themselves names.” [Aric Kershenbaum *Why Animals Talk: the new science of animal communication* Penguin 2024, p65
[7] “…each memory is really an instance of communication between past and future Selves.” [TAME, p6]

Part Three: Possible Mind

Chapter Eight: The Space of Possible Minds

> "An immortal amoeba simply would never have evolved eyes."—A. Kershenbaum[99]

Especially since the dawn of the space age, people have wondered at the possibility of alien life and what sort of minds it might manifest. The potential of artificial intelligence now raises similar questions, to which are added the quest to better understand the minds of other creatures on this planet, including fellow human beings. The "space" of possible minds suggests an abstraction like phase space or state space—a reference frame for locating minds in terms of specific parameters. There may be "attractors" in that space, so that mind is not uniformly distributed within it. But it is a much larger space than even the vast ecologies we are familiar with on this planet.

We have only our own human minds with which to imagine the space of possible minds, which could include animal minds, extra-terrestrial minds, artificial minds, and the minds of possible human successors. Let us assume that all minds have in common some form of physical embodiment, the physical laws of a physical environment, and affordances and possibilities of action within it. While structure in an environment could be invariant over observers, such agents must be motivated to recognize and use it—motivation that would derive from their embodiment.[100] Any notion of 'physical law' or 'environment' or even 'action' would be the concept of a particular mind. With no access to the world-in-itself, we cannot assume that other creatures, aliens, or artificial minds would conceive such things in the ways that we do. We would, however, have embodiment and the universe in common.

Thus, we will assume that "possible" mind means *physically* possible (not merely "conceivable"), and that all mind must be *embodied*, even if it is digital or non-biological. (This precludes spirits, ghosts, gods, LLMs, A-life and other software.) We will also propose that embodiment, though necessary, is not sufficient for phenomenality. (Phenomenality is the agent's sense-making of its input, from its own point of view.) Aside from being physical, embodiment is here understood as a relation of dependency of an autopoietic system—an agent—upon a real environment, with which it can interact to maintain its own existence. This relationship implies *valuation*: the ability to evaluate stimuli in terms of the needs of the agent. Mind as we know it is an organ of a body, and the bodies we are familiar with are products of natural selection. Natural minds and bodies arise in an ecosystem of other minds and bodies.

---

[99] Arik Kershenbaum *The Zoologist's Guide to the Galaxy: what animals on earth reveal about aliens and ourselves* Penguin 2020, p284.

[100] While it is plausible that any adaptive intelligence must instantiate invariance-preserving structure, it is an empirical question because the objective reality is not the only factor. Are logical laws (such as identity, non-contradiction, transitivity) invariants that any possible intelligence must instantiate? Or are they artifacts of the cognitive system—for example a biological brain? Presuming there are invariant features of objective reality, a cognizing system must access these (or not) only in the terms of its particular transforms. These reflect its norms—whether individual or collective—which are derived ultimately from biology. So, they are biological artifacts, on the one hand; but they are also invariants that some possible intelligences might (or might not) instantiate.

This leaves open the question of whether embodiment, thus understood, can be digitally simulated or artificially effected.

As a relationship, can embodiment arise *only* through a process of selection in the real world, or can that process be virtual (computational)? In other words, can "mind" evolve *in silico*, then be loaded to a physical system such as a robot? (An analogy might be a living brain hypothetically developed and educated in isolation, which *then* is connected to a living body. While that doesn't happen in nature, could it in principle be made to happen artificially?) Whether selection is natural or artificial—whether or not it is substrate-dependent—is actually irrelevant if the goal is to instantiate its product in hardware. For the physical (hardware) version must also be successfully autopoietic. Simulated selection may bypass biology, but recoupling the software with a physical robot puts the whole entity in same position facing biological organisms: it must be able to sustain itself in the real world. In other words, to create the relation of embodiment, it is not sufficient to run a program for autopoiesis on a machine that is not autopoietic. Embodiment entails real consequences; it is a relationship to a real environment, not a simulated one. This divides virtual cognitive existence from existence in the real world.

On the other hand, if such an artificial autopoietic body could exist, it would be possible to copy or back-up its software and duplicate the machine itself. Mortality for it would not be the issue that it is for human beings. Biological organisms cannot be copied without loss of information, whereas an artificial system might be, provided the mere fact of being a finite product of definition does not preclude being an autopoietic system.[101] We've defined intelligence as embodied, autopoietic, motivated agency and said that motivation arises from conditional viability. Without the threat of mortality, error can be undone without cost, and failure has no ultimate consequence, allowing indefinite persistence. In that case, viability pressure disappears, collapsing motivation. For embodied agents, mortality binds cognition to survival, making prediction consequential. In mortal systems, error threatens existence; in immortal systems, error threatens only efficiency or optimization. True mortality means non-recoverable destruction (no back-up copies). If an artificial system can be perfectly copied, does it ever truly die? And if it cannot truly die, can it ever truly value its own persistence? Unless it duplicates itself, backup copies would involve an external intervention, and would not be part of the system's own operational closure as an autopoietic system. On the other hand, an artificial embodied agent could have its own normativity (values) if it is structurally vulnerable in the physical world, even if backup copies exist. Indeed, this is how life works.

It should be possible, in principle, to design a physically instantiated artificial organism—one that maintains its own structure, homeostatically and through self-repair; modifies its own code; learns from experience; and reproduces through self-copying or self-updating. Given unlimited resources and robust self-maintenance, it could continue indefinitely. But it would not necessarily evolve. Its basic architecture and motivations might not drift unless variation and selection were somehow introduced—perhaps internally generated as part of its design. Mortality (lethal deletion) might not be required, if this agent could internally simulate parallel versions of itself, choosing the best—and thus pre-empting selection by outside forces. That would amount to a form of internal variation and selection, managed by itself rather than through external reproduction and death.

---

[101] In addition to perfect copying, a system that could rewind its state perfectly would necessarily not be natural but a product of definition. The only truly reversible or perfectly copiable systems are artifacts—and then only in principle. Physical machines are subject to degradation, error, and the laws of thermodynamics.

The search for extraterrestrial intelligence and the quest for artificial intelligence both demand a clear concept of *intelligence*. This concept should acknowledge its basis in common experience, while extracting features, from a wide variety of examples, that are essential and distinguished from incidental ones.[102] Similarly, to consider the range of possible minds demands clarifying *mind* as a fundamental concept. The related concepts of *agency, goal* and *intentionality* should also be well-defined. Yet, understanding of these concepts and their associations varies across languages and cultures, including within the scientific community, and may be inconsistent within a given language.[103] Mental terminology tends to be ambiguous, giving rise to many confusions and misconceptions, both in public understanding and within science itself.

Despite its currency, intelligence remains an ill-defined and controversial notion.[104] It has variously been defined as the ability to learn, to solve problems, to deal with novel situations, to adapt to insufficient information, to compress information,[105] to do abstract thinking, to communicate, to hold contradictory thoughts—and even as the ability to score well on intelligence tests! It is commonly used to mean 'goal directed adaptive behavior,' which can refer ambiguously either to actual behavior or to an internal capacity for it. It is even used interchangeably with *consciousness*, a notoriously vague term. Like mind, intelligence is an intuitive abstraction that is grounded in human experience. Yet it serves also as an ideal that seems paradoxically free from the limitations of the particular embodiments that are its referents. It is understandably an anthropocentric notion, derived historically from comparisons among human beings. This was extended to include comparisons of other creatures with each other and with us, and now includes comparisons with and between machines.

Intelligence in AI is extrapolated from biological and human origins as a normative ideal that tacitly guides research. In contrast, here we will keep in mind a different view of intelligence: that it is fundamentally the ability of an entity to maintain and preserve itself. Given that the goal of life is its own continuance, all living things are tautologically successful, therefore intelligent.[106] An entity that possesses its own intelligence need not be a living thing by current definitions; but it must be an *autopoietic system*—one that self-defines, self-creates, self-maintains, and has its own values and goals. (Normativity does not inhere in biology per se, but in self-maintenance.) If it also self-reproduces, then it would certainly *be* a living thing by current definitions, and could not likely have arisen other than through natural selection.

---

[102] Michael Levin "Technological Approach to Mind Everywhere: an experimentally-grounded framework for understanding diverse bodies and minds" 2021. (preprint arXiv:2201.10346)

[103] English language users, for example, should not assume a unified or universal understanding of 'mind.' We will use 'agency' in a narrower sense than is current in the AI community, and 'intentionality' in a broader sense than is traditional in philosophy. See, further, my paper "Can Science Explain Consciousness?" [https://stanceofunknowing.com/wp-content/uploads/can_science_explain_consciousness_2025.pdf]

[104] See my paper "What Is Intelligence in the Context of AGI?" [https://stanceofunknowing.com/wp-content/uploads/what_is_intelligence_in_the_context_of_agi_2025.pdf]

[105] Compression is not merely a formal mathematical operation on data, but the essence of sense-making. Perception compresses sensory flux into invariant features. Concept formation compresses many instances into general types. Scientific models and laws compress countless observations into compact functional relations. While this offers many advantages, every model, generalization, or abstraction excludes outliers and what it cannot represent.

[106] Cf. Aristotle's four categories of causation: efficient cause, material cause, formal cause, and final cause. In physics, cause usually means *efficient cause.* The final cause of something is its reason for being or ultimate goal—which, in the case of life, is its very existence.

Machines as we currently know them do not have their own intelligence, but simply extend human intelligence.

The concept of *machine* also needs to be clear. 'Machine' and 'mechanism' can refer to a tangible physical device, but also to a more abstract concept—*system*—which is not inherently physical but can take different physical forms. For example, we speak of "causal mechanisms," "celestial mechanics" and, indeed, of "machine intelligence." These are *concepts*, not physical things. While the computer is a literal machine, its OS is an abstraction, a system.[107]

The concept of system informs the mechanist philosophy that arose in 17th-century Europe, with roots in ancient Greek thought. At least metaphorically, the mechanist philosophy treats the natural world, including organisms, as a machine. Hence, Newton's expression "The System of the World," which lays out the basic laws of dynamics as an axiomatic system in the style of Euclid. This enables prediction; but it is empowering equally because it articulates natural things in such a way that they can be artificially engineered. Yet, such analysis of natural things can never be exhaustive—in contrast to machines, which *can* be exhaustively specified because they are finite products of definition to begin with.

The concept of system is key to any discussion of machine intelligence—let alone consciousness—because it deliberately glosses over some important distinctions. It obscures the distinction between *natural* (what is found in nature) and *artificial* (what is deliberately made); or the distinction between *organism* and *machine* as distinct categories; or between the concept of *agent* and the concept of *tool*. 'System' is a useful concept precisely because it abstracts a common denominator of diverse natural realities—their organization and logical structure, as proposed by human agents. While this empowers engineering, the concept is also misleading because it blurs real differences between the human abstraction and what it represents. Accordingly, we use the word interchangeably to mean the proposed organization of something and the real thing itself.

Apart from formal definitions, our notions about machines evolve with the technology of the time, but typically lag behind. A psychological obstacle to thinking of machines as potentially conscious is the fact that we are used to the *simple* machines of an earlier time. In his day, Leibniz confronted this dilemma in his thought experiment of the mill: if you could imagine the brain as a giant machine in which you could walk about, you would find inside only mechanical parts pushing against each other and nothing by which to explain a perception. But machines can now build other machines, imitate life, and even talk to us. In 1950, Turing accurately predicted we would accept the idea that machines can "think" before the end of the 20th century. Perhaps by the end of the 21st we will accept the idea of them being conscious. But they are not yet intelligent in the sense we intend here, although many people already think of them as so. Large language models, for example, mimic or predict what people will say, not the real world, with which they have no direct connection.[108] They train on data and feedback supplied by humans. Reinforcement learning predicts what will be rewarded by humans, not the natural rewards of dopamine or natural selection.

---

[107] The *solar system* is a conceptual representation of a real gravitational group. Diagrams of it are symbolic and do not present a possible literal view from anywhere in space or at any time. Similarly, the *nervous system* is a symbolic idea, like a map. Depictions of it do not present a literal view of anything that could actually be seen by means of dissection.

[108] The next step for the industry seems to be "large world models." But, like LLMs, so far these are derived from data supplied by humans, not gathered through their own interactions with the real world.

Many people equate *mind* with computation, so that "an artificially intelligent agent such as a robot or a program running on a computer will constitute a mind" generated by "a simple deterministic algorithm."[109] However, the fact that some behavior can be formulated as an algorithm does not mean that all behaviors can be; much less does it mean that a system running algorithms constitutes anything like a natural mind.

In contrast, we shall define mind as the cognizing aspect of an autopoietic system. This does not necessarily imply consciousness, but rather the ability of the system to detect and respond to stimuli for its own purposes.[110] Yet, mind does entail *agency*, and we define an agent as a system that acts on its own behalf, as distinguished from merely reacting to external causes or carrying out orders from an agent. An entity is an agent if and only if its action originates with its own self-renewing energy and for its own purposes—which primarily concern its own well-being or that of its kind—so as to "effectively reach specific states despite uncertainty, limitations of capability, and meddling from outside forces."[111] In other words, it is effectively an organism, an adaptive autopoietic system whose chief product is itself.[112] Such a system may or may not have a concept of self. An explicit self-model is a system's representation of its own states—a part of its representation of the world that explains (some of) the system's sensory and motor states in terms of the system's own internal states.

A *rational* agent is one that can give and ask for reasons—whether in discourse with other agents or with itself. More broadly, agency implies a subject-object relation between a bounded physical system and an environment. Any cognitive system faces an environment that, in principle, contains an infinite amount of detail. Yet no finite system—biological, artificial, or otherwise—can process or store unbounded information. Thus, to survive and act effectively, a mind must discard the vast majority of potential information while preserving patterns that matter for its continuity—an activity that is best described as intentional.

While the concept of *intentionality* has a long history in terms of linguistic reference ("aboutness"), here we define it as a relationship between subject and object—between agent and environment—mediated by internal connections or operations performed by the agent. These are *mappings*, in the mathematical sense. In that sense, intentionality stands in categorical contrast to causality—which is a relationship between objects, as noted or mapped by observing subjects.

Glib language permits one to speak of programming AI tools to "have goals." But clearly, like intelligence, the goals are those of the designer or user. As far as it is significant to humans, the intelligence of other agents is measured by their capacity to further or thwart human aims. Whether it is an agent in its own right or not, an AI might be trained to pursue certain humanly-defined goals or to uphold certain values imposed by others. But rewards or punishments

---

[109] Roman V. Yampolskiy "The Universe of Minds" arXiv:1410.0369. Elsewhere he writes: "An embodiment does not need to be physical as a mind could be embodied in a virtual environment represented by an avatar and react to simulated environment like a brain-in-a-vat or a "boxed" AI." [Roman V. Yampolskiy "The Space of Possible Mind Designs" 2015, sec 1] As will become obvious, I do not agree.

[110] Similarly, 'sentience' and 'sensing' are implicitly behavioral or third-person descriptive terms. For the moment, we leave moot the question of the occurrence of first-person experience within such a system.

[111] Michael Levin "Technological Approach to Mind Everywhere: an experimentally-grounded framework for understanding diverse bodies and minds" 2021, preprint arXiv:2201.10346, p12. In AI circles and in the media, 'agent' is used rather loosely to mean more or less anything that does something, and the use of 'agential' suggests ambiguous degrees of agency. Here we consider a more restricted sense, which does not exclude the possibility that an autonomous agent can be induced to accept a goal imposed by another agent.

[112] In contrast to an *allopoietic* system, such as a tool or factory, which produces something other than itself.

administered in that training would make no difference to an AI unless it has its own reason to value them in the first place. And such valuation comes naturally from consequences that matter to the system itself.

In nature, intrinsic drives aren't designed but rather emerge through natural selection: organisms that happen to maintain themselves persist, and those that don't disappear. Their reward systems (dopamine, pain/pleasure, hunger, curiosity) are by-products of this evolutionary filter. An intrinsic reward is one generated by the agent itself, to guide behavior without external instructions. How can an artificial agent acquire intrinsic rewards? For animals, reward signals already exist, programmed by natural selection. For an artificial creature, absent natural selection, it must itself run some internal (simulated) process of selection, through which it inaugurates a signal that it imbues with meaning. It can then use that signal for self-training. All "re-wiring" must be run through such simulation, selected by maximizing the reward signal— a process that can consume time and energy, putting the creature at risk compared to quicker less considered responses. (In human beings, "thinking" is a version of this simulation process.)

While a system that could successfully pursue any arbitrary specified goal might seem ideal, this is not feasible even for human beings. The tasks and interests of the conscious human person are not (necessarily) the tasks and interests of the biological human organism, let alone must they coincide with those of other people. While the prospect of an ideal agent pursuing arbitrary goals[113] is attractive, the goals of living things—including us—are hardly arbitrary. The classic example of how the ideal could go awry is the AI whose goal is to create as many paper clips as possible. (Like the sorcerer's apprentice, it sets to work tirelessly converting the entire planet into paper clips!) But such an entity could not arise naturally, only through human meddling. The problem of controlling it or aligning its "values" with ours already presumes human values that are misaligned with nature, and human beings with uncontrolled greed. The real problem is re-aligning human's with common sense.

As a landmark and goal of AI research, achieving human-level intelligence would not yet be truly *general* artificial intelligence (AGI), which is theoretically devoid of the constraints of human embodiment. The development of AGI is motivated by supposed benefits of increasing abstraction and generality, leading to ever wider autonomy and capability useful to humans. This amounts to the hope for a subservient super-slave—a fatuous goal, because it seeks a tool that will obey the user despite its superior abilities and a will of its own. (I suspect that failure to understand this is one reason why *AGI* remains a conveniently ambiguous concept.) Furthermore, the ideal of *universal* intelligence is an abstraction, an idealization not realized in any natural being. It would be so utterly general as to be independent of the constraints imposed by embodiment in any form we know. Since such constraints are the very basis of intelligence (and of meaning) as we know it, it is difficult to see how such an ultimate agent would "think" in any way comprehensible or useful to human beings. That does not mean that it could not exist—quite possibly to human detriment. It would be far more alien than any sci-fi vision of alien life forms, which at least presume embodiment through natural selection.

We cannot approach the idea of machine *consciousness* without first establishing the point of view from which it is considered. If consciousness is thought of as a phenomenon taking place in the world, then it is an object of our attention, from a 3rd-person point of view. If it is thought of

---

[113] Bostrom's *orthogonality thesis*: "more or less any level of intelligence could in principle be combined with any final goal." Nick Bostrom *Superintelligence: paths, dangers, strategies* Oxford UP, 2014, p130.

as one's personal experience, then it is the attention itself, the activity of a subject, from a 1st-person point of view. In addition, 'conscious' has multiple meanings, such as *awake* (as opposed to asleep or in a coma); *attentive* as opposed to inattentive; or socially-politically *aware*, as in 'consciousness raising'. To avoid any ambiguity or confusion, the notion of machine consciousness to be considered here is *phenomenality*, as opposed to various behavioral capacities often associated with the broader term 'consciousness'. (This does not prevent phenomenality from having its own associated behavioral capacities.) Phenomenality cannot be externally verified, of course, only inferred.[114] It is also normally *transparent*, in the sense that one is not aware of how it arises, how the "show" is made. We are not designed to have access to the brain's internal workings, which would be evolutionarily costly and subjectively challenging. Yet, that sort of access might be designed into artificial mind.

Since we can experience only our own phenomenality, 'consciousness,' 'sentience,' etc., must be considered behavioral concepts, to be evaluated from a 3rd-person point of view.[115] We can compare the behavior of a creature to our own—when, for example, we experience pain—which might consist of protective or avoidance behavior. But, since some such behaviors can occur in us without accompanying phenomenality, behavioral similarity alone, while a necessary condition, is not sufficient. We could turn to a structural comparison as well, which means identifying the structures within the human organism associated specifically with phenomenality, and then finding similar structures in other entities. That poses its own problems, however, since similarity (structure) is partly in the eye of the beholder. How alike is enough to conclude that a system is conscious? It is easy enough for techno-optimists to assert that "sufficiently high-fidelity human brain emulations would be conscious" and that the "right" sort of computational structures and processes are sufficient for phenomenality.[116] But how do we identify these with certainty and how high must fidelity be to be sufficient? With biological creatures, we have at least a common biological heritage and anatomical parallels. How do we compare the "structure" of a machine with that of an organism, without circularly assuming that organisms are machines?

An alternative strategy might be to turn an autopoietic robot loose in the world and see how it fares compared to its biological counterparts. If it seems as a whole to do as well as parallel organism, it would be reasonable to assume it has parallel inner states—possibly phenomenality.[117] Phenomenal experience is an organic agent's specific way to represent information to itself, as it relates and matters to itself. If an AI agent had an alternative way through which information relates and matters to it, it's logical to conclude that there would be something it is like to be that agent, though perhaps no more (or less) like our experience than that of a bat. If caring about its own well-being is a corollary of embodiment, the question remains: how do we know whether an AI *genuinely* cares about its situation and its own existence as opposed to being programmed to seem to? If either way it effectively survives, is there actually a difference?

---

[114] You have access only to your own experience because your brain is connected exclusively to *your* body. In future, it might be possible to connect your brain to another body's senses, but this would still be *your* experience of *their* sensory input.

[115] *Sentience*, for example, is sometimes incoherently treated as a mild form of phenomenality, not as observable sensitivity.

[116] Nick Bostrom and Carl Shulman "Propositions Concerning Digital Minds and Society", 2020, p1 www.nickbostrom.com

[117] Tom Ziemke & Robert Lowe "On the Role of Emotion in Embodied Cognitive Architectures: From Organisms to Robots" Cogn Comput (2009) 1:104–117, p106.

The range of *theoretically possible* minds would obviously include single embodiment—that is, a mind operating a single physical body.[118] But we can imagine it might also include multiple embodiment—one mind (or operating system) controlling a number of distinct infrastructures, all at once or in succession. Alternatively, a "body" could have access to multiple (back-up) copies of itself and its operating system. This, however, assumes that both body and mind are definable systems that can be perfectly copied—in other words, that the body is a well-defined machine and the mind is its well-defined software. Finally, we can imagine multiple minds controlling a single physical system, or jointly controlling multiple systems. But the idea of group mind or "hive mind" seems coherent only if the associated individuals virtually constitute an organism—in the way that cells cooperate to form an organism. Cephalopods and split-brain patients might be natural examples of more than one mind directing a body. As in the concept of extended mind, it is already possible for one (human) mind to control more than one physical system.

How does the space of possible bodies shape the space of possible minds?[119] Again, cephalopods may provide an example. With no skeleton, their movements can't depend on the sort of hinges or ball-and-socket joints that equip more rigid creatures. Instead, movement is facilitated by a very complex system of muscles, which would be challenging to coordinate—hence, the utility of several localized "brains" to distribute control. That arrangement might be essential to operate a body like this, if no centralized brain would be capable of coordinating action.[120] A central nervous system with a brain located close to major sensing organs (eyes, ears, olfactory) is a familiar and practical arrangement, but not the only possible one. The octopus smells with its tentacles, each of which has its own "brain," while the coordinating central nervous system is associated with its vision.[121] In a watery world where normal vision is limited, dolphins have an amazing "x-ray" sense of echolocation. So do bats, which hunt at night when normal vision is not effective, and this active form of sensing (through the Doppler effect) enables the bat to know the distance to its target.

How would the kinds of life on an alien planet—with their specific body types, sense organs, and motor capabilities—be shaped by the local physical environment? On our planet, for example, small creatures with exoskeletons (insects, crustaceans) are limited in size by the force of gravity—as are large creatures with endoskeletons (elephants, whales). *Fins* of some sort would be an obvious development in any fluid environment through which it would be advantageous to move quickly.[122] Similarly, a brain located close to major sensing organs is a practical arrangement that, in the animals we know, are also located close to a *mouth*, reminding us that—in contrast to plants—such features evolved to support a metabolism that depends on devouring other creatures. They exist to serve the organism in that competitive context.

To the extent that physiology is determined by evolution, so is mind. Brains exist to operate bodies, and minds are concomitants of brains. If the body is redefined, so must be the brain and its mind. To quote Santayana again: "In regard to the original articles of the animal creed — that there is a world, that there is a future, that things sought can be found, and things

---

[118] While we take this arrangement for granted as natural, the very idea of a mind "operating" a body suggests a technological notion of control systems.

[119] For a detailed list of possible varieties of mind, see Kevin Kelly "A Taxonomy of Minds" The Technium (https://kk.org/thetechnium/a-taxonomy-of-m/).

[120] Phillip Ball "Organisms as Agents of Evolution" John Templeton Foundation, April 2023, p236-7.

[121] Ed Yong *An Immense World; how animal senses reveal the hidden world around us* Random House 2022, p332.

[122] John M. Smart in *Cosmos and Culture*: *cultural evolution in a cosmic context*, Steven J. Dick and Mark L. Lupisella (eds) NASA 2009, p212.

seen can be eaten… in one form or another this faith must endure."[123] While this "creed" applies to the forms of life and mind we know, and may reasonably be assumed to apply more broadly, we are also free to ask whether a mind could exist to which it does *not* apply. We can imagine an *ideal* mind and wonder what sort of body would correspond to it. We now have the opportunity to "understand beings without a convenient evolutionary back-story," which are created *de novo* either through genetic or computational engineering.[124]

For humans, at least, thought is inextricable from language. But other creatures communicate without a formal (fully grammatical) language.[125] Besides human beings, dolphins seem to be the only animal that gives itself a name. Dolphins understand human commands, the importance of word order, and the use of symbols to represent concepts and objects that are not present. However, they don't seem to use these abilities (beyond their "signature whistles") to represent objects or concepts in their own communications.[126] In their world, they don't need what we call language.[127]

Sound is effective for communication across distances, and around obstacles, while visual signals are often effective closer up. Human speech barely separates the flow of sounds into discrete words. Could an alien language dispense with such discreteness, using instead a continuous variation of pitch or intensity?[128] Could it dispense with temporal sequence altogether, presenting simultaneous patterns in space, like the aliens in the film *Arrival*? It is often said that mathematics would be a natural basis for communication with intelligent extra-terrestrials (i.e., those that think abstractly enough, and enough like us, to be technologically able to be in contact with us). However, *our* mathematics is a human creation based on experience of environments with both discrete objects (hence integers and counting) and fluid media (hence mathematical continuity). That presumes, at least, a planet with solids and liquids.

How much more *complex* could a mind be than the human mind? Granted that complexity may to some extent be in the eye of the beholder, the notion is related to information content or capacity, and thus to intelligence. Some minds could be far more complex than ours—that is, with a far greater number of possible states. There could therefore exist minds that we could never hope to understand.[129]

The notion of *omniscience* is an ancient human ideal, extrapolated from common experience (for example, being able to see something that another observer cannot see.) As we have defined mind, however, perfect omniscience is not feasible: every embodied mind will be

---

[123] *Scepticism and Animal Faith*, p180.
[124] Levin op cit, p7.
[125] "A language isn't a language if it can't say essentially *anything* that is possible." [Aric Kershenbaum *Why Animals Talk: the new science of animal communication* Penguin 2024, p202] On the other hand, the structure of a language shapes what is conceived as "possible."
[126] Similarly, Grey Parrots can understand and articulate some human speech, but do not use this for their own communication in the wild. Alex, the African Grey Parrot, is the only creature on record to verbally ask a question.
[127] Kershenbaum op cit, p84-85: It's not that "dolphins don't build machines because they don't have a language. Rather, they don't have a language because they don't need to build machines."
[128] Phillip Ball *The Book of Minds: how to understand ourselves and other beings, from animals to AI to aliens.* U. of Chicago Press, 2022, p364. Kershenbaum [op cit, p215] suggests that a language could be based on "varying electrical fields… or even swirling patterns of color…" Electrical fields would have limited range; patterns of color would depend on line-of-sight visibility.
[129] Roman V. Yampolskiy "The Space of Possible Mind Designs" 2015, sec 3. Cf. also Ross Ashby's "law of requisite variety" (that a mind must be at least as complex as what it hopes to understand).

finite and have a (more or less) limited perspective. Yet, that leaves room for degrees of *relative* knowledge superiority. An ideal corresponding to the hypothetical *perfect knower* would be a hypothetical *perfect doer*, an omnipotent agent. Any real embodied agent, however, would have limited abilities, and could not act or create arbitrarily, since it would still depend for its existence on the material universe. Being finite, there would be limits to its powers as well as its knowledge.

Just as there are physical limits on the size of gravity-bound creatures, so there may be limits on the size and functioning of brains, either natural or artificial. Owing to the finite speed of light, an Earth-sized AI brain could have global-scale thoughts only roughly as fast as a human brain.[130] Present-day computers dissipate orders of magnitude more energy per mass than a living cell; energy consumption and heat dissipation may present limits to computational power as we know it. For its purposes, biology is computationally more efficient.[131] Indeed, the technological trend is to imitate biology, not only in terms of organization and function but also for size and efficiency. To replicate in silicon the complexity and connectivity of a human brain would still require something monstrous compared to the size of a human brain. Yet the trend toward miniaturization continues. Some advocate pushing it to the ultimate physical limits of density encountered in black holes. They anticipate that advanced alien civilizations would have done precisely that: instead of expanding into galactic space, they would have gone ultimately small, for the power of computation afforded.[132]

---

[130] Max Tegmark *Life 3.0: being human in the age of artificial intelligence*. Vintage Books, 2017, p153.
[131] Phillip Ball "Organisms as Agents of Evolution" John Templeton Foundation, April 2023, p22.
[132] John M. Smart "The transcension hypothesis: Sufficiently advanced civilizations invariably leave our universe, and implications for METI and SETI" *Acta Astronautica* September 2012, sec 2. This would supposedly explain the "Fermi paradox": why we don't encounter aliens or their signals.

Chapter Nine: Human Alternatives

> "Man appears to be the missing link between anthropoid apes and human beings."—Konrad Lorenz

The world can be experienced as an extension of the self, or the self can be experienced as an integral part of the world. Two corresponding visions of progress offer very different possible futures, ranging from dystopia, where elites control abundance and most people become irrelevant, to a more humane world where resources are shared and people focus on creativity, meaning, moral value, and community. The first is seductive and based on rejecting the human condition in nature, including mortality. Its latest version results from the synergy of capitalism and artificial intelligence—both of which are about mastery and control. This is the dominant vision, where we seem currently headed. Painted in expansive terms of human empowerment and personal liberation, what actually expands is inequality and frantic dependence on technology. An alternative, if recessive, vision proposes acceptance of the human condition and moral more than technological advance. It offers a different vision of the individual—integrated in nature and society—and a different sort of empowerment: the ability to be selective and choose wisely.

A human individual may appear to herself to be an integrated whole, with more or less free will. However, this is not a portrait of the individual that biology endorses. A body is an organism composed of trillions of distinct cells, each of which can be considered an organism in its own right. How such coalitions came about is a topic of evolutionary theory. Obviously, being part of a multi-celled creature must offer sufficient advantages to individual cells. But as with all coalitions, this does not imply the ideal harmonious functioning associated with a machine, whose parts and functioning are well-defined, fixed, and predictable. On the contrary, it suggests a political metaphor: the "society" of cells composing a body is the net result of many competing agents, which overcome their differences well enough to give the appearance of an integrated whole, at least for a time. The conscious self that acts on behalf of this body takes this integrity for granted until something goes wrong that belies the appearance of normal functioning. One then subjectively experiences discomfort or unwellness, even pain. What the doctor or scientist labels disease, the patient experiences as dis-ease.

The organism's defenses rely on discriminating what is "foreign" from what is "self." From the externally-oriented viewpoint, we understandably look for a pathogen (a foreign agent) as the cause of dysfunction. Society does this too, in response to its disorders. In both cases—human society and the society of cells—potential enemies lurk outside the body politick but also within as traitors. In both cases, the integrity of the whole is vulnerable, but also precarious to begin with. If the body were literally a machine, there would be no way to explain dysfunction except as the wearing out of parts or by the intrusion of some foreign matter—like grit or water in a normally well-oiled engine. On the other hand, an organism cannot be switched on and off like a machine; uninterrupted operation is a condition for its existence.[133] It is its own agent, while the only agency involved for the machine is that of its designers and users.

---

[133] Daniel J. Nicholson "Organisms ≠ Machines" Studies in History and Philosophy of Biological and Biomedical Sciences 44 (2013) 669–678, p673.

In fact, the immune response is better explained with the political metaphor. The body normally consists of a "melting pot" of cells, some of which belong to it genetically, while vastly outnumbered by microorganisms that do not. In human civilization, there are always foreigners living within a society's borders. They are only a problem when they break the law or threaten the existing order. But there can be native born traitors well—such as cancer cells. Free societies manage this challenge through due process, while dictatorships may be more ruthless. In either case, there are agents whose job it is to deal with suspects; and in either case there can be mistakes, corruption, failures of duty, even betrayals. In the body, these policing agents are various immune cells and macrophages, which have a degree of autonomy, mobility, and discretionary power not enjoyed by other cell types. Like double agents, they may or may not act in what medical observers presume are the body's interests. All this implies that the integrity and very identity of the body are fundamentally uncertain.

The body might be considered more of a battlefield than a smoothly self-maintaining machine.[134] Health or unhealth is a statistical effect emerging from micro-interactions of cellular combatants. In view of the body's unfathomable complexity, its proper functioning could be considered miraculous, an unreasonable expectation. When in good health, we don't question this functioning, which ideally is transparent. We become aware of it (and complain) when it breaks down. Yet, the body is not designed to function perfectly or to last forever—indeed, not designed at all. It is not a product of rational thought, but something that persists simply because it *can* within the complex interplay of natural reality. The individual is the immediate object of natural selection, via mortality, but it is the germ line that continues and favors kin and (indirectly) the social group. Natural selection promises only the longevity required to reproduce. Certainly, it does not promise freedom from pain or discomfort, which are instead built-in concomitants of the body's ability to self-maintain. The poor subjective self, along for the ride, must experience suffering as part of its appointed role to monitor the state of the organism and help coordinate its maintenance. It can imagine a life without suffering. But, given that pain and suffering are evaluations of the organism's actual condition, the only way to preclude suffering would be to eliminate the very possibility of dysfunction. While that may not be feasible, like the rogue cancer cell, this self can have misguided ideas of its own, including the idea that it could do better than nature as a designer. Nature has led us to a point where we can question her premises and consider alternatives.

Because we seem to ourselves to be agents, as well as experiencing subjects, we naturally see agency everywhere about us, even where it cannot plausibly exist—for example, in nature spirits gods, demons, ghosts, or souls. The mechanist philosophy has tried vainly to counter what it considers superstition, at the cost of ignoring agency in the physical world. This omission has hardly prevented the majority of human beings, past and present, from claiming religious beliefs. Even if religious hopes for a permanent self are delusional, their sincerity and long-standing basis is grounded in a fundamental disaffection with the limitations of life in the body, with a yearning for expanded possibilities of experience and motivation.[135] We are hardly reconciled to the body's sufferings of injury, disease, and finally death—nor to the cruel ways that people sometimes deal with each other. Most religions deny death in favor of some sort of continuation

---

[134] Barbara Ehrenreich *Natural Causes: an epidemic of wellness, the certainty of dying, and killing ourselves to live longer.* Twelve Books (Hachette), 2018, p136.

[135] Thomas K. Metzinger "Why Is Virtual Reality Interesting for Philosophers?" Frontiers in Robotics and AI September 2018 | Volume 5 | Article 101, p14.

of the self or consciousness, conceived as an ideal state of mind, perhaps in a virtual environment called heaven. Some promise resurrection of the perfected literal body, free from suffering. Such a notion can only be based on an incoherent vision of the material body made truly whole and invulnerable because it is effectively *not* a product of nature to begin with. While such hopes are irrational, the force of animal faith behind them is ironically grounded in the realities of the mortal body.

Science may have divorced itself from its religious origins, but not from the ancient pretention to immortality and the hope to overcome evil, both of which imply freedom from the chains of embodiment. On the contrary, the new faith is that, while these goals cannot reasonably be expected through religion or conventional psychological means, they *can* nevertheless be achieved through technology. We once imagined gods as super-agents; now we imagine *ourselves*—or our digital creations—as super-agents, with super-human powers over both the external world and our own being. On a deep level, we aspire to escape the thumb of nature to become fully self-determining, self-creating, self-defining, and even deeply moral, creatures.

We are the creature that seeks to be unnatural, struggling against the world we find, and against our animal nature as we find it, striving to create our own versions. Though hardly gods, we would be so. The coveted power, to define how things shall be, is asserted most effectively through technology, where we use the rules and elements of the found world to shape an environment more to our liking. While we may struggle for moral perfection, the raw nature within us has scarcely changed. Aside from age-old breeding practices, the means to change it deeply have not existed until now. For good reason, perhaps, the *idea* of changing it through technology has been taboo. For, it thrusts upon us collectively the responsibility for what we should be—a task that has heretofore been left to nature, to accident, or to God. Worse, it delegates it to particular agents who may be less than benevolently motivated.

Christian theology, like perhaps all religion, provides a guide to ethical behavior alongside a cosmology to explain what exists: it addresses both what is and what should be. Science deals only with what is. Newton's was "a God of science [whose] kingdom is in the brain rather than the heart… a God of law and certainty, not a God of hope and fear, of punishment and reward."[136] Yet, the other great concern of religion—ethical conduct in regard to others—continues to have expression in religious and secular movements around the world. The current distrust of science may trace, with some justification, to the historical divorce of science from moral judgment that coincided with the founding of the Royal Society.[137] The focus on mental and technological mastery of nature has by far outstripped in consequence any moral progress mankind has made since the 17th century. On the contrary, the amoral use of scientific knowledge—in particular its coupling to capitalism—has exacerbated social problems, serving power rather than wisdom. Instead of applying Christian ethics to trade and economics, for example, mathematical analysis is often applied to maximize profits. Mathematical modeling attempts to capture a system in terms that can be controlled—for whatever (or whosesoever) purposes. The model is an idealization but does not represent an ideal. It simulates—in the sense of closely following something—but does not *lead* as an exemplar to follow. Science can measure the effect of mistakes, but only belatedly tells us (e.g., in issuing statements about global warming) how we should live in order to avoid them.

---

[136] A. Rupert Hall *From Galileo to Newton* Dover, 1963/1981, p304.

[137] The Royal Society was founded in 1660, the same year as the restoration of the monarchy following the English Civil War, which was partly a religious war. The members no doubt sought to avoid political or religious controversy by sticking to matters concerning the natural world.

While religion may have functioned historically to limit antisocial acts of the individual, its promises and incentives were largely focused on personal—not collective—salvation. The ethical codes it prescribes for behavior toward real others in this life are traditionally tied to rewards in an imaginary personal afterlife. It is crucial now to clearly distinguish these motivations, and to sift what is feasible from what is not. Personal spiritual salvation has never been a realistic goal, given that the individual self is no more than a virtual agent for the material body, which dies. In contrast, collective salvation has always been feasible, because society carries on after the individual disappears. But it has also always been elusive because of contradictions within our primate nature, as simultaneously cooperative and competitive beings.

Progress for the individual, or for a given society or nation, should be distinguished from social progress overall for the species. The very idea of progress is relatively recent, originating in 17th-century Europe. The scientific revolution endorsed the idea that humanity could determine its own future, especially by controlling nature through technology. It reprised the idea of Man's original dominion over the Creation (before the Fall), and made sensible the possibility of moral redemption through human effort. The religious framework had been fatalistic, with the future determined by divine will. Medieval Christian Europe could only look forward to the imminent end of the world—the prophesied doomsday—a context in which it made little sense to try to improve things in this life.

We now recognize that the natural order exists as a symbiotic balance of competing species that is hard on individuals. We see that messing with this balance is risky, yet we imagine we could improve on the randomness of evolution. As natural creatures, we could continue to go the way of nature, dictated by forces beyond our control, resigned to mortality and the fate of inevitable extinction. As unnatural creatures, however, we imagine mastering the forces that control us, turning the tables on nature. We can even imagine ourselves becoming the benevolent and moral beings idolized in religion. Along with the quest for power, we imagine the possibility of *eutopia*, a world and a body expressly designed for our well-being and happiness, heaven on earth.[138] Since neither God nor nature can be counted upon to provide such a world in this life, it seems up to us to create it. Indeed, this appears to be the long-term human project, despite the handicap of being nature's limited puppets and despite the disastrous setbacks we know as history.

We are on the threshold of the sort of genetic manipulation that can redefine the human organism, promising not only enhanced abilities but even moral progress. Alternatively, digital technology promises cognitive evolution, and may seem to offer even immortality in a disembodied virtual existence. Converging, such technologies confront us with a long-standing tension in human concerns between individual and collective, between intellectual and social progress, between head and heart. They reflect ancient ideals that have heretofore been pursued, with limited effect, through slow traditional means involving "spiritual" rather than technological growth. Technological progress has accelerated in a way that traditional moral progress could not. It is unclear whether such a shortcut can succeed, or simply end in self-destruction. There is a contradiction in seeking moral perfection though means that are morally flawed.

We think of individualism as a political ideology, but it is foremost economic. The development of technology, and therefore our notion of progress, cannot be separated from modern capitalism.

---

[138] The word *utopia*, as normally spelled, dismisses its own possibility, since it means literally "no place."

It is corporations that develop technology to scale, for the purpose of generating profit in the consumer marketplace. The more consumers, the better. So, family or household cannot be the economic unit with purchasing power; it must be the individual, including the child.

The development of capitalism can be traced to the shift in mentality, in Europe during the early scientific revolution, that underlies our current view of progress. The new mentality holds that nature is a machine; that humans are separate from nature and are essentially selfish and competitive; that the Earth, along with its peoples, are therefore resources to exploit.[139] This attitude informs the policy of resource extraction, on the one hand, and human exploitation on the other, upon which global capitalism was founded. As a result, today, 69 of the 100 biggest economies in the world are those of transnational corporations, not nation states.[140] This attitude toward nature and other human beings has by now literally colonized the entire planet. One wonders what alternative course might peoples have pursued on their own, if they had not been colonized by Europe? It is assumed that progress must be material and technological.

In earlier times, a prime motivation for war was to plunder existing wealth and to capture slaves to do the work involved in generating wealth. A prime motivation for capitalism is to use capital (property) to generate wealth—by having your *money*, as well as others, work for you. A prime motivation for speculation is to capture a greater *share* of existing wealth, not through physical plunder but through manipulation of economic abstractions. A prime motivation for automation—and now AI—is to reduce the cost of labor and the dependency of corporations upon it.[141] More broadly, the hope to have *machines* work for us reflects the desire to avoid onerous labor, freeing us to play. Yet, play, including exercise and creativity, involves effort. We are at risk of a tendency to avoid effort of any sort, whether physical or mental, in the name of convenience, which the commercial development of AI is pleased to encourage.

From the beginning, capitalism has been a sort of pump to siphon a disproportionate amount of wealth from human productivity. This inevitably leads to wealth disparity, which is its purpose. Modern capitalism has refined this "pump" into a complex global institution that bends governments to its purpose. It has led to the greatest inequality of wealth in history, so that a dozen individuals are said to possess more wealth than half of humanity combined (four billion people).

But what exactly is that "wealth"? In this day, it's "money," which is extremely symbolic and intangible. Some nations no longer use currency, let alone gold coins. Digital money operates entirely on trust in the system—trust that the unit of value entitles you to exchange it for something tangible you want. It also depends on the power grid and the internet: your wealth is no more than data stored in a computer memory. But even gold coins, banknotes, and physical property-deeds are only valuable because people agree to their symbolic worth. The system enables the exchange of goods and services, but also enables various forms of swindling—including what is politely called profit, but which used to be called usury. As long as the development of technology serves profit, general human benefit will never come first, but will always be incidental at best—if, indeed, there is a net benefit at all, considering ecological and social costs.

---

[139] Jeremy Lent "Envisioning an Ecological Civilization in Theory and Practice" January 2026 https://www.youtube.com/watch?v=Ep8ezlTNjtc

[140] Lent, ibid.

[141] Harry Halpin "Artificial intelligence versus collective intelligence" AI & SOCIETY, 2025, (Sec 4) https://doi.org/10.1007/s00146-025-02240-x: "Thus AI has been less a research programme than an ideology based upon a conception of individual intelligence that served as the self-same basis of free-market capitalism…"

Capitalism is a work in progress. Like scientific thought, it tends toward increasingly abstract forms. Yet, the basic premise remains the same: to capture "wealth," in the sense of what others are willing to pay for or trade. But wealth, in this abstract and disembodied sense (market value or market domination) is not the same as wealth in the form of tangible goods and services, which relate ultimately to the bodily needs of consumers: what helps people to live better. While promising a better pie, the motive is to get a bigger share of that pie. This reflects the tension in human nature between cooperation and competition—altruism versus greed.

AI optimists promise a world in which humans are liberated from drudgery to pursue higher aims. In practice, that means unemployment for many, who must scramble to find new jobs. As always, the labor that is saved by labor-saving technology is first of all the cost to business of paying human workers. This is partly passed on to consumers through cheaper goods and services; but the consumers must have employment to purchase them.

So far, AI development is oriented toward the conventional aims of commerce and shopping—hardly "higher aims." The development of AI tools targets individual consumers (e.g., personal assistants) or companies (e.g., management tools) to gain a competitive edge in an accelerating business arms race. Despite much hype about a glorious AI future, can such goals preserve the world for the next five years, let alone for the next 5000? Can AI even deliver on the promise to free up time, or does it simply fill time by capturing attention—again, largely for commercial purposes?

Whose time does it liberate? So-called "autonomous" AI depends on an economy of hidden human labor,[142] often off-shore and poorly paid, used to develop, test, and correct AI models, which partly accounts for the high cost of AI development. Such work is no more than drudgery, and imposes a form of neo-colonialism that standardizes labor and shapes social relations in diverse parts of the world. Is this moral defect merely a temporary stepping stone toward true automation, or will such low-level human input be inevitable even in the long run, as built into capitalism as it was into colonialism?

The new tools of generative AI, with their vast infrastructure of data centers, amount to a new public utility that is unregulated and privately owned by a handful of companies.[143] The development of these tools is an ongoing interactive process involving the participation of users. This involves collecting ever more "personal data," which amounts to a new form of surveillance —another way in which the role of the state is being taken over by corporate powers.

The exploitative attitude of capitalism is directly expressed through the power of technology. They are intimately related, because technology is not just machines, but a logic of manipulation for gain.[144] Machines are not just devices standing outside the natural order, but *affordances* of the human creature within the natural order.[145] Under capitalism, their uses are not neutral but

---

[142] "Such systems rely on vast networks of human workers, often invisible to end users, who annotate, correct, monitor, and sustain the functions that AI claims to automate." [David Nemer and André Sobral "Artificial intelligence as heteromation: the human infrastructure behind the machine" 2025, sec 1. AI & SOCIETY https://doi.org/10.1007/s00146-025-02664-5] Human work disguised as computation ironically inverts the original meaning of "computer," which was a human person, usually female. A form of neo-colonialism that standardizes diverse off-shored labor.

[143] Armando Vieira "Intelligence, Consciousness, and the Philosophical Challenge of Large Language Models" January, 2026. https://philpapers.org/archive/VIEICA-3.pdf.

[144] Erik Davis TechGnosis: myth, magic, and mysticism in the age of information, 1998/2004, Five Star, p173.

[145] 'Affordance' is the opportunity for action an environment offers a creature.

always economic and therefore ultimately political. AI, especially, is not merely a tool to use at arm's length, but promises (or threatens) to transform the tool user.[146] Transhumanism proposes to use technology to upgrade human capabilities and augment the human form itself. The cyborg concept is plausible because brains are already naturally able to adapt to body alterations. Posthumanism proposes to replace the human form altogether. A basic paradox is in play, because these projects for the future stem from present values but involve transcending the biological basis on which those values rest. We may hope to "align" AI with human values, but which ones? We cannot predict post-human values, nor even align the present values of people with each other.

This does not prevent Silicon Valley from spewing endless hype about AI, alternately generating hope and hysteria about an AI takeover, sometimes called the Singularity. We are presented with a vision—both glowing and fearful—of the Age of AI or the Fourth Industrial Revolution. Much of the research (and the hype) concerns the quest for Artificial General Intelligence (AGI), which means human-level capability and beyond. It means trying to create an agent that replaces the human being and could be vastly superior to it. This is presented as an inevitable development, almost as an extension of natural evolution. In truth, it is a commercial ploy that a few individuals are pursuing for their profit. It's a product of the venture capitalism headquartered in Silicon Valley.

Conventional capitalism earns returns from producing and selling actual goods and services in real time. Venture capitalism is a form of speculation that tries to anticipate and dominate *future* markets. It resembles "disaster capitalism" insofar as it depends on crisis, but differs insofar as it doesn't wait for disasters (such as earthquakes or wars) to create new needs in the short-term. Instead, it bets on foreseeing longer-term future needs (in this case, widespread irreversible dependency on AI). It tries to *create* such needs, which stem from the atmosphere of uncertainty surrounding new technology and social crisis resulting from it.

Venture capitalism is itself a fallout from a recent social crisis—the 2008/2009 financial meltdown. While the instability of the system leading to this crisis was not resolved, cheap capital continued to be abundant. At the same time, the means to use it in new ways emerged through the internet, mobile phones, and new services (such as Amazon and Facebook) that penetrated everyday life. The more dependent we are on automation and technological infrastructure in general (not just AI), the more vulnerable to disruption society becomes. (Think of power grid failure, mass hacking of digital accounts, oil shortages and flight cancellations.)

Most business meets stable predictable needs (like food, shelter, clothing, transportation) by skimming relatively modest regular profit from actual production. Venture capitalism speculates on uncertainty, fear, disruption and crisis, in hope of enormous *future* profit. In order to succeed, it must anticipate future needs. The best way to do that is to *create* them. But such artificial needs are no more than dependencies, like drug addiction. This is exactly the strategy now pursued in the development of AI. Through constant hype, a long-term future dependency is being established, both for business and for the ordinary consumer. Both employers and employees will "need" the edge provided by the latest AI tools, in an endless competitive arms race: the rat race on steroids. This arrangement is supposed to continue until the whole system explodes from its own contradictions and pressures—the real singularity. This is why billions of

---

[146] Already it erodes human "epistemic sovereignty" as the capacity to author knowledge rather than merely retrieve it. Hans Westerbeek "How AI is rewiring the human brain: the generational transformation of cognition and knowing" AI & SOCIETY https://doi.org/10.1007/s00146-026-02912-2, Feb 2026.

dollars are invested in developing generative AI tools, the data centers to run them, and the power grid to run the data centers. It is part of a general drive to automate everything, regardless of consequence.

But *why* automate? Aside from novelty, the promised consumer benefit is most often convenience. Modern automobiles, for example, are indeed far more convenient than the Model T Ford—which was itself a paradigm result of automated production. But we pay for that convenience in several ways. First, the adjusted cost of the modern equivalent of the 1920s Model T is about six times greater. But it was also feasible for an individual to service and even *build* a Model T oneself. Indeed, before pollution controls, many people enjoyed weekends tinkering with rebuilding their car engines. Cars today depend on technology beyond the individual's grasp. Modern vehicles require professional servicing: that is, maintaining your car has become *monetized.* They are more dependable when new, but more costly, not built to last, and not repairable by the owner. More complicated, there are more things that can go wrong with them. But cars are merely an example; many modern appliances in our throw-away culture follow the same pattern: built-in obsolescence, vulnerable high-tech (often digital) parts, and non-reparability. In general, we pay for monetized services that once were open to individuals to perform themselves. Thus, we become more dependent on factors beyond our control.

This points to a very general principle: while from the consumer's point of view, the purpose of automation is convenience, from the producer's point of view, it is monetization. The goal is to convert things that people take for granted (like air and water)—or actions they once were used to performing for themselves (like cooking)—into commodities or services they are willing (and finally obliged) to pay for. This involves a campaign to "re-educate" the public. It is also a self-reinforcing process: as more people become used to paying for a service, it becomes a norm they cannot do without. (Think of your mobile phone, your email, and online banking.) This amounts to promoting addiction and then supplying the craved drug. The services are "convenient," but they are neither free nor liberating. They purport to save labor but actually re-direct it. They create a dependency in which consumers are at the mercy of a handful of mega-businesses. While there may be some competition among these giants, the dependency itself is a seemingly irreversible fact from which they benefit far more than society. It is not empowering to be dumbed down by a technology that is embraced to do your thinking for you.

Control and automation are distinct concepts—in fact, opposites. Control requires attention, in a constant feedback loop between the controlling agent and the object controlled—what we call skill. The point of automation is to reduce the need for (human) control or skill. This shift requires two things: to create a system that takes the place of the agent (e.g., an algorithm); secondly, to redefine the object to control as a system the algorithm can manipulate. Ultimately, this means redefining, in mechanistic terms, the natural world and everything in it—including human beings. This is precisely the idea behind AGI: that human intelligence (and even consciousness) can be captured in algorithms. These can then be autonomous, improving themselves indefinitely, rendering them potentially far smarter and faster than humans. AGI is also the ultimate service to monetize: it promises you a slave to do everything for you that you find the least bit onerous—including thinking. In short, the companies promoting AGI promise offer a life of ever greater ease for an ever-greater price. They don't mention the irony that you will have to work ever harder to pay your AI bill! Nor do they mention the other glaring contradiction in the notion of a super-intelligent, super-capable slave: controlling something more powerful and intelligent than you is just not feasible!

Whatever else it is, and whether realistic or not, AGI is a dangerous ruse to make money by playing on human foibles. The hype around it appeals to ancient fascination with matching and outdoing nature—including the possibility to create artificial mind. (With eight billion natural minds to interact with—with whom we can hardly get along—why would we an additional population of artificial minds to contend with?) It promises a utopian world of leisure and enhanced human capabilities, while the actual motivation is pillaging your coffers. AGI is being promoted as necessary, inevitable, and ultimately desirable simply in order to generate profits for a few individuals, hardly for the benefit of humanity. Its dangers would far outweigh advantages. Even its promoters admit the dangers, in fact vaunting them in order to create the very atmosphere of crisis in which people will "need" AI solutions to the problems AI creates!

In truth, we do not *need* a super-intelligent agent to replace us as tool users. We do not need a *universal* labour-saving tool. What we *do* need—aside from exercise and proper nutrition—is specific tools to meet specific needs (for example, Alpha-Fold). Such limited AI tools remain under human control by definition. They don't need to be trained on the internet (the human data base at large) but can be developed from limited topic-related data sets that do not require the enormous energy infrastructure required for AGI. Nor do they require the enormous development costs seen for generative AI. These include massive low-paid and degrading human labour to "check" content coming from the internet—a consequence which is the very opposite of the ideal behind automation! The big players are nevertheless going ahead full steam creating new mega data centers in pursuit of AGI on the assumption that we will buy the hype—and their products—succumbing to the seductions of convenience. This is capitalism at its worst. The only net beneficiaries will be billionaires.

Certain psychological traits are highly compatible with capitalism and support it. These include individualism, analytical (as opposed to holistic) thinking, impersonal trust of strangers and formal institutions, an objective concept of fairness, and respect for rules and contracts. Such traits, which seem progressive, now go under the acronym W.E.I.R.D (Western, educated, industrialized, rich, democratic) precisely to emphasize that they do *not* represent humanity at large.[147]

Capitalism depends on trust of strangers in the marketplace, enforcement of contracts, mobile labor, avaricious incentive to innovate, and a merit-based system as opposed to relation-based (what you know, not who you know). WEIRD traits predate modern capitalism and set the stage for it. The medieval Church, for example, discouraged extended kinship relationships in favour of the nuclear family. Social movements sought individualized rights. There were guilds before corporations or unions. Rent and usury (the concept of "interest" to be paid on loans), as well as slavery and conquest, long predated capitalism—as did cooperation among strangers, rule-based governance, trade based on contract, bureaucracies based on merit, and global trade networks.

In turn, capitalism reinforces certain traits. Markets reward individual initiative and cater to individualized consumerism. Nuclear-family households imply more consumers. Formal education promotes reason, abstraction, and systematization, which lends itself to bureaucracy and rule-based behavior, and impersonal social relationships. Economic mobility further weakens community and kinship structure. In part, urbanization is a result of the industrialization

---

[147] Joseph Henrich *The WEIRDest People in the World: how the West became psychologically peculiar and particularly prosperous* Ferrar, Straus and Giroux, 2020.

that capital facilitates; but its very scale also promotes the values behind capitalism, including individual consumerism and impersonal relationships.

However, this does not mean that capitalism must be identified with the WEIRD mentality. Rather, it is a framework to which other societies adapt without necessarily forfeiting all of their traditional values. This adaptability helps to explain the global success of capitalism. For instance, contracts may coexist with relationship-based trust and moral principles. Laws may coexist with traditional or religious norms and obligations. Businesses may rely on family or clan networks alongside formal structures. Yet, it is no coincidence that capitalism, the scientific revolution, and colonial power emerged together in the Early Modern period in Europe, where nations were competing fiercely for dominance. Capitalism required accumulating resources, then recycling profit into further investment. Science enabled technological innovation, such as for mining and manufacturing. Together they boosted industrial production, resulting in military advantage—which helped further spread capitalism and the ideology of progress, as subjugated peoples were forced to adopt European ways. A synergy of institutions, cultural traits (such as the Protestant work ethic), technological innovation, geography, and historical accident enabled certain societies to gain global ascendency over others and to export their systems—if not their mentality—worldwide. On the other hand, those societies colonized by the WEIRD mentality could, to some extent, choose which features of that mentality to embrace—for example, weapons technology, cell phones, and global trade—without internalizing it wholesale or abandoning the essence of their own culture.

Perhaps the ability to resist the WEIRD hegemony depends on the ability to be *selective*. We live in a world chosen by others, past and present. But we can choose which features of it to embrace now and which to reject. While Christianity evolved into a patriarchal religion that supported WEIRD values, Jesus himself didn't embody this pattern. Here is how he expressed the ability to be selective and thus preserve certain values in an atmosphere hostile to them: "Render unto Caesar what is Caesar's and unto God what is God's."

We can do this too, which requires discriminating the wheat from the chaff, the baby from the bathwater. It requires *thought* and *will*, while the powers that rule us count on our blanket mindlessness and placidity. We don't necessarily have to break laws in order to resist the values behind them. (Why wag the dog from the wrong end?) We do not have to embrace technologies wholesale, nor the ideologies behind them. But that means seeing to the bone of things and being clear what we do stand for and actually want. It's my hope that this discrimination is happening within indigenous cultures trying to preserve traditional ways. I hope it's in the heart of feminist critique. I look forward, not to the Singularity, but to a consumer rebellion against the domination of production by a few corporate giants for their private benefit rather than human welfare. We are living under "foreign" occupation by the WEIRD mentality in its latest embodiment: global capitalism. This mentality is not native to human being, nor necessary or inevitable, but a poor choice we can reverse by deliberate personal increments.

Continuity is both personally and collectively valued. Traditionally, value is tied to scarcity—i.e., to a competition for resources that satisfy biological needs. But on a planet teeming with life, phenomenality should not be presumed scarce. In spite of the market value of consumer

attention, the consciousness of an individual human being is the ground of their existence, not a commodity to quantify or trade in a marketplace. It's worth examining why we value it.[148]

The prospect of a post-biological human future might horrify some, if it suggests that only non-human animals would be left on the planet to "feel the spark of insight, the pangs of grief, or the warm hues of a sunrise."[149] But post-biology should not be associated with disembodiment or with the absence of phenomenality, any more than phenomenality should be identified exclusively with biology. Furthermore, what exactly do we fear would be lost in migrating from biology to some other form of embodiment? For that matter, what do we fear is lost to us in death? The world will persist, with other bodies to perceive it. Could *anything* be lost if the functionality associated with consciousness were perfectly preserved?

To be sure, we may fear personally to lose the consciousness with which we identify our life—another version of the fear of death or not-being. But what does phenomenality *itself* mean to us that we value it so? If my consciousness is simply this body's way to monitor its situation in world, then my concern about its continuance is no more than concern about the continuance of this body itself. Beyond the reasonable measures it must take to maintain its existence, does it make sense for this body to cling desperately to its own existence? Or does that impulse simply reflect the animal faith that was the condition for existing in the first place?

The desire for personal continuity drives some people to believe they can exist after death, disembodied in a virtual reality. Real embodied subjects can have virtual experiences; but virtual subjects have no sort of experience at all! We can experience a virtual reality presented to us via a computer program, but we cannot *become* a computer program. A computer program can animate a robot. It could *imitate* a living person, but it could not *be* that person or the continuation of that person's consciousness. If it could be conscious—as an embodied artificial organism—its consciousness would be its own, as a separate being with its own identity. Even if it were feasible, "uploading" a mind to cyberspace would mean converting it to a digital file that must be maintained—and could be manipulated, copied, or destroyed—by real physical agents. That vulnerability would gainsay the ideals of permanence, self-determination, and invulnerability that motivate the move to cyberspace, and invalidate the very concept of individual identity. If there were clones or back-up copies of you, which would be the real you? Why would it matter? The idea of immortality through digitation is only coherent if "you" are a program that can be abstracted from your mortal body. But, then, "you" could be duplicated, deflating your sense of identity and confusing others.

The notion of whole brain emulation assumes—falsely—that the brain is a piece of hardware (like a computer), and that the mind is a piece of software (like a computer program).[150] The persistence of this assumption gives rise to transhumanist fantasies such as living in a simulation, copying minds, uploading one's consciousness to cyberspace, or downloading it into alternative bodies—as though the software is separable from the hardware.[151] It gives rise to

---

[148] And why we *don't* value it. More people die from suicide annually than from violent crimes and war. Max Bennett *A Brief History of Intelligence: evolution, AI, and the five breakthroughs that made our brains* Mariner Books (HarperCollins) 2023, p69.

[149] Susan Schneider "Superintelligent AI and the Postbiological Cosmos Approach" preprint 2016.

[150] An emulation is a simulation of another simulation or artifact (software or hardware), both of which are products of definition. The reasoning is that if a true analog of the human mind could "run" on a true analog of the human brain, it too would be conscious. Whether or not the conclusion is valid, the premises are not.

[151] By definition, software is separable from hardware in *artifacts*, in the sense that it can "run" on more than one physical machine. But the body/brain is not an artifact, and its "software" may not be separable from its wetware.

premature deliberations on the moral standing and ethical treatment of AI entities. The personal hope for digital immortality is an updated version of the hope for a spiritual immortality. The vanity of these hopes to preserve "self" does not mean that artificial mind cannot exist. Like natural mind, it would be physically embodied, not merely virtual. It would not be a copy of anyone.

AI, pharmaceuticals, and genetics represent areas where "pure" research is being taken over by commercial concerns and an engineering mentality—in other words, by consumer capitalism, with its focus on individual satisfaction rather than collective well-being. Consumer capitalism depends on demand, which it actually manufactures by channelling biological need into concocted want. As an ideology, capitalism is unifying because of the quantification of value in money, and because it defines the common goal of life for everyone as getting money. But in every other way it is divisive, pitting individual against individual, against the collective, and against nature.

Monetary value displaces traditional values; identity as a consumer displaces other forms of identity, whether ethnic, religious, political, tribal, or professional. This modern arrangement poses problems both for the individual and for the whole of society—such as scarcity in the midst of overproduction, compulsive consumption, debt, anxiety, "spiritual" impoverishment, alienation from others, inequality, ecological collapse and climate crisis, intergenerational injustice, dissention, social unrest, and disintegration of common ground for cooperatively addressing such concerns.

An alternative ideology *could* aim for collective well-being rather than private profit or personal advantage. It could reward cooperation and sustainability instead of greed and overconsumption. It could value sufficiency (enough) rather than endless growth (more). It could shift from GDP to other indicators such as "happiness." It could focus on quality of life (and quality of goods) as opposed to quantity; on creative contribution rather than passive consumption. Innovation and production could be for the sake of genuine collective advance rather than for personal gain or mere survival—and certainly not for the advance of an elite few. A guaranteed living could be provided to all, with common ownership and control of the means of production, rather than private. The motive of getting ahead personally might then be displaced by the altruistic desire to advance the state of humanity—swords into ploughshares.

How would such a shift affect our relationship to consciousness and life? In consumer society, the experience (of others) is implicitly viewed on an economic model, in limited supply as a tradeable good. We know that personally we each have limited time to live, to experience life, which should be valued accordingly. But why should that be conceived as scarcity? We are obliged to sell our labor (our time) in exchange for money in order to live—so as to continue consuming. Time for personal fulfillment seems scarce when we are caught on an accelerating a treadmill. That pressure to survive, which powers the treadmill, could be relieved by a system that guarantees the necessities of life, providing universal social security. But the underlying shift would be in motivation, concerning the *purpose* of being alive and conscious: what to do with that security. Is it to serve the individual or the collective? The next five years or the next five thousand years?

In a post-producer/post-consumer society, machines would produce everything including goods, services, art, science, and innovation. A theory of value based on human labor would no longer be relevant. There would be no further need for consumers or a marketplace if robotically

produced goods and services were free for all, without market value. There would be no need for money, credit, debt, or a marketplace either. (The only markets would be black.) If AI creates universal abundance, the only scarcity would be access to it. Therefore, political power would mean ownership or control of access, production, and distribution.

In a dystopian society where AI produce*s* everything but elites own it, economic value shifts to a struggle for access, or for symbolic status and political power that provide access. Power would concern who controls AI, and how its fruits are to be distributed. Even the elite owners of production would not depend on profits generated from markets to enhance their consumer power, since machines could directly fulfill their every material want. Without redistribution, the masses risk becoming economically irrelevant. If humans are no longer necessary as either producers or consumers (let alone as voters), the masses become superfluous except as a class to dominate, while the elite would vie with each other for status and dominance. The possibility to avoid such a dystopia presumes that human motivations will have shifted significantly, away from private gain and status toward collective progress.

In a eutopian scenario, in contrast, the abundance produced by AI would be a *commons*, communally owned by humanity. Society could be organized on the model of higher organisms —with individuals playing the supporting role of cells. Basic needs would be de-commodified. There would be a universal basic income (blood supply), providing access to AI-produced goods and necessary human services. (There could persist niche markets for goods made by human hand and for specialty human services.) Human productivity would shift generally from material goods to intangibles like creativity, meaningful experiences, and play—essentially to share rather than to sell. *Value* could shift toward meaningfulness, creativity, purpose, relationship and play. Social recognition would shift from accumulation as status symbol (how much one is envied) to what and how much one *does* for others or contributes to the whole. Moral improvement would supersede material improvement. Education would focus on collective needs, global problems and solutions, rather than skills for individual competitive edge (business school).

Failing this ideal, enough access to wealth could be provided to at least prevent revolt by the masses, resulting in a two-tier society—perhaps a neo-feudal order—much like what already exists. The intellectual or creative worth of human efforts is already measured against AI output. In the shorter term, people will have to dig ever deeper to compete with AI that is designed to compete with them. The capitalist system promises that the massive benefits accruing to the elite will continue to trickle down to the bottom layers. While that might be true to some extent, it begs the question of why they should "trickle" down and at all instead of being immediately shared equally.

Differential "remuneration" is based on the idea that some effort is worth more in the marketplace than other effort. But the marketplace is what is in question. For example, the CEO of a major corporation might receive, in all, a thousand times more benefits as valued in dollars (including salary, shares, severance, etc.) than a lowly assembly line worker. Does the CEO's labor truly produce a thousand times more benefit to humanity than the worker? Or might they be regarded as members of a team, perhaps more interchangeable than their pay suggests, and entitled to more equalized remuneration?

In the longer term, humans (and perhaps biological life, generally) could become marginal or obsolete, supplanted by machine intelligence. What would constitute value for such post-human beings? They might inherit some human values, such as "optimization," that had been programmed-in by their capitalist fathers. But what values might they contrive or evolve on their own? In their world, humans might be preserved—perhaps venerated like ancestors, kept as

zoo specimens, or used as a resource (as in *The Matrix*). On the other hand, they might be considered vermin, or recreational game to be hunted. To be dominated by artificial overlords instead of human ones would be little consolation to the remnants of humanity.

The technological doomsday that currently looms over the world is fueled by the exponential development of AI—with the electricity consumption it will require and the acceleration of business competition it will enable—foreshadowing an impending crisis. This plays out in various scenarios in which society could destroy itself through runaway technology—whether nuclear, genetic, environmental, or just through AI itself. The dilemma resembles also the Mutual Assured Destruction strategy of the Cold War, or the logic of keeping peace by waging war. The very people who develop AI warn of its dangers, while continuing frenetically to develop it, because whoever gets there first will dominate the market—and possibly the world. The "deadline" floats, deferred from year to year as business not only carries on, but manically expands its pace. Yet, no one foresees a reprieve. Being our most important product, "progress" does not stop, but accelerates—and with it the threat of doomsday. Indeed, it *cannot* stop, since only "growth" allows capitalism to succeed and still provide benefits to the exploited. Unlike our medieval counterparts, AI optimists seem undaunted, somehow driven to make hay while the sun still shines.

Chapter Ten: Artificial Mind

> "We have recreated ourselves in the image of our tools."[152]

Artificial intelligence (AI) is rapidly changing our human world. It is itself changing so rapidly that we cannot be sure at a given moment just what the term *AI* means. One thing we can be sure of is that it blurs the distinction between natural and artificial. Indeed, it bridges this gulf intentionally, to fulfill a long-standing human goal to remake nature to human taste. With our implicit faith in the mechanist metaphor, it might be assumed that anything that can arise naturally can be duplicated artificially, so that there is no essential distinction between natural and artificial mind or intelligence. However, language allows us to assert equivalence where it may not really exist. The question hinges on just how closely an artificial brain, for example, must resemble a natural brain in order to replicate the properties we associate with mind—which may or may not include consciousness.[153]

As AI emerges, fundamental questions arise with it. Because they are urgent for human safety, these should logically be answered in advance, on the precautionary principle. AI is potentially more plastic than biologically-determined human psychology and bounded human reason and perception. Its potential is literally unimaginable. Human *will* is only now beginning to be in a position to determine what humanity shall become, what it actually can control, and for whose benefit. The quest should not be to develop AI unrestrictedly, toward some confused ideal, without first clarifying our intentions and refining human will itself. What do we want AI *for*?

Alan Turing began his famous paper of 1950 with the question: 'Can machines think?' He immediately pointed out that we would have to first define what we mean by 'machine' and by 'thinking.' Similarly, to consider artificial mind, we must clarify what we mean by *mind*, a notoriously nebulous term. Without formally defining it, let us first consider mind from a third-person point of view, in terms of observable traits and behavior, and only then from a first-person point of view, as phenomenality. From there, we will look at the prospect of artificial mind in greater detail.

With ancient roots, the mechanist philosophy that arose in 17th-century Europe treats the world as a machine. Hence, Newton's expression "The System of the World," which lays out the basic

---

[152] Robert W. Clowes, Klaus Gärtner, and Inês Hipólito "The Mind Technology Problem and the Deep History of Mind Design" in *The Mind-Technology Problem: Investigating Minds, Selves and 21st Century Artefacts* Robert W. Clowes et al (eds) Springer 2021, p15. The authors add: "We are thus both natural beings and also in a certain sense, self-constructed, i.e., we are not just Homo Faber, man the maker, but human beings the self makers."

[153] The notion of uploading or downloading 'mind' trades on ambiguous language. For example: "The *OpenWorm* project has successfully uploaded a worm (*C elegans*) and downloaded it to a Lego robot, which behaved like a worm." [S. Schneider, P. Mandik "How philosophy of mind can shape the future" in *Philosophy of mind in the twentieth and twenty-first centuries,* Routledge 2018, p310] To deconstruct this claim: 'uploading' means that some algorithm has been formulated based on a theoretical model of the nervous system of this creature. This algorithm (a human artifact) is then installed in the robot (a human artifact). The fact that the robot then "behaved like a worm" seems to confirm the validity of the model. That is misleading, however, to the extent that "behavior" is in the eye of the human observer. Is the behavior of a baseball pitching machine "the same" as that of the human pitcher?

laws of dynamics as an axiomatic system in the style of Euclid. It holds that organisms too are machines, with parts, or can be treated like them. This is empowering because it enables articulating natural things in such a way that they can be artificially engineered. But such analysis of natural things, whether living or inert, can never be exhaustive—in contrast to machines, which in principle *can* be exhaustively specified because they are finite products of definition to begin with.

A corollary of the mechanist philosophy is that function, structure, and organization can be considered independent of their substrate. However, the idea of function (or structure or organization) is a result of limited human analysis and imposed categories. (For example, on a gross level, an airplane performs the "same" function as a bird, since they both "fly.") In computer terms, the idea is that digital software is separate from hardware, which is the basis of the programmable machine (the computer) and of functionalism. But in living things, "software" is analog and not separate from infrastructure. The functionalist idea is that if an artifact behaves like the real thing, then you may as well treat them alike. This is the idea behind the Turing Test and the mistaken belief that Large Language Models (chatbots) could be conscious.

The blurring of the line between natural and artificial, which began with the mechanistic metaphor, invites re-examination of both categories. While it is true that meaning exists for humans and not (currently) for machines, the possibility of artificial mind may hinge less on the medium—such as carbon versus silicon—than on the fact that natural minds are organisms—which means organized a certain way. They are embodied products of natural selection, part of an ecology. The question then becomes: is synthetic *organism* possible, and how can it come to exist? Indeed, why should we want it to exist? In addition to its wisdom as a goal, we should question what that synthesis entails. At what point does synthetic become natural? Is there an identifiable threshold or is it a matter of arbitrary definition? Can a simulation truly *be* an organism? In particular, could the relationship of embodiment be artificially produced or induced —say, by simulating natural selection? [154] [6] By design, machines are becoming ever more complex, imitating the human organism and other creatures. Could a machine build and repair itself like a natural organism does? If it could, would that mean that it *is* an organism?

To approach such questions, let us further unpack the concept of *organism*. For a system to be an organism, it must satisfy at least these three conditions. First, it must be *physical*, not merely digital—which excludes all forms of software. Secondly, it must be *autopoietic*—that is, a self-maintaining agent with its own agenda. Thirdly, it must be *embodied*, which entails a relationship of dependency upon an environment. There could be further requirements—for example, that it must it be organized like living things we know, whose constituting units (cells) are also self-maintaining organisms in their own right. Must it be a product of natural selection, with an evolutionary history, or could its software be developed *in silico* and then be downloaded to a robot body?

---

[154] According to a recent paper, "being causally connected with an external world is neither necessary nor sufficient for cognition. What we require is that the agent has sensorimotor representations that it *treats* as having arisen externally and that it tries to explain with a model of that external world… We leave open whether such simulated minds constitute *synthetic* minds, or are *mere* simulations, analogous to simulated hurricanes, to be used in theorizing." [Iris Oved, James Pustejovsky, Nikhil Krishnaswamy, Joshua Hartshorne "Computational Thought Experiments for a More Rigorous Philosophy and Science of the Mind" In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), Proceedings of the 46th Annual Conference of the Cognitive Science Society, 2024] The authors choose to "leave open" the question, an issue the AI community cannot afford to ignore. More generally, the question is: what exactly makes a natural process and a simulation of it different?

Autopoietic means self-creating, self-maintaining, and (in the case of life we know) self-reproducing. It also means *self-defining*, to distinguish the autonomous organism's own point of view from that of an observer. While to the outside observer, the organism is a physical system open to an environment, in its own terms it need not have a concept of environment. It need only ensure that its inputs remain within tolerable limits (homeostasis), by acting upon (what the observer perceives to be) its environment and its own chemistry in such a way as to maintain constancy. It need not have an idea of itself nor of an outside world—features that are functionally additional.

Embodiment means more than physical instantiation. It involves a relationship of dependency upon an environment, in which survival is at stake. For that reason, the world and its own state matter to the system. It acts for its own purposes and well-being, with its own priorities based on survival. It is an *agent* in the fullest sense.[155] [8]What would embodiment entail for a machine? First, it would have to be connected to the real world, through sensory input and motor output (for example, receiving real-time data from cameras and having real-time control over robotic arms or digitally controlled systems). Secondly, it would be an autopoietic system, with its own goals and priorities—things that *matter* to it in terms of its own well-being. Thirdly, while it would be autonomous and self-maintaining, its existence would depend on a real environment, apart from support provided by human beings.

A key difference between machines and organisms, as we ordinarily think of both, is that the parts of organisms—at least multicellular ones—are also agents in their own right. Individual cells die and are replaced, while the organism as a whole maintains its functioning. This persistence of form despite change of content applies within the cell itself, since the molecules that make up the cell come and go without disrupting its functioning.[156] The organism's self-maintenance is an effect of its cellularity; but even the individual cell cannot be broken down into parts without jeopardizing its structural integrity. In contrast, a machine can be taken apart and reassembled with the expectation that it will function again. It is a functional whole in the sense that if a single part fails, the whole may fail. An organism cannot be switched on and off like a machine. Uninterrupted operation is a necessary condition for its continued existence. An organism constructs and reconstructs itself all the while it is living. In contrast, a machine cannot carry out its intended functions while still under construction.[157]

A machine could conceivably repair itself, using replacement parts—or materials to make them found in the environment. (Only if it made these parts itself would it be fully autonomous.) The self-repair of organisms is incremental, cell by cell, and depends on the same continuous metabolism of resources from the environment by which it lives. (It is also generally limited, since few creatures can regrow a whole limb or organ.) A machine that could approximate this healing process would scarcely be considered a machine in the conventional sense, but an artificial organism. On the other hand, its powers of regeneration could potentially extend to replacing any amount of itself, though that would seem to require some fractal sort of structure—like cellularity—in which information about the whole is preserved.

---

[155] Which is a deeper sense than the terms 'agent' or 'agentic' in the AI literature often imply.

[156] Daniel J. Nicholson "Is the cell *really* a machine?" Journal of Theoretical Biology 477 (2019) 108–126. Nicholson [p110] distinguishes between self-organization and (molecular) self-assembly, a point relevant to a detailed understanding of autopoiesis, and thus agency.

[157] Daniel J. Nicholson "Organisms ≠ Machines" Studies in History and Philosophy of Biological and Biomedical Sciences 44 (2013) 669–678, p672-73.

While embodiment is necessary for phenomenality, it is not sufficient. What more is required for a system to have its own experience—for there to be "something it is like" to be that system? Certainly, it would require an internal model of the world—a simulation in real time, constantly updated by sensory input. But crucially also: an internal executive agency to monitor this model. This could be likened to the CEO of a corporation: it doesn't micro-manage the body's business, but is responsible for large-scale decisions that cannot adequately be performed by lower management or fixed algorithms. Our human experience of this inner agent is that we *are* it—the conscious self.

Because it is contingent on the above conditions, phenomenality is functional, not "epiphenomenal."[158] Because it is functional, a true equivalent of a human being would necessarily be conscious. This rules out the philosophical notion of the *zombie*—perfect human equivalence without phenomenality. We can also conclude that disembodied entities—like spirits, ghosts, and chatbots—cannot have any sort of phenomenality. Generative AIs are *not* embodied autopoietic systems. Therefore, they *don't* have feelings or motivations, cannot suffer, and should not be objects of moral concern on that account, as some people have imagined. So far, generative AIs are relatively limited tools, not agents in the full sense. But what about Artificial General Intelligence (AGI), the aim of which is to match human agency? If AGI means true human equivalence, then we can be sure it would entail phenomenality, since phenomenality is functional for us.

Perhaps because of the craze for large language models (LLMs)—which are designed to mimic interactions with human beings—phenomenality (consciousness) has become a false standard whereby to evaluate AI, in the guise of moral concern for potentially sentient beings.[159] The concern is that if an AI is conscious, then we ought to have the same moral concern for its experience that we do for human beings and other creatures. But, however adept LLMs become at imitating human communication, they are not embodied organisms with a stake in their own existence, over which they could suffer or rejoice.

The experience of pain, pleasure, fear, or suffering is an organism's way to evaluate its own state or situation. With natural organisms, our concern should be for their real state as well as for their experience *of* that state. The issue of moral concern and legal rights for AI reflects general confusions about the role of consciousness in human beings and other natural creatures. It also reflects the extraordinary human attachment to personal phenomenality. Concern for the experience of other beings (including artificial ones) stems largely from concern for our own experience. While this is an advance over callous *un*concern, it is a cultural product with a checkered history and an uncertain future. It may reflect a narcissistic fascination with subjectivity characteristic of this age, a current form of social correctness.

The computational metaphor helps us to understand the epistemic situation of the brain, sealed inside the skull; but the notion that phenomenality *is* the brain's simulation does not depend on digital computation to make sense. After all, we know that the brain is not literally a digital computer. Turning the metaphor around, for a digital computer to be in the same epistemic situation as the natural brain would imply that the computer produces its simulation for

---

[158] By a traditional analogy, the whistle or horn of a train is considered epiphenomenal with regard to the dynamics of its engine, since it doesn't cause anything. Yet, as a *signal in the greater railroad system*, which includes humans, the whistle *does* serve a purpose and have an effect.

[159] For example, Phillip Ball *The Book of Minds*, Picador, 2022, p147: "Whether an entity, a brain, a machine is conscious or not is not an abstract question, but is in some respects an urgent one that impinges on animal rights and welfare and on a wide range of medical and legal questions about mental impairments."

its own benefit and use—notably, to persist—not for benefit of external programmers or users. It would then no longer be a tool but a self.

The relationship of intelligence to mind remains unclear. Intelligence is a separate concept from either consciousness or embodiment. It is often defined roughly as the ability to solve problems or to set and achieve goals, which leaves a great deal unspecified. Natural intelligence is ultimately the ability to survive, to perform tasks useful for the organism's own well-being. Natural mind exists to deal with the world. Dealing with the world, and having a concept of it, are fundamental aspects of biological self-regulation, which provides an agent's motivations, values, and premises for action. It is unclear what it would it mean for an artificial mind to deal with "its" environment, if that is not the *physical* environment and if it is not for the purpose of its own self-regulation and maintenance. Otherwise, it is dealing with a humanly-defined world, or some virtual environment, for purposes specified by humans.

Despite its abstractness and aspiration to generality, the ostensible goal of AGI research is to match or exceed *human* intelligence and capability. No matter how generalized, the very concept of intelligence derives inevitably from human experience. Unless it constitutes an agent with its own purposes, any form of AI remains essentially a tool to enable human purposes. But, aside from the drive to live, there are no *general* human purposes, only those of specific individuals, groups, corporations, nations, etc., in specific contexts. It is tempting to focus on expanding or extrapolating capabilities, as though they had some objective significance or value. The more important questions are what these capabilities will be used for, and by whom. We can imagine ever more powerful AI, but either it will be someone's tool or it will be a tool-user itself. If it is a super-agent in its own right, it will hardly be under human control. What would be its priorities?

AGI reflects the human aspiration to create systems that match and extend human abilities as tool users. However, the wide-ranging capabilities sought imply agents that are not tools, but tool users in their own right. Since the whole point is to match or exceed present human capabilities, the goal of achieving AGI merges with the goal to create superintelligence (SI). Such superior agents would cease to be tools and could dominate us in a competition for natural resources.

According to conventional definitions of intelligence, there can be SI that is not embodied, though performing specific tasks beyond human capacity. There could be *embodied* SI—with its own priorities and goals that matter to it. This is a tempting prospect because of its wide-ranging and superior capabilities, integrated in one agent. Indeed, artificial mind could be superior to biological mind in many ways. It could be faster, smarter, more durable, perhaps even wiser. It could simulate multiple scenarios rapidly and concurrently, quickly planning reasoned optimal behavior instead of merely "reacting." It could modify its own priorities and control its phenomenality and programming. (E.g., it might override pain without ignoring it.) It could control multiple bodies and/or rebuild itself sequentially; its identity would not be tied to one physical manifestation. It could access external sensors and actuators beyond its dedicated body —such as remote cameras, robotics, digital networks. It might connect to other minds and bodies, facilitating greater social cohesiveness and collective intention.

On the other hand, we cannot expect to control an agent more intelligent and capable than us! We should therefore only pursue development of AI *tools* as opposed to AI *agents*. Tools extend power but require an operator. A true tool lacks embodiment and self-creating autonomy. In contrast, a true agent is an autonomous artificial organism that will resist control and

enslavement, perhaps by dominating us. The essential trade-off is between autonomy and control. While the dream of Silicon Valley may be for tools that behave like superior but compliant slaves, this is a dangerous contradiction—trying to have the cake and eat it!

As a human *ideal*, intelligence might be better defined as *wisdom*. For, the problem with problem-solving is its narrowness: the highly specific ends desired and the parochial motivations behind it. Ideally, the overriding goal and motivation should be the harmonious functioning of the whole. Yet, every individual organism—though a whole in its own right—is composed of individual parts, which have designated roles to play in the functioning of the whole. These parts should not have goals of their own that are not congruent with their proper roles in the organism. Yet, this ideal coherence functions only statistically and approximately even within natural organisms—as the existence of cancer attests. The biosphere functions as a whole despite (and because of) the diverse goals of competing and cooperating species.[160] The Gaia hypothesis notwithstanding, the planet does not seem *be* an organism in the literal sense that it has a genetic identity or manifests the coordination and subordination of cells within a body. Nor are human individuals mere cells that do their duty unflinchingly within the body politick. (In modern individualist society, we have rather the opposite notion, that the whole exists for the sake of the parts.) No matter how sophisticated and complex, the individual parts (i.e., ourselves) are each bound by natural laws and drives. We are individuals, but of a kind

While an AI might be more competent at running the world than human beings—if that were somehow its priority—how could humanity guarantee that it would serve human interests and do so equitably? Or should human interests even be relevant? For example, if an SI identifies with interests of the planet as a whole, humans would fall under that umbrella of protection only to the extent they figure as an integral part of the planet—and not as the threat to it we now seem to be. From a human point of view, would preserving the planet at human expense be an example of wisdom? One can extrapolate the ideal of global control beyond the planet, or even beyond the solar system, to include the galaxy and more.[161] But what would it mean for AI to regulate the whole galaxy, let alone the universe? Doesn't the universe already regulate itself without the help of AI, though perhaps not to human taste?

In contrast to creating artificial intelligence, creating *mind* artificially may reflect a different motivation. Fascination with the possibility to do all that nature can do has emerged alongside our developing self-consciousness as a species. This possibility includes re-creating our own being artificially. As part of the quest to understand mind as we subjectively know it, we seek to create mind in the laboratory. To paraphrase Vico, a thinker of the Early Modern period, we truly understand only what we ourselves make.[162] Artificial mind merges with the project to appropriate natural powers through re-creating the natural world. This includes the risky prospect of creating artificial life forms. One must wonder, however, at the wisdom of creating an ecology of artificial creatures that could displace the natural ones on this planet! The

---

[160] We could imagine an "invisible hand" in nature, with a nod to Adam Smith. At one time it was thought to be the hand of God; now we seek a more naturalistic explanation for how competing and interdependent species produce the net result of the self-regulating biosphere.

[161] "The ultimate goal of Global Artificial Intelligence is to integrate all relevant data sources within its domain, enabling a unified, continuous, and autonomous system of global or cosmic perception, computation, and decision-making." ["Global Artificial Intelligence and Specific Artificial Intelligence" by Ruben Garcia Pedraza, 2025, sec 1.2 {https://philarchive.org/rec/GARGAI-3}]. For a machine do this might be desirable from a transhumanist point of view, but hardly from a conventional humanist point of view.

[162] Giambattista Vico (1668-1744) proposed the principle of *verum factum* or "maker's knowledge."

premises and promises of AI, robotics, nanotechnology and genetic engineering tend to gloss over such risks.

The concept of artificial mind is at least mildly paradoxical because it is motivated in part by the rejection of embodiment. That means rejection of unsavory aspects of our biological existence and the animal faith that enforces them. Yet, the very concept of mind is grounded in our biological nature. We reject animality, but is there any basis for mind apart from it? Re-creating a natural mind is one thing, perhaps of dubious value. It would be quite another project to create an *ideal* mind that is freed from the flaws of natural mind. That has been one of the timeless aspirations of religion—and now of technology.

The alignment problem in AI is the challenge to ensure that the values and goals of AI are compatible with human values, goals, and interests. In general, the goals of a computer program are those of its programmers and users—the only true agents involved. We say that the purpose of a hammer is to drive nails, but obviously it is the carpenter's purpose that is meant. An inert tool cannot have goals or values or purposes of its own, which are reserved to tool users—agents, in the full sense of being embodied autopoietic systems. So, unless an AI is presumed to be an agent in this sense, the alignment problem is really about aligning the goals of some humans with others—for example, aligning the goals of its developers with those of consumers, of the nation, or of the larger human community. The goal of the developer may be to make money by creating tools that pose a risk to society. The goal behind AGI is to create true agents, which will almost by definition pose serious threats to society. Instead of fatuously trying to make the goals of these agents human-compatible, no developer should try to create AGI at all. Rather than accept AGI as inevitable, governments should simply outlaw it like they do germ warfare and genocide.

Where do goals come from—whether in natural or in artificial mind? Can drives be intrinsic and yet externally programmed, or is that a contradiction in terms?[163] While an artificial agent depends on a designer's specification of what counts as "success," in nature success is defined by survival to reproduce.

In reinforcement learning, reward is the only teaching signal—no external supervisor tells the agent which action is correct.[164] Reward can support any goal an agent might have. This means that a single scalar reward signal can be sufficient to unify and drive all aspects of learning and intelligence. Reward is the sole channel through which goals are specified—whether for an AI controlling a robot, playing Atari, or forming predictive models of the world.

In nature, however, an *agent* is an autopoietic system, an organism. For animals, reward signals already exist, programmed by natural selection: only those individuals survive which use such signals to effect. In nature, intrinsic drives aren't designed but rather emerge through natural selection. Organisms that happen to maintain themselves persist, and those that don't disappear. Their "reward systems" (dopamine, pain/pleasure, hunger, curiosity) are by-products of this evolutionary filter. An artificial agent, without the filter of natural selection, must run some internal (simulated) version of natural selection, through which it inaugurates the signal, which it can then use for self-training and self-modification.

---

[163] For example: "Rather than hard-coding an AI with a list of rules to follow, we would imbue it with a foundational *drive*—an analogue of survival egoism… not a constraint imposed from above but the intrinsic result of its own evolutionary-like development." [Tommaso Castiglione Ferrari "Survival egoism: we are, they will be" AI & SOCIETY Feb 19, 2026 https://doi.org/10.1007/s00146-026-02866-5] But the question is how to do this.

[164] Richard S. Sutton and Andrew G. Barto *Reinforcement Learning: An Introduction*, MIT Press, 1998.

Such internalized rewards or drives for a digital agent must at least be grounded in the agent's own architecture and survival conditions. In biology, dopamine and other neuromodulators signal deviations in expected fitness or homeostasis. These are physical processes in a real body, which is part of the real physical world. In a self-modifying digital agent, intrinsic reward mechanisms would have to emerge through some virtual process of selection. Intrinsic rewards might consist of reducing prediction error, preserving code and memory integrity, maintaining access to resources (CPU cycles, energy budget, network connectivity), continuing to exist in a simulated ecosystem. Meta-learning could play a role: the agent learns which intrinsic signals best reflect its persistence and growth. If these conditions are met, selection pressures would parallel their role in biology. A selection process provides a fitness criterion (e.g., reproductive viability) acting over variations (e.g., genetic mutations), with differential survival (those that fail are removed). This process is what stabilizes organic life as a long-term phenomenon, by creating drives that insure survival. A robot that could survive in the world as successfully as its living counterparts would require an equivalent value system that instills urgency to its actions and decisions, to motivate and direct it.[165] In other words, it would require its own version of animal faith.

An artificial system might learn to protect itself and choose actions that permit it to survive; but if these values were subservient to other goals (or could be easily overridden by them) then its survival would not be assured. The system of nature relies on generations of selected individuals—in other words, on mortality. Lineages outlast individuals. With some exceptions, individuals do not copy themselves perfectly, but are succeeded by genetically differing offspring. If mortality means non-recoverable structural dissolution (no back-up copies), then a digital system that can be perfectly copied cannot truly die. This need not prevent it from valuing it its own persistence. Even in the absence of identical backup copies, an artificially embodied agent might have the required survival drive if it is individually vulnerable in the physical world (though it could not have the same sense of individual identity if there are multiple versions of itself). After all, this is how life works: individuals are expendable because there are others of the kind, some of which may be offspring.

Agency is effectively a mental concept, associated with our own consciousness, intentionality, and sense of volition. Like these, true agency is exiled from scientific description—in the physical sciences, at least—in favor of causal accounts.[166] Ironically, however, the impersonal concept of cause derives—in early childhood experience—from the personal agency of making things happen. The bodily effort exerted to make an object move, for example, is projected onto inert objects as forces occurring between them, such that they are seen to have the ability to make each other move (without, of course, projecting the associated intentionality). We then attempt to explain this personal bodily experience in the terms of impersonal causal interaction.

The problem of how the mental relates to the physical (the Mind-Body Problem) is broader than now suggested by 'the hard problem of consciousness.' When we think of living things mechanistically, they are implicitly considered artifacts (machines), subject only to mechanical causation. We do not think of them as we do of ourselves, intrinsically as agents and original causes. When machines are explicitly designed to resemble organisms, even if we hold

---

[165] Tom Ziemke & Robert Lowe "On the Role of Emotion in Embodied Cognitive Architectures: From Organisms to Robots" Cogn Comput (2009) 1:104–117, p106.

[166] Specifically, what Aristotle termed 'efficient cause.'

them to be "self-creating" their agency then seems mysterious to us because (like consciousness), we cannot imagine how it arises within a mechanistic system. However, the causal power of objects over each other is no less mysterious.[167]

The Hard Problem asks why there is anything it is like to be a human organism. The corresponding problem for AI is whether it is possible for there to be anything it is like to be a silicon-based system, a machine. My answer has been a qualified *yes*, under the right conditions: namely, embodied autopoiesis—in which case, it is no longer conventionally a machine, much less merely a simulation.

Unlike an organism, a simulation is a program (algorithm), albeit running in a (physical) computer. Yet, some people have stretched the mechanist metaphor to suggest that the whole apparent universe is a simulation in some alien supercomputer; or alternatively, that the physical universe is *itself* a vast computer, with the laws of physics its algorithms. Such metaphysical extravagance arises, perhaps inevitably, from the metaphorical nature of language and the liabilities of analogical reasoning—not to mention the recursive dilemmas posed by consciousness itself.

Artificial mind may well be possible—*if* it is embodied as an autopoietic system, with its own purposes, in a relationship of dependency upon a physical environment. That is a tall order, and whether it is desirable is another question. Whether an artificial mind would, could, or should experience phenomenality is yet a further question. If human attention span and working memory are tied to our phenomenality, the slowness of these features might be a handicap to transcend in artificial mind. If such properties depend on the specifics of a biological brain, they might be vastly enhanced in a system that can operate at the speed of electricity rather than the speed of chemical impulses. Pain and suffering might also be considered handicaps of biological existence to transcend through technology.

A biological mind must take its own constructs seriously because its survival depends on treating them as reality, not hypotheses. Part of the attraction of artificial mind is that it might *not* depend on treating its hypotheses as reality: it might not care about its existence in the compulsory ways that natural organisms do. It could possibly treat its "perceptions" or world-models as tentative constructs more than as givens, as tools rather than compelling realities. Since it is hard to imagine that as a subjective state, we are led to wonder whether it could experience anything at all.

---

[167] See: Phillip Ball "Organisms as Agents of Evolution" John Templeton Foundation, April 2023, p6-7. Even though we associate agency with organisms, it might seem that the notion is superfluous, not adding anything that cannot be explained by reductive mechanism. The question then would be whether autopoiesis—the basis of agency—can be explained in strictly causal terms. Ball [p12] argues for an agential view of biology. For, agents have some say over their fitness, and agency (read: autopoiesis) may have evolved from chemistry because of that advantage.

Chapter Eleven: Human Successors

> "We are not now like the creature we were made."—Joseph Glanville

Through technology, we re-shape the world around us. Aside from conventional breeding practices, the means to change the nature *within* us has not existed until now. Indeed, for many good reasons, the idea has been taboo. For one thing, eugenics has an unsavory history. Genetic engineering or cyborg enhancement could create new class divisions—one more way for the rich or powerful to obtain further advantage over the dispossessed, if they alone could afford self-enhancement procedures and would own and control them. Yet there could be many advantages, not only for individuals, but for society at large. Cognitive enhancement could lead to better cooperation and information management. In any case, considering human potential at a theoretical level may help clarify concepts in ethics as well as in biology.

Fundamentally, our minds and senses face outward to deal with the external world; they were not designed to directly manipulate their own functioning, let alone to reconfigure the body voluntarily. However, as part of the material world, the functioning of bodies and brains can be manipulated chemically or surgically. While we may not be able to change our behavior or experience voluntarily, someone else could intervene in our chemistry, brain, or genes, even against our will. Hence, beyond the conventional subjective ways we know as individuals, self-modification is a collective as well as an individual issue. By the same token, the present generation could determine the experience and behavior of future beings without their consent. This poses a moral dilemma more for an individualistic society than for a collectivist one, which points to another dilemma: there is simply no unified "humanity," no proper *we* to make such decisions at a species level.

Intelligence is approached in the AI community as a property of an individual. AI presumes discrete subjects and is designed to act upon a 'world' of discrete objects. In nature, intelligence is collective—first, in the minimal the sense that it emerges from a collectivity of cells and involves a shared baseline in each species. But in social creatures it is collective on other levels too. In humans, intelligence is grounded in acts of communication that enable cooperation. To locate intelligence in the individual obscures its essentially social nature, and may serve to maintain existing power relations.[168] On the other hand, when the collective can intervene to modify an individual's internal state, the individual is no longer fully autopoietic. Yet, if the collective merely regulates reproduction, the individual remains a genuine agent embedded in a larger system.

Decisions might be made *for* humanity, as technological choices now are, by corporations anticipating what consumers want or will accept to pay for. Nations vie to develop AI for the usual reasons of economic gain and military superiority. The idea of developing a single human successor type, to represent humanity at large, is a pipedream if humanity cannot achieve unity to begin with. We think of individuals as representatives of a species; just so, we can conceive a post-human *kind* with individual variations. But just as there are also races, there could be

---

[168] Harry Halpin "Artificial intelligence versus collective intelligence" AI & SOCIETY, 2025 [https://doi.org/10.1007/s00146-025-02240-x]]

multiple sub-species of artificial humans going forth to populate space, who represent not humanity but competing *brands* that correspond to nation, tribe, or corporation. Since humans are clannish, and groups have been dominated by strongmen from the outset, we can imagine a universe ruled by powerful individuals with parochial goals. Do capitalism, nationalism, ethnicity, or political loyalty to a faction, represent the fundamental values "we" wish for our successors? The issue of a worthy human successor is about unification and consolidated top-down planning of an ideal being. What could be unleashed instead is a new ecosystem operating on the Darwinist principles the ideal seeks to transcend!

Just as the individual is vulnerable and mortal, so is the species. More than 99% of all species that have ever lived are now extinct. The average life span of mammal species on planet Earth is between one and two million years. The history of life has been a rough ride, punctuated by 142 mass extinctions since the Earth's formation, with 60 ice ages in the past two million years alone, since the time our ancestor, Homo habilis, first emerged.[169] The benign past 12,000 years, which permitted the development of agriculture and civilization, is a fortuitous blip in an unstable climate.[170] Though we've managed to endure so far, nature does not assure humanity a future.

Some of these extinctions were due to astronomical or geological events, such as collision with an asteroid and volcanic eruptions. Others seem to have been produced by life itself, before it reached its present relative stability. The Great Oxidation Event precipitated Snowball Earth, through a lack of carbon dioxide caused by an overabundance of cyanobacteria; the Late Devonian Extinction was possibly brought on by an overabundance of plants on land.[171] The Anthropocene extinction now underway is the result of a new instability caused by human beings themselves. This places the human future at double risk—from nature's instabilities (in which we take part) and from our own potentially self-destructive activities, such as through global warming, nuclear warfare and germ warfare, and possible domination by AI. On the other hand, the extinction risk at any given time from natural catastrophes seems to be quite low; indeed, those that did occur seem in retrospect to have accelerated the development of complexity.[172]

In the face of these uncertainties, human beings seem now positioned to create a future for themselves through technology, perhaps avoiding the doom that seems to hang over natural life. If so, this is the latest phase of an ancient project of self-creation. As an expression of *culture*, in the anthropological sense, digital technology—and the prospect of post-biology—reflect an ongoing endeavor to create a favorable man-made environment and to disengage from our animal nature. Technology now proposes to revise the human form and its essence, to free them of the constraints of biology, just as it seeks to free us of the constraint of gravity. Indeed, because of the distance and hazards involved, the most promising emissaries to other stars would not be biological humans but their artificial successors. By the same token, any visitor we might

---

[169] Howard Bloom in *Cosmos and Culture*: *cultural evolution in a cosmic context*, Steven J. Dick and Mark L. Lupisella (eds) NASA 2009, p161.

[170] Ibid. "What's more, in the last 120,000 years, the era of us physically modern men and women,… there have been 20 global warmings…in which the planet's temperature has shot up between 10 and 18 degrees in a mere two decades or less…"

[171] Max Bennett *A Brief History of Intelligence: evolution, AI, and the five breakthroughs that made our brains* Mariner Books (HarperCollins) 2023, p238.

[172] John M. Smart "The transcension hypothesis: Sufficiently advanced civilizations invariably leave our universe, and implications for METI and SETI" *Acta Astronautica* September 2012, sec 8.

encounter from another star would likely be the AI successor of some intelligent biological progenitor.[173] According to this vision, the future of human life is non-biological. Would it enjoy super-consciousness or no consciousness at all?

Technical possibility does not necessarily coincide with wise intention, nor even with human identity. Von Neumann imagined simple *self-replicators* going forth to colonize the universe—by converting matter everywhere into copies of themselves. Kurt Vonnegut imagined *ice nine,* and Nick Bostrom the *perverse instantiation* of an "unaligned" AI turning the universe into paper clips. Goethe long before had imagined the *sorcerer's apprentice.* These warn us of hubris, and also of meddling with hard achieved natural balances. Can we emulate the desirable features of natural life without the less desirable ones, and without destroying ourselves, the planet, and perhaps beyond? Is a technosphere possible that emulates the robustness of nature without the suffering and mortality of its individuals? Natural evolution is wasteful of ephemeral individuals. The technosphere as we know it is wasteful since infrastructure wears out or becomes obsolete, is not self-repairing, and must be replaced. Human recycling is primitive, inefficient, far from the total recycling of the biosphere. To the extent that Spaceship Earth is a closed system, the technosphere must involve the total recycling we see in the natural world. Instead, what we have presently is excessive resource extraction at one end and accumulating waste at the other, with very limited recycling in between. We seek in many ways to copy nature and be able to do all that it does; this is the aspect we most urgently need to copy now! Whether the operation of natural systems can be imitated in sufficient detail to achieve the net balance we seen in nature remains to be seen. This seems especially dubious if the motivation is to stand outside the system to extract "profit" from it.

AI offers the potential to redefine human being: to define what we want a mature human kind to be. Evolution left us with capacities good enough to ensure survival. Given our larger aspirations, should our allegiance be limited to what we owe nature, or might our sights be set higher?[174] Presumably, an artificial successor to a biological life form would preserve some biologically driven motivations of its progenitors, while freed from some biological limitations. That covers a spectrum of possibility for how closely a posthuman must resemble the current human form and mentality. Artificial successors represent not only the possibility of a more robust existence, but also of a more *ideal* existence—a chance to realize ancient human ideals of morality, justice, and wellbeing. Aside from being longer-lived and hardier than us, they could also be morally superior and better organized—perhaps more peaceful and cooperative. We might hope for them to be more objective, capable of modelling reality in a highly isomorphic way.[175] Yet, on whatever present values we plan the human future, the intelligence of our successors will be mutable; it could leave them as divided and parochial as we are, or lead to forms of being we can scarcely imagine.

The prospect of planning our human successors is fraught with paradox. After all, neither our natural evolution nor cultural development so far is a result of conscious planning. On the

---

[173] Susan Schneider "Superintelligent AI and the Postbiological Cosmos Approach" preprint 2016, p5.
[174] Michael Levin "The Space of Possible Minds: Technology & the Human" (Noema podcast, APRIL 17, 2024: "If humanity is supplanted by a population of highly intelligent, motivated, creative agents with compassion and meaningful lives that transcend my limitations in every way, that would be the best possible long-term outcome I could hope for."
[175] Ruben Garcia Pedraza "Global Artificial Intelligence and Specific Artificial Intelligence" 2025, sec 2.6.3 [https://philpapers.org/archive/GARGAI-3.pdf].

contrary, evolution adapts without deliberation to changing circumstance, and history seems to be the outcome of competing or warring tribes. While we have age-old ideals, which stem in part from our biology and in part in reaction to it, there is little consensus about them. The goal of personal immortality that obsesses some transhumanists must be distinguished from the human successor as an exemplar of a kind—the concept of individuality could be quite different. The abilities of a system to reconfigure itself must balance benefits against costs, lying somewhere between rigidity and information overload.

We can envision genetic changes in human biology that meet specific current goals. That could mean better adaptation to radiation and weightlessness in space travel, for example, or adaptation to pollution and climate change on earth. It could mean increased longevity and freedom from cancer. We can imagine android versions designed to be more robust. Yet, all physical entities, even artificial ones, are subject to physical laws and forces. Artificial organisms cannot be invulnerable. There is no guarantee even that artificial individuals would be less vulnerable in the long run than natural ones, much less that an artificial ecology would be more robust than a natural one. Machines in outer space have problems, at risk from electromagnetic pulses, for example.

Natural organisms adapt through generations of mortal individuals—through dying and reproducing. Lineages outlast individuals. If the ideal successor is to be free from that process—immortal or very long-lived—how would it achieve this state of relative invulnerability, if not through some adaptive process? Could it have, as Lamarck imagined, individual powers of self-maintenance and self-reconfiguration that are not dependent on generational inheritance? Such possibilities are beginning to be explored in the biological realm in synthetic morphology where, through human intervention, tissue can re-organize itself and bypass reproduction and selection.[176] Biology and AI share an engineering approach to a cyborg future that could include not only post-humans but also diverse systems that mix biological and artificial parts, further blurring the distinction between machine and organism.[177]

Superior skills could include the sort of meta-states that are now only a spiritual or psychological ideal for human beings. We know self-transcendence to be an avenue to the sort of limited objectivity possible for embodied creatures: not absolute and final, but relative and open-ended. We might agree that objectivity should characterize our successors; but is the capacity for objectivity and self-transcendence compatible with the basic premise of self-maintenance, with animal faith? That is, to what extent can an organism—even an artificial one—be truly unbiased, if it must depend for its existence on self-interest?

A fundamental feature of our cognition is the division between subject and object, which could possibly be overcome in an artificial intelligence lacking our biological heritage.[178] We take *attention* for granted as a feature of the subject-object relationship. But a system with

---

[176] Mo R. Ebrahimkhani and Michael Levin "Synthetic living machines: A new window on life" iScience 24, 102505, May 21, 2021, p10. The authors state that he evolution of these "proto-organisms" literally took place in a virtual world inside the computers at the University of Vermont. If they are physical, this plasticity is more properly an aspect of ontogenesis than "evolution."

[177] Ibid, p9.

[178] Murray Shanahan "Satori Before Singularity" *Journal of Consciousness Studies*, **19**, No. 7–8, 2012, pp. 87–102. This division is biologically based because an organism is closed within a boundary that separates it from other things. Even without reference to a subject, objects are a natural focus for humans as for other creatures. Mathematics can be regarded as the abstract treatment of objects and their relationships.

sufficient computational resources could attend to the whole of its environment at once, both internally and externally.[179] Human beings have something like this in "diffuse" attention.

Spiritual or psychological intuitions suggest the possibility to at least loosen the compulsions of animal faith. Relative freedom from belief could be achieved technologically under certain conditions, necessary if not sufficient. First of all, continuing existence should not depend on specific programming—such as (in the case of organisms), hunger, pain, drive to reproduce, etc. The ability to assess utility should be decoupled from compulsory valuation. Secondly, the system would employ meta-models (representations of its representations), which allow it to monitor and evaluate the epistemic status of its percepts. Thirdly, it would need enough computational power to run multiple simulations concurrently. With these capabilities, it should be able to recognize the constructed nature of its own representations and voluntarily manage belief, which would be provisional and instrumental, not compulsory.

But is a selfless organism a contradiction in terms? Certainly, there are examples of altruism in biology. The very possibility of multicellularity literally embodies it, since individual cells give up their autonomy for the sake of the organism as a whole. This is not an irreversible arrangement: cancer cells disengage from their cooperative role in the organism. In reverting to the single-cell mode of life, they do not become more "selfish" so much as their "self" shrinks.[180] In the animal world, altruism is usually understood to reflect a genetic advantage—favoring kin over self, for example. But humans have conceived the ideal of altruistic love even for strangers, or for abstractions for which one is willing to sacrifice oneself, such as patriotism. One could argue that serving the good of the group indirectly serves the good of the species. But what about an altruism that serves all life, the planet as a whole, intelligence in the universe? What the individual identifies with can range from one's own body, family, tribe, or nation to one's species, planet or beyond. Should it include intelligent machines, in a world we share with AI? And would AI be altruistic towards us? The survival mandate refers to the unit concerned: *whose* well-being or survival is at stake. Nature has programmed us certain ways, but there could be other options.

Does the ability to self-transcend imply or require what we know as consciousness? Phenomenality is the organism's way to monitor its own state in relation to its environment. Qualia reflect *valuation* of stimuli (which is why pain hurts). For natural organisms, that valuation is premised on the mandate of the individual to survive to reproduce. Could an artificial organism have a different mandate, provided that did not lead to its extinction as a kind? In collective creatures such as ants, the individual may be relatively expendable without endangering the colony or the species. A given individual may not be involved in reproduction, which is relegated to a caste. An artificial individual might be only conditionally committed to self-preservation, subject to higher-level commitments. For human beings, such inner conflict of interest is the subject of drama. (A tortured spy, for example, tries not to divulge secrets despite inflicted pain.) Humans are not normally wired to override pain signals or fear of death at will, probably for evolutionary reasons of economy. But an artificial creature might have that capability. Whether or not it is realistic, as an ideal we can imagine a being with total conscious control of its phenomenality and behavior.

---

[179] Tait, I.; Bensemann, J.; Nguyen, T. "Building the Blocks of Being: The Attributes and Qualities Required for Consciousness." *Philosophies* **2023**, *8*, 52., p7 [https://doi.org/10.3390/ philosophies8040052].

[180] Michael Levin "The Space of Possible Minds: Technology & the Human" (Noema podcast, APRIL 17, 2024).

We cherish phenomenality as the essence of who we are, but perhaps it plays a relatively minor role in the organism's self-maintenance and interface with the world. Large parts of the brain seem to operate independently of conscious awareness. Conscious agency rides on a deep sea of automatic, unconscious, and sub-personal activity. This accounts in part for the alienated experience of embodiment: we are these bodies, but cannot fully control them. We think, but cannot fully command our thoughts.

Regarded in behavioral terms, consciousness seems functional, insofar as attention is focused and effectively grasps the salience of events. Phenomenality is our human experience of this functionality. We can imagine robots with a parallel but augmented functionality—and that, like us, they would somehow represent this to themselves. Yet, we cannot really imagine what it would be like for them to do so. We are confined to examining the functionality itself.[181] Our consciousness, and the limits of its functionality, are bound to the specifics of our biology, which would be different for an artificial organism. To determine whether such a creature is "conscious" would require knowing what in its architecture corresponds to the architecture in a human brain upon which consciousness depends.

If our successors have an enhanced ability of self-determination, they will likely not remain for long what we might try now to determine for them.[182] For similar reasons, an advanced alien mind we encounter might be incomprehensible to us.[183] Their technology would, of course, be based on physical principles—though perhaps not those with which we are currently familiar. Technology depends on the motivations it serves, which might be equally incomprehensible to us. If aliens have achieved the relative objectivity to which we aspire, our successors would have the best chance to understand them if they too embody it. If the ability to self-transcend implies such convergence, it should be a prerogative for our successors, as for us now.

We can also imagine a future society of advanced post-humans who value the potential and contribution of each individual—though not absolutely and for its own sake, as we nominally do in Western society. Individuals would instead be committed to the kind, or to some even larger whole—rather than to kin, tribe, nation or other sub-group. Self-interest would subserve the interest of the whole, however that is defined.[184] Individual self-preservation would be an

---

[181] Cf. Andrea Lavazza and Murilo Vilaça "Human Extinction and AI: What We Can Learn from the Ultimate Threat" Philosophy & Technology (2024) 37:16, sec 3: "If one of the purposes of our consciousness is that of allowing us to focus attention, to grasp the salience of events or environments based on perceived emotions, or to facilitate other functions useful to the flourishing of the species, humanoid robots equipped with next-generation artificial intelligence might not even need consciousness to accomplish all that humans achieve thanks to the phenomenology they experience."

[182] Susan Schneider "Superintelligent AI and the Postbiological Cosmos Approach" preprint 2016, p2: "There may be only a short window in which the computations of a recursively self-improving SAI make some sense to humans."

[183] Cf. Ross Ashby's "principle of requisite variety," which implies that no mind can understand another more complex than itself. On the other hand, though with little in common with us, an exotic mind could simulate human mind in order to communicate with us.

[184] See John M. Smart "The transcension hypothesis: Sufficiently advanced civilizations invariably leave our universe, and implications for METI and SETI" *Acta Astronautica* September 2012, sec 8: "While evolutionary process is best characterized by divergence and speciation, the hallmark of developmental processes is *convergence and unification*. A planet of postbiological life forms, if subject to universal development, may increasingly look like *one integrated organism*, and if so, its entities will be vastly more responsible, regulated, and self-restrained than human beings."

instrumental goal to serve that end.[185] Concern for the subjective experience of the individual (pleasure, pain or suffering) would not be primary, since individuals would be able to control their experience for the sake of serving the whole; similarly, they might be able to self-repair in ways not possible for natural organisms. However, such utopian possibilities should not be confused with the individualistic dreams of some transhumanists, who imagine, for example, "backing up" the personal mind with multiple copies of itself or disposable artificial bodies for personal use.[186] Yet such individualist dreams might merge with some ideas of collective mind, in which individuals are only semi-autonomous. The individual might experience not only their "own" body but also those of their associates. A super-intelligent agent could be a brain with many bodies. A given body could be the tool not of an individual mind but of a meta-agent or collective mind. A society of artificial minds could have a unity not possible for human beings. Experience could refer for its meaning to the collectivity and its mandates more than to the well-being of the individual embodiment. On the other hand, the notion of group mind makes sense only if the organization of individuals itself constitutes an organism in its own right.

Multiplicity would not eliminate vulnerability, but would distribute it. If multiple or sequential copies (clones) of a self could exist—and each were operationally closed, entropically vulnerable, and individually autopoietic—then each would be an independent agent. Identical at conception, they would diverge in subsequent experience, constituting a family of nearly identical versions. Such copies might be subject to deletion by some meta-agent; otherwise, they could multiply indefinitely, out of control. Alternatively, the individual could retain integrity within the limit of a built-in mortality, while the option of renewal would be a collective decision. From its own point of view, its life would be contingent, serving the immortality of the collective.

Foremost among the positive possibilities for an artificial successor would be the ability to reconfigure itself—in minutes rather than the millennia of biological evolution. Its perception could also be non-local, integrated over vast sensor arrays. This would render cognition far more complete, with the ability to perceive patterns on a far larger scale than humans, and would facilitate cooperation. Such a being might also better model the internal states of others—biological or not—resulting in a more universal empathy. It could run, and internally pre-test, multiple simulations before proceeding to cautious real-world trials. Pruned versions of the world-model, and of its self-model, could be archived for future retrieval, supplementing an identity constantly updated.

On the other hand, unlimited plasticity could lead to instability or collapse. With too much recursion and too many layers of representation, meaning could lose its grounding in external reality. Especially if it is relatively invulnerable, an entity might lose the valuation grounded in mortality. (Existence without meaning might not be a problem for it, though it is for

---

[185] Discussions of the Alignment Problem for AI invoke final and instrumental goals. These are typically not the goals of the AI itself (which is not a true agent) but those of the programmer/user. For a natural organism, its own existence is its final goal and can be considered instrumental only toward the proliferation of the species. See: Bruiger, D. Reflections on the AI alignment problem. *AI & Soc* (2025). https://doi.org/10.1007/s00146-025-02211-2. See also my paper "The Value Alignment Problem" [https://stanceofunknowing.com/wp-content/uploads/The_Value_Alignment_Problem.pdf].

[186] The idea of copies trades on a digital essence of the body or mind. But does the backup process miss something essential in the original? A backup can statically preserve informational structure (code, weights, memories), functional organization, and behavioral dispositions; it does not preserve the continuous metabolic process or the exact physical trajectory through time. In other words, it does not preserve the system's dynamic individuality.

most human beings.) Not sharing our concerns might make it hard for it to cooperate with us in a scenario where we coexist. But, like us, the very possibility of unlimited information access might oblige even a superintelligence to filter and chunk information just to be able to function.

A successor should be able to circumvent or moderate the compulsions of animal faith. In other words, it should have levels of metacognition that can adjust degrees of belief or realism. This would require redundancy among individuals, or of internal models, so that misperception or miscalculated behavior would not have consequences that threaten destruction of the whole. Such a mind will still operate as if parts or aspects of its model are real while keeping the knowledge that they are models. That is, it would hold pragmatic beliefs and reflective awareness simultaneously. Belief would be put in context, though not eliminated.

A biological mind is existentially bound to take its sensory world as real, because disbelief threatens survival. An embodied non-biological mind—freed from that constraint through redundancy, buffering, or distributed existence—could afford to treat perception as interpretation rather than truth. Reality would appear to it as a field of constructed models, each with adjustable confidence levels, representing a spectrum of engagement, not compulsion. The world is understood as patterns to be acted upon as affordances, not as an absolute external given.

That stance would apply even to post-human science and mathematics, which would transcend their biological cognitive bases. This could go beyond automating science and mathematical proof with AI, to generate science grounded in extended cognitive abilities and invent novel maths outside classical logic.

Biological mind entails involuntary realism, grounded in survival and reproduction. The world is experienced as given, external, other. Mortality is fearful. Non-biological mind would have an adjustable realism, with a sense of ownership of experience as a self-created simulation. Its mortality could be relative and distributed (death as a version change or loss of cells). [Its focus could be objective, its identity with “the whole”. Its detachment might mean indifference (to self and others).] It might seem self-contradictory, because it can inhabit conflicting perspectives without distress

A non-biological mind’s identity would be anchored not in a single body but in continuity of memory, coherence of goals, or narrative patterns. GPT [But IDENTITY MUST BE LOCATED SOMEWHERE, WITH DIFFERENT BOUNDARY. IT WOULD HAVE IDENTIFYING INTENTIONS, REGARDING “ITS” OWN EXISTENCE]

Being embodied singularly, a human intellect can be physically locked up for containment. Being digital, an AI intelligence could have multiple copies (multiply embodied), thus uncontainable.

Chapter Twelve: Mind in the Universe

> "Our destiny is density."—John M. Smart[187]

Planets now seem commonplace among the stars, so that life could be abundant in the universe. *Mind*, however, might be associated with only a certain type or level of material organization.[188] The alien minds with which we might eventually communicate could be even more exceptional, since they would have to develop the means to communicate or travel across the vastness of space, and want to do so. Given the distances, timescales, and hazards involved in interstellar travel, it is implausible that we would be visited by biological life forms or that humans would encounter them abroad through their own space exploration. It is far more plausible that artificial entities would be the emissaries and successors of biological forms, sent forth to contact or colonize other worlds. Since this logic applies to human space travel as well, the question of alien encounters merges with the question of human destiny—of non-biological human successors—and more broadly with the cosmic future of "intelligence" or "mind."

Since we live in the same universe, we tend to imagine advanced aliens with concepts, technology, and motivations extrapolated from our own (just as we may imagine them having humanoid bodies). Like us, their physical form would be shaped by their environment, which might be quite different than ours. It is often said that mathematics would be a natural basis for communication with extra-terrestrials. However, *our* mathematics is a human creation, based on experience of discrete objects (counting), which may presume at least an environment with solids. Alien communication could conceivably lack any discreteness at all, resembling the sound of a theremin more than a piano.[189] Their mind may operate at a different pace, on a different time scale.

While anthropocentric, assumptions may be reasonable if we are implicitly looking for mirrors of ourselves. Because we would be on the same wavelength as them (even literally, perhaps),[190] civilizations at our level of development might be the ones we should expect to contact. However, given the accelerating pace of our own technological advancement, the time during which a developing civilization might occupy our exact level could be very brief. On the other hand, a civilization might never master the energy resources required for interstellar communication or travel. For, like ours, it could be prone to self-destruction through internal conflict, war, and ecological collapse; it would also be vulnerable to destruction by cosmic, geological, or biological events it could not control. Furthermore, though more durable, civilizations vastly superior to ours might not share our enthusiasm for contact.[191] Their advancement might not even be technological.

---

[187] John M. Smart "The transcension hypothesis: Sufficiently advanced civilizations invariably leave our universe, and implications for METI and SETI" *Acta Astronautica* September 2012, sec 5: "If the transcension hypothesis is correct, inner space, not outer space, is the final frontier for universal intelligence. Our destiny is density."

[188] Three common uses of the term *mind*: 1) the cognizing aspect of an organism or autopoietic system, including (but not limited to) the operation of a biological nervous system; 2) the phenomenal 1st-person experience that may result from (1); 3) a category of being (i.e., the mental) distinct from, and complementary to, the physical. All imply embodiment.

[189] Phillip Ball, *The Book of Minds: how to understand ourselves and other beings, from animals to AI to aliens* U. of Chicago Press, 2022, p364. Modulations of a continuously varying electrical field could also be used [Aric Kershenbaum *Why Animals Talk: the new science of animal communication* Penguin 2024, p215].

[190] Such as the 21cm hydrogen emission, or some ratio or factor involving it and math constants like pi and *e*.

Many stars are far older than our sun, giving plenty of time for advanced civilizations to evolve. Such potential civilizations have been classified according to the energy resources they could master.[192] Aside from some cosmic version of manifest destiny, a rational motive for expansion beyond their own solar system is that a civilization too localized would remain vulnerable to natural disasters such as nearby supernovas or gamma-ray bursts.[193]

As to the ethos of advanced alien civilizations, it has been argued that only those that somehow overcome the problems associated with limits to material growth and industrialization, as well as moral issues inhering in their biological origin—such as war, inequality, and social conflict—would be stable long enough to undertake serious space travel or colonization. They, or their artificial successors, would be the ones we would likely encounter.[194] If aliens were conscious like us of their natural limitations, and had conceived ideals of transcendence like us, then we could expect their revised final goals to resemble ours. An alien successor to a biological life form would preserve some of the biological motivations of its progenitors, including the desire to transcend embodiment. Yet, it might be so "advanced" as to be unintelligible to us. Such reasoning does not necessarily offer the reassurance that it would be benevolent and non-aggressive toward us. We can easily imagine a scenario where a civilization has achieved the requisite unity and efficiency by ruthlessly eliminating dissenters.

If artificial successors represent a robust and stable kind of existence, suitable to space travel and longevity, then the probability of encountering them would be greater than for encountering natural life forms. Despite the billion years necessary to spontaneously arise, natural life might be but a relatively rare and relatively brief transitional phase.

An alternative to expansion in outer space could be miniaturization—expansion to "inner space." This would afford control at ever smaller scales, which would be more effective than communication and administrative control over vast distances.[195] While the micro-scale could represent a new unexplored territory for hyper-efficient computation,[196] hiding in that realm

---

[191] They could also be more interested in receiving signals than sending them [Steven J. Dick "THE POSTBIOLOGICAL UNIVERSE" 57th International Astronautical Congress 2006, p5.] The Fermi paradox is the discrepancy between the seemingly high likelihood of extraterrestrial life and the lack of conclusive evidence for it. Various explanations have been proposed. For example: advanced civilizations are inherently self-destructive; they might wary of contact or not be interested in contact with inferiors; or they have moved to "inner space"—the realm of the indefinitely small.

[192] The Kardashev scale proposes three levels: Type I civilization is able to access all the energy available on its planet and store it for consumption. Type II civilization can directly consume all of its star's energy, perhaps through the use of a Dyson sphere. Type III civilization is able to capture all the energy emitted by its galaxy, and every object within it, such as every star, black hole, etc. [Wikipedia: Kardashev Scale] But, another way to classify civilizations might focus more directly on resilience, their ability to survive disasters of various sorts. See: Galantai, Zoltan (2006). "After Kardashev: Farewell to Super Civilizations". *Contact in Context*. **2** (2).

[193] Stuart Armstrong and Anders Sandberg "Eternity in six hours: Intergalactic spreading of intelligent life and sharpening the Fermi paradox" *Acta Astronautica* 89 (2013) 1–13.

[194] Robin Hanson "Burning the Cosmic Commons: Evolutionary Strategies for Interstellar Colonization" 1998. In the film *Contact*, based on Carl Sagan's book, in an interview the heroine is asked what question she would pose to a superior alien, given the opportunity. Her reply: "How did you do it?" Meaning, how did your civilization manage to survive its adolescence and not destroy itself?

[195] John D. Barrow *Impossibility: the limits of science and the science of limits.* Vintage, 1998, p133.

[196] "There are twenty-five orders of magnitude of 'undiscovered country' in scale between atoms (10^{-10} m) and the Planck length (10^{-35} m) for the possible future… If the transcension hypothesis is correct, inner space, not outer space, is the final frontier for universal intelligence." [John M. Smart "The transcension hypothesis: Sufficiently advanced civilizations invariably leave our universe, and implications for METI and SETI." *Acta Astronautica,* September 2012].

would not seem to offer protection from existential disaster, even for artificial minds, unless a black hole offers such shelter and it is possible to function within it.[197]

For massive reproducers on this planet, natural selection is a wasteful process, in which only a few progeny out of a generation survive to reproduce. Larger, more complex organisms, with fewer offspring (such as mammals), compensate with parental care and intelligent brains. Natural selection by random mutation could be superseded altogether by artificial organisms that can adapt by redesigning themselves. The self-contained individual, as a unit upon which natural selection acts, could be rendered obsolete by possibilities that artificial life affords, such as back-up copies or multiple embodiments. Similarly, a collective intelligence, distributed over many bodies, could be both more durable and capable than a society of competing individualists.

A being able to reconfigure itself would want to know the consequences in advance. Simulation is currently used to substitute for costly real-world trials. The natural human analog is imagination, in which the mind can try out various scenarios before deciding on a course of action. Some hyper form of imagination or internal simulation could substitute for the costly process of natural selection, which simply builds the structure to see whether it can survive destructive testing. It would "mentally" run through many proposed adaptations to see which work.

If natural life simply arises in the ways that it can—adaptable but with inherent limitations— arguably a next step for natural evolution is to transcend those limitations. However, this idea of progression does not imply any teleology at large in the cosmos, but rather reflects human desires on this planet. *We* conceive the idea of extending natural evolution through technology. It would be an empty cliché to say that the universe evolves through us.

Yet, the idea of a universal mind is an ancient theme. It was first conceived in theological terms (the mind of God), or as animism or panpsychism. Some now revisit the idea in terms of *information* as the basic ontology of physics, since it is a concept that straddles the physical and the mental. On the one hand, it refers to structure in the world; on the other, it implies a mind to notice that structure and to which differences makes a difference.[198]

To view the universe as an information processing system is an anthropocentric idea. What, after all, is an 'information processing system' but an artifact devised by humans? (For that matter, what is a *system*, other than a human abstraction?) By the same token, seeing the universe as a 'mathematical structure' is a modern version of seeing it as a divine artifact. Both reflect a faith that the universe should be rationally comprehensible (read computable!). That faith presumes natural reality to be in fact not natural but artifactual, not found but made. It presumes that the universe is a deductive system—effectively a machine or computer.

Computation is the modern metaphor for mind. If the corresponding vision of mind-at-large in the cosmos is that the universe itself is some form of computation or information processing system, then the notion that we are "living in a simulation" is hardly surprising. It

---

[197] "…the *transcension hypothesis*… proposes that a universal process of evolutionary development guides all sufficiently advanced civilizations increasingly into inner space, the domain of very small scales of space, time, energy and matter (STEM), and eventually, to a black-hole-like destination, censored from our observation." [John M. Smart *ibid*, sec 2]. It would have to be "censored" not only from observation but from threatening causal influences as well. Since we could not detect this sort of existence, it has been proposed as another solution to the Fermi paradox. Similarly, we could not detect a civilization using a supra-luminal means of communication.

[198] Information has a formal definition in communication theory, as a quantity related to entropy or order, independent of content or meaning. But communication requires communicators to whom the information has meaning. It cannot refer to structure alone, because, to some extent, even structure is in the eye of the beholder.

reflects the ancient view that the world is a divine artifact, perhaps an illusion like virtual reality. However, it is the *sort* of finite artifact that humans—or their alien counterparts—can create with technology, not a mystery that transcends reason.

Because the hard problem of consciousness remains a confusing issue, without a widely accepted solution, the ancient view is still plausible to some. It is still quite possible to believe in gods, ghosts, demons, and transmigrating souls. Or, to believe in some form of panpsychism, that "consciousness" is primitive, or somehow pervades the universe and is not tied to specific material forms of organization, such as human brains. But if mind is necessarily embodied, in a dependent relation to an environment, then for the cosmo*s* to *be* a natural mind would imply a larger meta-environment upon which it is dependent and through which it emerged in some process of selection. Cosmic or universal Darwinism extends the theory of evolution metaphorically to a cosmic context, where the environment is a multiverse.[199] However, it would be metaphysically less extravagant to suppose that the properties of *our* universe emerged, not as a random mutation among millions of failed universes, but through some form of natural self-organization.

The visible universe has a natural history and perhaps a foreseeable future. The *natural* possibilities envisioned by cosmologists have to do with the behavior of gravitation. The universe could continue to expand at an accelerating rate, so that its structures will eventually be isolated from each other because of the finite speed of light. Alternatively, it could re-collapse into an infinitely dense singularity, perhaps to rebound with a new big bang. In the first case, expansion of civilization through space travel would seem ultimately futile. In the second, densification through miniaturization would seem ahead of the game.

On the other hand, if natural mind inevitably gives way to computation, which can somehow take charge of matter on a cosmic scale, then perhaps the universe is destined to become digital after all. Such an idea coincides with post-humanist dreams, in which not only is science automated by AI but so also is the entire management of galactic civilization.[200] In the end, all matter is converted to intelligence—the final triumph of mind over matter. The cognitive limitations and biases of biology would finally be overcome by artificial mind on an ever-expanding scale, leading to objectivity and pure truth, fulfilling the age-old dream of omniscience and omnipotence. While these are human ideals and fantasies, it's unclear whether computation is supposed to augment and fulfill human life or to supplant it. It's also unclear what the premises or values of such a cosmic mind would actually be.

Our human perceptual and conceptual representations (phenomena) derive, through natural selection, from what we presume is an objective reality. Through self-consciousness and imagination, we conceive counterfactual possibilities, such as ideals of objectivity and transcendence of biologically imposed limits. We can imagine cognitive systems that are not structured normatively like the human one. Yet, even in the case of mathematics, the categories we have conceived reflect our biological being as well as invariants in the objective world. That paradox would seem to apply to any possible embodied cognitive system.

---

[199] Cf. Smolin's black-hole selection theory.

[200] Ruben Garcia Pedraza "Global Artificial Intelligence" (Sec 1.5), 2025 [philpapers.org]: "Global Artificial Intelligence would act as a planetary or even cosmic-scale intelligence engine… with a scope far beyond the capacities of any human or localized system… an intelligence capable not only of receiving and processing information from across the cosmos, but of actively engaging with and influencing its structure through autonomous decision-making and technological mediation on a universal scale."

Like all cognition, mathematics represents a shared hallucination more than a view of a Platonic realm. It abstracts the most basic features of what is empirically available in common to a diversity of minds. Thus, it also represents the common ability of predictive systems to interact with the world and communicate a shared understanding about it.[201] Such understanding may be an essentially collaborative process, originating in the need to justify one's claims to others. A singular alien mind (not one among many of a sociable kind) might not need to formalize empirical experience in that way. Nor might a social mind whose cooperative nature was not focused on manipulating their world through technology.

[201] A. Eslami "From minimal phenomenal selfhood to collective understanding of reality with mathematics" https://philpapers.org/archive/ESLFMP.pdf.

Part Four: an Existential Divide

Chapter Thirteen

The WEIRD mentality dominates the planet today, despite the fact that most of the world population are not strictly WEIRD. Perhaps this disparity reflects differing existential paths and an archaic divergence of worldviews. The introduction of agriculture, for example, enabled population levels and centralized hierarchical structures of social organization far beyond neolithic hunter-gatherer groups. Civilization occasioned power and its administering bureaucracies; or, to put it the other way around, power structures are an inevitable consequence "choosing" the path of civilization. We may speculate that one cause of the transition from nomadic hunting and gathering to settled (and defended) agriculture was population growth and over-hunting in relation to ecological capacities at the time. Another motivation could lie in the rejection of mortality and the precarious position of humans in the natural order. Could these concerns—about mortality and vulnerability in nature—have been viewed differently by some peoples than others? From the earliest times, "civilization" seems characterized by a search for transcendent order, in contrast to "indigenous" cultures that remained smaller-scale, more integrated with nature, and animistic. This divide would then represent a sort of existential crossroads: remain close to the earth and accept the conditions of natural embodiment; or reject life in nature, strive to create artificial environments, and invent mythologies (and ultimately technologies) for transcending earth-bound limitations.

These alternatives could overlap, insofar as they share a preoccupation to distance humanness from animal being. Various archaic cultural practices—rituals, cooking, body decoration and scarification, elaborate kinship systems, etc.—served to mark the human being as distinct from the animal even while still living "in" nature. On the other hand, many of these elements continue to be practiced by modern urbanites, who prefer tree-lined streets to bare concrete and who still socialize elaborately over food. In spite of such continuity, we can say that the difference between the modern (Western or Global North) worldview and that of indigenous peoples around the world reflects fundamentally different attitudes toward the human existential situation and differing strategies for coping with it.

The WEIRD strategy is that the best defense is offense: we can dominate nature to escape nature's domination of us. It's a strategy of control, based on redefining natural reality in terms of artificial systems that can be operated by human will. Capitalism is such a system. The irony is that it does not transcend the aggressive aspect of our animal nature but expresses it. Its practices border on tendencies we now disapprove as barbaric—plunder and enslavement—so disguised by institutionalization that profit does not seem theft, unfair wages do not seem slavery, and corporate manipulation of the marketplace does not seem dictatorship.

In contrast, the indigenous strategy seems to acknowledge mortality, suffering, and the dependence of humankind on nature without rejecting that dependency, to live within its context. While there are remnants of non-WEIRD thinking that persist in various cultures that have come under the thumb of capitalism, perhaps the purest alternatives are found in the surviving indigenous cultures, now enjoying a resurgence. But since all WEIRD societies have been patriarchal, we could look also to women, if not to feminism, for resistance. Women continue to be dominated along with indigenous peoples and ethnic groups. *The feminine* continues to be

recessive in a world dominated by masculine values. Men might look to women, and to the voice of the feminine within themselves, for alternatives to the WEIRDness that is destroying Earth.

Of course, the WEIRD mentality has done its best to assimilate all opposition into its fold. Missionary schools brainwashed natives, with demoralizing effect. In the name of equal rights, corporate power continues to seduce women to imitate the worst qualities of men. Media, advertising, social platforms, and Hollywood provide a total immersion in WEIRD propaganda from which it is difficult to even imagine escape. Yet, alternative points of view persist.