ANIMAL FAITH AND HUMAN DESTINY © 2025 Dan Bruiger

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

Through experience and education, I began to wonder and ask questions. When one eye is closed, why does the world look a slightly different color viewed through the other eye? How is it that I can see faces and familiar shapes in clouds? What is it like to be a bug making its way through the wilderness of the backyard lawn? It occurred to me that I too might be a creature making its 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, the intricate nature of the world, and the relationship between themselves as perceiving subjects and what they perceive. They alone have a science that allows them to grasp that the senses do not offer a transparent window on reality. 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 revulsion—meat machines who aspire to be gods. This inner conflict shows up in culture and in daily life as a tension between the natural impulse to believe our perceptions and thoughts and a hard-earned wariness in regard to them. Whether literally true or not, perception normally serves us well—but not always. Humans have gained their ascendency on the planet by being able to question it. Science is a sceptical 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.

As a philosophical concept, naive realism asserts that the world is exactly as we perceive it to be: objects exist as they appear, independent of our minds or perceptions. In this view, the mind's internal processing is treated as a direct and unmediated representation of the external world, an accurate portrait. Few philosophers today would endorse such a view. But naïve realism is more than the name of a philosophical doctrine. In truth, we are *by nature* naïve realists, taking for granted that our perception of the world is reality itself, without acknowledging the mind's active role in constructing that perception. This transparency of perceptual processing is our natural default state. When we see a tree, for instance, the tree is just *there*, objectively and independently of our minds, whether we look at it or not. We don't recognize the mental processing involved in *seeing* it. Indeed, we cannot see the internal workings behind seeing, which we can know only 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 internal sensory simulation of the tree as though it was the tree itself: to *believe* the simulation.

My aim in this book is to examine that fundamental mental impulse that makes us believe our senses, thoughts, and feelings and which renders scepticism necessary. The philosopher Santayana called that impulse *animal faith*.[1] Even to define that concept invites a discussion of scepticism. One makes sense only in relation to the other, and so he appropriately titled his book *Scepticism and Animal Faith*.[2] When faith is blind, it can be a liability. As we shall see, scepticism serves as an antidote to faith; yet, animal faith pervades cognition regardless of scepticism, which itself requires a form 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." It is faith in the truth of what we perceive and believe about the world—in short, we are compelled to believe our own minds and experience. It 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. As the name implies, animal faith is a biological necessity, as basic as the need to feed on other creatures. To begin with, the animal engaged in action instinctively must believe in the existence of a knowable external world. It must tacitly assume a real field of action in which to act.[3] As opposed to a spiritual identity, and godlike ideals of benevolence and objectivity, 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 *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* it a certain way and *act* within it

[1] To be clearly distinguished from "study of animal behaviours that suggest proto-religious faith." [Wikipedia]
[2] George Santayana, *Scepticism and Animal Faith: introduction to a system of philosophy*. Charles Scribner's Sons, 1923.
[3] Herman J. Saatkamp Jr. *A Life of Scholarship with Santayana*, 2021, Brill, p10.

a certain way. The creatures that exist, including us humans, are only here because they take the external world seriously and have learned to deal with it in the specific ways that allow survival and reproduction. Those ways include the fundaments of perception: belief in the reality of objects, substance, space and time; and the compulsion to evaluate stimuli, to judge things good or bad for oneself or one's kind.

One may protest that objects and space and time must exist anyway, objectively, as the pre-existing stage upon which our actions play out, upon which our sensory information depends. The point, however, is that we are not—and cannot be—in direct touch with that world, which Kant called the noumenon, the world-in-itself. But mustn't our perceptions at least reflect that reality if we are to survive? The answer is both yes and no. Presuming that there *is* an external reality, there must be *some* relation between that inaccessible objective world and our experience of it. 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. 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. Map and territory cannot be viewed side by side for comparison.

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 like some people must learn to operate and identify with a prosthetic limb it. Throughout life we are obliged to negotiate the world in terms of the needs of this body and as seen through its eyes. This natural state of affairs—the state of *embodiment*—must nevertheless seem awkward to a consciousness that can imagine limitless possibilities. It is an unwelcome and disturbing realization for a 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 is the insult of death.

For a mind that can conceive itself apart from the body, life in an animal body is problematic in several ways. There are, of course, the ongoing challenges: disease, dysfunction, aging, the ever-present possibility of injury and pain. There are the psychological problems of inner conflict, grief, anxiety, 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, tribalism, fanaticism. These liabilities of biological embodiment render us ambivalent toward nature and the body, which seems at best an instrument to use and at worst an enemy or prison. But they also leave us highly perplexed. How can a physical object be a conscious subject? Or, how could a conscious subject not be a 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; for, *mind* is a nebulous catchall for a subject's experience, and *body* refers not only to one's particular carcass but to matter generally. While it is a challenge is to understand how mind can exist in a physical universe at all, more specifically the challenge is to understand how the biological brain can give rise to ones' personal subjective experience—now usually called the Hard Problem of Consciousness. The designation "hard" reflects the basic frustration involved, reminding us that the puzzle consciousness poses to itself remains unsolved.

The problem, of course, is uniquely an intellectual challenge for human beings, who know themselves to be "conscious." It asks how a person's physical brain can produce their first-person experience. This is hardly a problem *for the brain*, of course, which undeniably manufactures that experience on a daily basis. The challenge for philosophers, scientists, and the rest of us is to understand *how* it does that—how *that* product comes from *those* ingredients. It is

an odd perplexity, since the outward-looking human mind habitually tries to understand things as causal processes in the external world, which includes neural processes in the brain. But those neural processes somehow give rise to the experience of that external world, which includes the neural processes in the brain that it tries to understand…and so on in an endless recursion. In speculating about consciousness, the conscious subject is caught in an epistemic loop. The whole thing bites its own tail.

Furthermore, the circularity of our epistemic situation is general. The truths sought by science are no more independent of the inquirer than the truths sought by ordinary cognition. Both ultimately are survival strategies: we see in ways that allow us to exist. While normal perception *seems* to us a transparent window on the world, which we take for granted, we also *know* that it is a product of the nervous system, as much shaped by the biology and the needs of the organism as by the external world. Logically, we must surmise that scientific cognition too is likewise a function of the observer as well as of the world observed.

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. While that does not imply that no real world exists outside the brain, it does maddeningly complicate our understanding of it since our only access to that world is (circularly) through the brain's simulation of it! The brain cannot get outside itself. Our scientific vision of the world 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—for reasons that are only obliquely related to objective truth.

The recent interest in "generative models" is the latest development of the computational metaphor to shed light on the natural operations of mind. Their inventive power resembles that of the brain—to imagine, dream, fantasize, and hallucinate.[4] The brain's commitment to what are clearly its own creations is the very thing we are calling animal faith. It might apply even to machines. But what exactly is that commitment?

Chapter Two: Fiat, or Intentional Connection

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

Causality is necessary to explain mind, but it is not sufficient. 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 on a causal level, 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

[4] 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.

structure and why they behave as they do, as though from a designer's perspective.[5] Of course, natural organisms are not human artifacts and we do not assume intelligent design. 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 subjective phenomenology. Apart from any observer's analysis, organisms manifest their *own* intentionality. Their purposive behavior cannot be reduced to physical causes, even if it depends on them.

The concept of intentionality has a convoluted history in philosophy as "aboutness," which is essentially a linguistic notion entailing reference. Since only humans use fully grammatical language, let us broaden the concept and reframe intention outside the context of language. Intention, in this broadened sense, is an internal connection made within an *autopoietic*[6] 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 also be a connection made within an artificial system—if, indeed, it is feasible for such a system to 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 creature, which does not simply *react* to stimuli but actively *responds*. Something far more complex is going on than simple action-reaction, for instance.

The concept of intentionality proposed here includes conscious intention but is not limited to it. Nor is it tied to linguistic or symbolic reference, which has been the traditional focus since Brentano. The reference of words to things is merely one example of the (human) organism's ability to internally make connections. (We will see, however, that language can help us metaphorically to understand the nature of such connection.) More abstractly, *intending* is the internal act of an *agent,*[7] which 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 embodied being is a complexly organized physical system. But *embodiment* implies something more than being physical: a certain *relation to the world.* Embodiment is a relation of an autopoietic system to its real environment, in which events *matter* to it, ultimately in terms of its continuing existence.[8] Every living organism stands in this relationship to the world, entailed

---

[5] Causality and intentionality are alternative ways of looking at the behavior of the system, of explaining how it works. As ways of looking, each projects aspects of the *observer's* mental processing onto the system. Both are attributions by an observer—that is, 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. 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.

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

[7] 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."

[8] No AI, robot, or other artifact is yet an autopoietic system, with an embodied relationship to its world. Can that relationship be simulated? A simulated organism is virtual, not physical. One simulation can exhaustively simulate another, since both are products of definition to begin with. But no model can exhaust any portion of natural reality, which is not a product of definition.

by its participation in the system of life we call the biosphere. It is a relationship inherited through natural selection and maintained and refined by the individual organism. The survival mandate implies priorities held by the organism, which reflect its needs and relation to its environment, and which motivate its behavior.

Sensory input thus serves more as a suggestion than as a command to a complex organism that can consider options.[9] But the organism as a cognitive system then issues commands within itself, by making internal connections, and thus programs itself, so to speak. This act of assertion or internal command I call *fiat,* which means, literally, "let it be done." Animal faith is then the willingness or compulsion to obey such commands, which Coleridge called 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 it *judgment.* Schopenhauer speaks of *will*, Helmholtz of *unconscious inference.* More recently it's been called "the ego tunnel" (Metzinger) and "the reality illusion" (Rausch).[10]

Fiat is the exercise of agency. It is the very basis of consciousness. Like gods or monarchs, we simply declare the inner show into existence, moment by moment, and naturally believe our own creation. This "show" is continually updated and guided by input from the senses. Like reading tea leaves, the patterns one discerns 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 *meaning* of what naturally appears to us as real. *Realness* refers to our dependency on a world we did not make and whose rules we did not choose—a dependency against which we may also rebel, having imagined a freedom beyond it.

The essence of a cognitive system lies in the very plasticity of its internal connections, which are in themselves arbitrary, like words are arbitrary symbols for meanings. That is, there is no pre-existing or absolute meaning inhering in a given connection, as there is not in any symbol, nor any inherent reason why it should be made or endure.[11] The significance of the connection 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 set 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 as conscious or capable of having experience if isolated from the brain.

This plasticity is both a boon and a bane. Internal connections must not preclude survival, but otherwise are not obliged to correspond to the external world. 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, 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 make its existence impossible. Its 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.

---

[9] Phillip Ball "Organisms as Agents of Evolution" John Templeton Foundation, April 2023, p20.

[10] 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, etc.

[11] If only for this reason, there cannot be a one-to-one relation between input and output, or between the "map" and the "territory" it represents.

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 real world of consequence outside the skin, not a movie running inside the head. (The idea that there nevertheless *is* such a movie is rather modern, reflecting the birth of scepticism.) Even today it serves 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 compulsion to believe it, quite apart from whether or how well it represents objective reality.

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. The brain has surprising artistic license to select connections and make gratuitous ones. If that was all there is to it, we could dismiss religion, magic, myth, and even science as no more than forms of arbitrary human creativity. But animal faith adds a dimension. Realness implies the need to take something seriously and even literally—to believe it so—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 credits realness where it chooses; the dilemma is that it sometimes chooses 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. 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 the literal truth does. While nature *permits* a latitude of fancy in how we perceive, the longing for freedom *motivates* us to be fanciful. I believe this helps account for the prevalence of magical thinking throughout human existence. Ideas can be unrealistic while not hindering survival. This sheds light on the ongoing importance of storytelling, in literature and film as well as the media, and even in the narratives of science.

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, for example, 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 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, which gaps *in turn* must be "filled in" phenomenally, but on a finer scale (and temporally as well as spatially, so that there is continuity of motion, for example). 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 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!

I propose that *qualia*[12] in general involve the same sort of acts of fiat as 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 this

[12] "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]

act, the brain translates linguistic symbols (written or aural) into mental images, thoughts and feelings, or vice versa. It assigns meaning to its own *internal* language, thereby evoking phenomenality[13] in the way that words evoke mental images or the way the continuity of the blind spot is evoked. The self-effulgence of qualities in sensation (such as the redness of red, the hurtfulness of pain) emerges in much the way that the 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.[14]

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 through 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 totality of 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 our own internal connections, to which we do not have conscious access. Yet, it is only from an outsider's perspective that they can appear arbitrary, merely 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, they are a *version* of that information, which an internal agent presents to itself synoptically in phenomenality. Qualia, in other words, are a way the embodied subject first-personally presents to herself information that an observer also might detect by means of laboratory equipment and describe in terms that are third-personal, physical, propositional, and quantitative.

The creation of sensory experience is *like* the creation of meaning from abstract symbols. Specifically, it is like the creation of mental imagery in response to language. Such mental images resemble their full-blown sensory cousins, to a degree of vividness that varies among individuals. Yet, the differences between them provide clues to what is required for actual sensory phenomenality. For one thing, mental images convey only the detail they already embody, based on prior sensory input. Unlike a live sensory image, a memory or visual imagining cannot be searched for more information than it already graphically presents and stores. A retinal image, in contrast, is constantly updated in real time (or nearly), and so 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

---

[13] I use this term, 'phenomenality,' to mean anything that can be consciously experienced—including dreams, hallucinations, mental images, thoughts, etc., as well as ordinary sensory experience.

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

sensory input makes the key difference between a memory or mental image and vivid real-time sensory experience.[15]

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

Animal faith follows inexorably from embodiment. It directs attention outward upon objects that reflect creaturely needs. 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."[16] In other words, animal faith implies the subject-object relationship.

Descartes had rationally (though illogically) concluded from the mere fact of *thinking*[17] that he must exist as a "thinking thing," quite distinct from the existence of his body, which he viewed at best as a machine, at worst as illusory. Had he focused on *feeling* he might have concluded, more like Santayana, that he should identify rather with his biological self and its drives.[18] 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."[19]

---

[15] In contrast to aphantasics (who do not 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 true, life would be very confusing for them! Indeed, it no doubt *is* for some people, labelled schizophrenic, who experience very real-seeming hallucinations. The point is that the intentional system is inherently plastic. If it is typically organized a certain way, it should not be surprising that there is some variation. The default state of the brain is perhaps like a seething cauldron of random potential connections, most of which are usually supressed in favor of "standard" connections that don't interfere with the business of survival. Certain drug-induced experiences support this idea.

[16] Santayana op cit, p214.

[17] 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."

[18] In fact, the Mind-Body Problem may reflect the differing 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 well be the source of the conflictual 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 qualia in particular.

[19] Santayana, op cit, 215-16.

This expresses well the longstanding ambivalence of the human psyche toward embodiment, experienced both as the source of carnal pleasure but also of pain, and as a prison for a soul imagined to be potentially free of material constraints.

The idea of disembodied mind abstracts and reifies the notion of consciousness (phenomenality) as a possibility separate from the life of the body. Human mind allows the indulgence of such abstraction. But experience without a body is an oxymoron. Minds exist to serve bodies, not the other way around; nor can minds exist independently of matter. Like Darwin, Santayana makes clear that our nature is rooted in biology. Our consciousness is shaped and limited by the needs of the body. This has not prevented us from conceiving ideals of "pure" consciousness, "pure" intelligence, and "pure" reason. But, by and large, it does inhibit grasping how utterly parochial our intelligence actually is, and how dependent it is on the body and on animal faith.

One expression of that dependency is addiction. The very concept of *experience*, as a subjective realm distinct from objective reality, introduces and rationalizes the pursuit of experience per se, for its own sake, as something to consume regardless of consequence for the body or the world at large. Addiction is hardly a uniquely human phenomenon; the point is that it is an *animal* phenomenon that takes us over despite our powers of reason and abstraction.[20] Reasons are often no more than rationalizations to justify compulsions. At worst, we simply succumb to the compulsion and its 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, making our animal faith far more insidious and dangerous.

Against common sense, human beings can believe some very strange things. One marvels at the ingenuity of the human imagination—not only the things that make practical sense, like houses, agriculture, machines, technology—but above all the things that make little sense to a rational mind, like gods and demons, ghosts and magic. Yet, religion and magical thinking have characterized human culture far far longer than what our secular culture now defines as rationality. The ancient Greeks we admire as rational seekers of order seemed to actually believe in their pantheon of rowdy and absurdly human-like gods. The Pythagoreans believed in reincarnation and sacred numbers; they used mathematics and music for spiritual training. Plato believed in a metaphysical realm of Ideal Forms underlying material reality. 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. The early scientists were literally Creationists. There are scientists today who believe in the Trinity and the transubstantiation of the eucharist. My point here is not to disparage religion as superstition, but to marvel that superstition is so endemic to the human mind in the first place.

In part, the mystery comes back to language, which confers the nearly magical power of fiat to define things into being—as we imagine and wish them. Outrageous beliefs are possible because a story can easily be preferred to truth; it is, after all, a human creation and not something foist upon us by reality. A story can make sense, be consistent, clear, predictable. Reality, on the other hand, is fundamentally ambiguous, confusing, elusive. Reality only makes *sense* to the degree it can be assimilated to a concept or story. (Despite the facts, it made sense to many ancient cultures that a year should have exactly 360 days.) In general, what we experience

---

[20] 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.

is sensory input assimilated to a story that is supposed to make sense of it, and upon which actions can be based that help us live. The story does not need to be true; it only needs to permit our existence. That gives a wide latitude to imagination and belief.

In a certain sense, no concept or story *can* be true. It can only correspond to reality in the way that a map can correspond to the real territory it represents, or that words can correspond to some reality they describe. Descriptions are always incomplete. No matter how detailed, a map is always selective and symbolic. A photograph is a representation of a real scene, but it is not the scene itself. Unlike the map, a photo seems to be a literal depiction, except that it is two-dimensional instead of three-dimensional. But that resemblance is misleading, because the brain already processes the photo image the same way it processes the optical image presented on the retina, which is also two-dimensional, and which serves as the input to a mapping process in the brain of which we are not aware. That process is *necessarily* symbolic and selective, like the map. It's meaningless to think that neural connections in the brain produce an image in the mind that *resemble* the external world—to which we have no access apart from those processes themselves—any more than words resemble the things they represent. At best, insofar as it facilitates survival, the brain's map of the external world can be said to be *adequate*, not accurate.

Even our three-dimensional perception of space, along with qualia such as colors and smells, is a construction of the brain. The senses are like remote detecting devices wired to the brain, not like open windows on the world. What we experience consciously is like an interactive map, which charts properties indirectly inferred about the territory, from data remotely collected. It has also been likened to the desktop of a computer: what we see are icons, which represent not things in the world but functions inside the computer.[21] In any case, everything about organisms, including their possible consciousness, results from natural selection. While that directly reflects fitness, it is not obliged to reflect reality, provided only it does not prohibit the organism's life as individual or as species.

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. In its awareness (if it has any), the objects of its actions will appear to it, appropriately, as desirable, fearful, disgusting, near or far, etc. It has faith in such appearances, which relate to its needs. It has no direct need for truth apart from appearances. Human beings, of course, 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, but we do not escape the need to believe *something* or perceive or opine something definite. We may be simply driven to higher ground, often more abstract, on which to stake belief. We remain animals, who need to act on certainty.

---

[21] "Properties such as 3D shape and color, for example, are representational formats that have been crafted by natural selection in order to support more effective interactions with the environment. They are part of our own species-specific interface, and not of the objective world. In this sense, evolution can fashion perceptual systems that are ignorant of the objective world because natural selection depends only on fitness and not on seeing the 'truth'." [Prakash, C., Stephens, K. D., Hoffman, D. D., Singh, M., and Fields, C. (2021). Fitness Beats Truth in the Evolution of Perception. *Acta Biotheoretica*, *69*, 319–341] A problem with the desktop metaphor, and the Interface Theory of Perception, is precisely that the icons on a computer desktop do not represent external reality at all, but merely functions within the computer, which presumably stands for the brain in the metaphor. But from the brain's point of view, its "icons" (perceptions) do represent the external world.

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 so 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 whatever the brain tells us. Human freedom consists in the ability to be wrong while utterly convinced that we are right.

We've already mentioned addiction as 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. Of course, we have extended body-related meanings to include intellectual pleasure and emotional suffering. In fact, humans have abstracted phenomenality away from its ties to the body, so that experience has become a sort of private entertainment to pursue for its own sake, apart from its relevance to bodily needs.

There are other compulsions, such as obsessive behavior, even avoidance in place of attraction. And, then, there are also artificially applied stimuli—for example with electrodes to the brain—which can stimulate specific experiences or memories. Similarly, transcranial magnetic stimulation can change your perception, for example altering the apparent color of things or draining them of color altogether.[22] On the other hand, sensory deprivation causes outright 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 of your experience. On the other hand, if you have ingested a psychedelic drug, it may affect your judgment as well as your perception, and you may fail to disbelieve your hallucination. What is helpful to keep in mind is that the brain hallucinates *all of the time*, while *some* of the time its hallucinations are dominated and guided by legitimate sensory input. We then call that 'reality' and feel justified in believing the hallucination.

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

Even in the most abstract realms of speculation, we tend to have faith in our mental constructs. For the most part, that faith is justified, at least as a provisional measure that can be updated by further observation. In the seventeenth and eighteenth centuries, 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 a form of energy—especially the 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, not *substance*. But to reify energy conceptually as substantial seems to be useful even though energy

---

[22] Donald Hoffman *The Case Against Reality: why evolution hid the truth from our eyes.* W.W. Norton and Co., 2019/2022, p11.

is rather a quantity, a function of certain measurements. Even more derivative concepts, like entropy and information, are reified as quasi-substantial, attributed their own causal powers.

To objectify is no doubt a built-in and useful tendency of the mind. After all, our primary orientation is toward objects in space. We literally experience the world as a real space outside our skulls, filled with interacting things. Since language and thought are essentially metaphorical, it is natural (if not logical) for us to think of abstractions—indeed, anything that can be named—as at least vaguely substantial or thing-like. We ontologize everything, more or less compulsively.[23] The fact that this includes 'mind' and 'consciousness' leads to the infamous Mind-Body Problem, as we then ponder what sort of thing mind must be compared to physical matter. Descartes posited a dualism of physical thing and "thinking thing." Others, before and since, have proposed a monism instead: that everything is physical; that everything is mental; or that mental and physical amount to the same "thing." Underlying these philosophical dilemmas, concerning what is ultimately real, remains the fundamental need to believe in *something*.

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

Self-consciousness is the capability to be self-aware: to be conscious, in the moment, *that* one is conscious. This capability adds a critical layer to the projective capacity of the mind, which I call the *realizing faculty* and which Santayana characterizes as a "vital compulsion to posit and to believe." When we're self-aware, we may reflect on how our perceptions and experiences do not fully capture objective reality and how they can be mistaken. We realize, in a general way, that our minds are actively constructing the world we perceive. Knowing this can lead us to question or deconstruct our experience and to allow for the possibility of being in error. While cognition normally assumes that the world is directly and truthfully experienced as it is, metacognition allows us to recognize that our sensory experiences, thoughts, feelings, and interpretations shape what we experience. Animal faith requires taking experience at face value. But, to understand 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 as a direct view of the external world. Since perception can be wrong, this ability to question experience helps us to look (again) before leaping. It helps us to notice our assumptions and revise our opinions, to question our feelings and reactions. While self-consciousness can thus compensate for the default stance of animal faith, it does not prevent the brain's natural commitment to project its simulations as reality. Even when we know that the brain is producing our experience, we still normally *have* that experience. Then, however, we may doubt its truthfulness or adequacy.

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

[23] Just as I am now, admittedly, ontologizing the need to ontologize, as an aspect of a "thing" called animal faith!

paranoia that "you are probably living in a simulation." Here we affirm that you are *definitely* living in a simulation—the one continually supplied by your brain! This is not in itself cause for scepticism, but to grasp that sometimes this real-time illusion serves us better and sometimes worse. It is the *persuasiveness* of the illusion that Santayana questions, and calls animal faith because of its natural function. As embodied creatures, we are here at the price of trusting our senses and our minds, regardless of truth.

We moderns can grasp the biological functions of perception. This 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 realization, tenuous and scarcely to be taken for granted. The natural focus of mind—in the service of survival—remains outward, toward the world external to the dependent organism. We are thus obliged to perceive the world as *real*, 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. But it does add other wiring that can partially compensate.

That compensation is one of the functions of self-consciousness—not in the sense of social awkwardness but as metacognition. We are able to step up a level, so to speak, examining the lens itself as an object, rather than viewing the world through it. This allows us to see the limits of cognition, at least in the abstract if not in the moment. The challenge in real time is more personal and humbling. We naturally enjoy self-confidence and having faith in our view of things. The very seductiveness of that faith calls for 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 can be viewed differently through other eyes. It underlines the ideal of objectivity as a double-edged sword, which can cut through self-deception but also can amplify it when we assume that our point of view is the objective truth.

Objectivity is desirable, of course, yet the natural tendency is to mistake our actual perception for objective reality. In order to maintain this illusion, we then tend to avoid reference to our own subjectivity, protesting that our perceptions and conceptions are objectively true. While the ideal of objectivity is to transcend the merely subjective, this can only be accomplished by claiming responsibility for the subjective basis for one's cognition. 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." But, both literally and figuratively, all views are perspectives from somewhere and are the views of someone embodied and limited. Natural realism trades on denying this fact and ignoring the responsibility it implies. Thus, the challenge is often to question or bracket an apparently obvious point of view.

As a form of cognition, science focuses on a world it presumes to exist independently of itself. 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 the embodied observer, even when it is a public act of communication accessible by others. 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 resembles, and perhaps recapitulates, the brain's natural processes of unconscious inference.[24] The epistemic situation of the scientist is no different than that of the brain's natural situation. It is 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 (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?

We are animals who can conceive being superhuman. Recognizing the limits imposed by physical reality and by our biological nature, we imagine freedom from those constraints and are driven to resist 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; we may realize with some horror that it must apply to oneself if it applies to others:

> "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…"[25]

Scepticism can be personally useful, if inconvenient. We have not only consciousness but also conscience. Regret is a form of moral realization, after the fact, that one did not behave well enough in a situation or missed some opportunity. Often such failure can be traced to a wrong assumption or a biased perception, some knee-jerk reaction that would not have occurred had one been more conscious or conscientious, or had been less judgmental, less caught up in one's own state. Sins against individuals accumulate over time, magnifying guilt. (Perhaps one of the motives for believing in a benevolent personal God is the hope for a consolidated blanket forgiveness.) 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.

Scepticism is especially warranted in a world where information comes to us digitally mediated—no longer face to face but Facebook to Facebook. No longer firsthand but *nth*-hand. The ancient dream is now imminent: to re-create reality ourselves, to convert the found world into a made world, manipulated and refined. The power of digital media to deliberately fool the senses at last fulfills Descartes' paranoid intuition. His solution was to trust that God would not permit systematic deception. Until recently, we could trust that *nature* would not permit it. Now, having abandoned God, distrusting authority, and appropriating the authority of nature, on what grounds can we know what is real?

Mechanisms can now be so complex that they appear natural, even organic. And simulations can be so sophisticated and "realistic" that we cannot see through them. 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

---

[24] Of course, Helmholtz's idea of unconscious inference is a metaphor modeled after conscious inference, then applied backward to the brain's 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.

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

models. Behind that intention is the drive to create artificial general intelligence. Moreover, the distinction between fake and real is only meaningful because we already have a category of the *made* to contrast with the naturally *found*. The power of simulation may already exceeds the natural power of the senses to tell the difference, which depends also on memory. As we become ever more immersed in simulations, will we be able even to remember natural reality?

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

Knowledge is a process and certainty is always conditional. The process involves a dialectical cycle: thesis, antithesis, synthesis. The last term then serves as a new "thesis," beginning a new cycle. We see this in formal knowledge processes, like scientific theory-making. An idea is proposed (thesis) to explain data or to make up for a deficiency in current theory. This idea is published in a journal, for example, which invites 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 the synthesis, which becomes a new thesis to eventually be challenged. Experiment, however, is never totally decisive, but always involves probability. And new theories are rarely universally accepted, even over time.

The demand for proof and for formal decision procedures, as opposed to blind acceptance on faith, formalizes the scepticism and relativization characteristic in ordinary self-consciousness. On the other hand, through the mind's ability to step beyond any defined bound, self-reference paradoxically implies a transcendent ideal of truth. At the same time that the mind insinuates higher absolute conceptions, the sceptical role of subjectivity undermines naive absolutism, in a dialectical cycle, from which there can be no final rest.

Parallel to science, ordinary cognition involves a cycle whose result similarly involves probabilities of error. But the 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. A primary function of animal faith is to ensure this ability to act; if, despite actual uncertainty, our perceptions were not definite, we would be paralyzed by doubt. Yet, the knowledge cycle would be incomplete and less reliable if animal faith alone (the thesis stage) were in play. There must be a balance 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 desirable, the natural tendency is to mistake one's actual 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 one's own views, but to retrenchment of them and scepticism in regard anyone who disagrees. Nevertheless, the fact that opinions can differ plays a positive role in the epistemic cycle. We tend to think of reason as a private tool to evaluate information. But its origin may lie elsewhere, in the ability to convince others; its main use even now may be

to evaluate the arguments of others more than to evaluate one's own.[26] In terms of a thinking process, it may be more efficient to be challenged by others than to challenge one's own thinking.[27]

Whether spontaneous or forced by others, the recognition of one's own error or subjective limits enables mind to evolve at once both toward humble 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.

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

> "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."[28]

In other words, our natural condition as organisms is to see and know what we need to see and know. And this is not simply a matter of selective attention or reduced information flow—a filter between the mind and an otherwise transparent window on the external world. In more modern terms, "each perceptual system is a *user interface*, like the desktop of a laptop."[29] What governs their use is not truth but evolutionary fitness.[30] Yet, the very possibility of scepticism and curiosity renders knowledge open-ended, implying possible greater objectivity.

Indeed, science manifests that search for objectivity, in the name of which it distances itself from the perceiving subject. 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. Science is about general patterns in the world, not about anecdotal accounts or the idiosyncrasies of individual scientists. Protocols must be standardized; observers must be interchangeable. In order to better explore the external world in ways useful to human purposes, these purposes remain unspoken. The observer stands ideally outside the system observed, a fly on the wall.

Science is intrinsically idealizing. The dominance of mathematics (whose essence is idealization) means that physical phenomena are idealized in such a way that they can be treated

---

[26] Hugo Mercier and Dan Sperber *The Enigma of Reason* Harvard UP, 2017, p7.
[27] Ibid, p11.
[28] Santayana, op cit, p.104.
[29] Hoffman, op cit, p.75. The metaphor does not quite work, since the desktop represents functions within the computer, not in the real world.
[30] Ibid, p.55. Hoffman pits the goal of truth against the goal of fitness as competing strategies in repeated simulations, with "fitness" consistently driving "truth" to extinction. 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.

effectively with math: as idealized systems. This aspect of science leads to an analysis in terms of the idealized parts of a conceptual machine. The functioning of the whole is to be understood as reducible to the functioning of these parts. Thus, mechanism and reductionism are essentially by-products of idealization. The natural system is treated as a mathematical idealization. Since determinism is an important property of such systems, real systems are often treated as deterministic, when in fact it is the *equations* used to describe them that are deterministic.

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 makes use of unconscious perceptual models, according to the body's needs and goals. Scientists consciously model observed phenomena, according to society's needs. 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 for its truth value alone, but also for its ultimate contribution to planetary well-being. Good science should support a human future.

Unlike the individual brain, isolated in the skull, science is a collective social process. It is a communication among scientists—a (mostly) polite form of argumentation through which ideas are justified to others.[31] 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. This includes paradigm shifts, but also alternations of more general fashions such as positivism and deductionism.[32]

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 the phases of a cycle. The pendulum of history swings back, fashions return; we move in spirals if not circles.

Throughout history, there has been a dialectical relationship between the playful, embroidering, subjective, ironic side of the human spirit and the heroic, serious, goal-oriented, earnest, realist side. The ironic mentality embraces limits and delights in playing within bounds. It understands limits to be arbitrary, 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, straight-lined, passionately simplistic, rectilinear, naive, square, 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,

---

[31] 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."

[32] 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.

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.[33] 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 ol' days. Instability inspires conservative longing for structure, certainty and control—until an excess of *those* inspires revolt again. Generally, too much of anything breeds its opposite, an aspect of the homeostatic search for balance.

Cycles acted out in time may reflect an endemic logical circularity. If space and time themselves are products of the brain, how can it 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.[34] While science aspires to describe the world objectively, it is another of the brain's constructions, whose fundamental purpose is advantage rather than truth. In trying to picture the unpicturable face of the world-in-itself, we have little recourse but to mistake it for its appearance to *us*. One is forced either to take the map as the territory or to remain silent.

Part Two: The Pilgrim's Progress

Chapter Six: Epistemic Epochs of the Biosphere

> "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 only 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.[35] 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 long preceded the possible existence of life. We could backtrack further, of course,

---

[33] 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.

[34] 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.

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

to imagine the selection of our universe from a multiverse of 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 might include the early shift to downward causation in self-replicators. For, a definition of true life involves bimolecular storage of information in a dedicated structure.[36] This was a precondition for the development of nervous systems. Following the Cambrian explosion, arthropods were the first creatures to leave the highly competitive environment of the sea for dry land.[37] 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 visual 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; and, then, specifically the capacity to self-refer. Among those creatures with vocal calls, the development of fully grammatical language was a major breakthrough. We may speculate that a future landmark in the evolution of mind might be transcending the limitations of biological embodiment—perhaps supplanted by artificial intelligence. Just as warm-bloodedness enabled faster nervous systems than for cold-blooded creatures, so digital brains—if they are feasible—would be vastly faster than biological ones. In any case, since mind is a function of embodiment, its development follows the changing needs and environmental circumstance of particular body designs.

Mind arose as an attribute of organisms that move about to feed on other organisms. It is natural for such creatures to be oriented toward *objects* in their environment. But there are two other types of life on this planet, with different survival strategies: plants and fungi. Plants stay put to transform sunlight, water, air and minerals into their own being. Fungi *wait* for some decomposing organic matter to feed on.[38] In contrast, animals actively pursue their food source, often other animals. All three forms developed multicellularity and became involved in the arms race of mutual adaptation that drives evolution. Mobility creates uncertainty, which broadened the scope of agency and favored the emergence of mind.[39] In the usual view, natural selection is a passive process operating on random genetic changes. But animal agency may have developed as a speedier way to adapt.[40] 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 agency.

All such developments should be viewed in the context of our planet's history, with its multiple mass extinctions, global warmings, and glaciations.[41] That is, evolution is not some kind

---

[36] 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.

[37] Max Bennett *A Brief History of Intelligence: evolution, AI, and the five breakthroughs that made our brains* Mariner Books (HarperCollins) 2023, p157.

[38] Bennett, op cit, p.28.

[39] Phillip Ball *The Book of Minds*, Picador, 2022, p.357 and p.417.

[40] 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."

[41] "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]

of steady inevitable progress. More than 99% of all species that ever lived are now extinct. The present climate crisis and mass extinction differ only in being caused by one of the planet's species, unchecked by the others. Given the overall patterns, our shorter-term prospects do not seem cheery. Yet, some argue, at least in the shorter term, existential risks such as asteroids, gamma ray bursts, pandemics and wars should not be exaggerated, and that past global catastrophes have only stimulated the development of complexity on Earth.[42]

Certainly, the ability to model the environment would have been an important breakthrough. An ongoing internal model of an environment would provide more flexibility than a fixed set of responses triggered by specific stimuli. Judgment and relative choice are implied, along with an organ to exercise them. A model of the world, however primitive, would necessarily be treated as an external space, since it represents the possibility of movement toward or away from objects to which the creature assigns a value. It would be important also for it to tell the difference between moving toward something and something moving toward it. Such a distinction, by itself, does not imply a sense of self, only proprioception. Nor does modelling an environment by itself imply the conscious exteroception we know in experience.

A major landmark was the development of speech and, then, "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.[43] Human infants also demonstrate shared attention and back-and-forth vocalizations with the parents, which other apes do not.[44] Some creatures (for example, song birds) appear to use a syntax that may be simple or relatively complex. Humans seem to be unique in using recursion (subordinate clauses) in the syntax of their speech.[45]

Chapter Seven: Collective Human Breakthroughs

> "Custom does not breed understanding, but takes its place, teaching people to make their way contentedly through the world without knowing what the world is, nor what they think of it, nor what they are."—Santayana

The development of ideas about the world we live in has been punctuated by a series of *grand realizations*. These are collective changes of perspective that transcend current notions and perceptions by situating them in a larger, more adequate context. A grand realization opens to a broader view, abandoning former assumptions, which in hindsight seem limited and naïve. It *negates* a prevalent view in favor of one that is more empowering and objective. Objectivity is an ideal, however, and not the natural goal of cognition. But understanding this is itself an overarching realization. Objectivity is a matter of refining or adjusting beliefs. Grand realizations are *sweeping* adjustments of worldview.

---

[42] 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.
[43] Max Bennett *A Brief History of Intelligence: evolution, AI, and the five breakthroughs that made our brains* Mariner Books (HarperCollins) 2023, p299.
[44] Bennett, ibid, p318-19.
[45] 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.

On the individual level, people may reflect on their own cognitive processes and come to realize how their perceptions are shaped by mental models. The concept of collective consciousness refers to the shared beliefs, ideas, values, and knowledge within a society, group, or humanity at large. It suggests that the minds of individuals are not just projecting their own internal models of reality, but that these models are shaped by, and contribute to, a larger shared framework. This shared mental projection can 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. Individuals in a society collectively project their shared understanding of the world onto the external reality, often treating this projection as objective truth. Whole societies or cultures can undergo collective self-examination, which leads to shifts in how people in that culture perceive and interpret reality. There are both individual and group-level shifts in metacognition, and these shifts can sometimes converge, leading to profound historical breakthroughs. As more individuals within a society begin to engage in metacognitive reflection, their collective consciousness may shift, leading to social movements that change the understanding of reality.

The first grand realization, perhaps in time and certainly in importance, is the idea of subjectivity itself. This is the compounding realization that the world is not just as it seems, and that apparent objects of perception—including other people—crucially involve our creative participation as perceiving subjects. In other words: the advent of collective self-consciousness. We credit the early Greeks with the earliest exposition in writing of this cognitive relativization in the West. In the East, it is the Vedas and Buddhist writings. Later reformulated by Kant, Plato's metaphor of the Cave expounds the problem of cognition: we do not perceive reality directly, nor as it is, but only, as it were, as a shadow cast by real things on the wall of a cavern in which we are as prisoners from birth. A more contemporary version of this simile is that our conscious experience, in the hermetically sealed cave of the skull, is 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. A precursor for such shifts in the modern era was no doubt the development of writing—then printing and eventually digital storage—as an external and public form of memory. Some obvious examples of epistemic breakthroughs, initially in the European context, include the Copernican and scientific revolutions, the Reformation, the Enlightenment, Darwinism, the rise of secularism and human rights, the feminist movement, the civil rights movement, the environmental movement, the digital revolution and the Internet, genomics and AI. We might include the development of relativity and quantum mechanics, the discoveries of galaxies outside the Milky Way, of the expansion of the universe, of exoplanets; the first view of the Earth from outer space and landing on the moon. Future breakthroughs might include the creation of life and mind from scratch, 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 dogmatic lenses toward understanding it through observation, experimentation, and reason. People became more self-aware of their own thinking, how they formed knowledge. In the Enlightenment, critical thinking, reason, and individual autonomy came to the fore 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 comprehensive and instantaneous, resulting in a similar questioning of authority.

Early modern humanism was itself a kind of grand realization. It contested the authority of the Church to rule over life and mind. In religious terms, this was expressed directly in 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. Humanism emphasizes universal human empowerment and self-definition as opposed to prescription and proscription from on high; self-determination as opposed to the authority of kings and popes; and the value of this life as opposed to the next.

The advance of scientific thought has involved a progressive dethronement of Man from any special or central place in the scheme of things. 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 observations implying that the earth was but another celestial orb, like the sun, moon, and other planets. Newton furthered this rebuttal of geocentrism with 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. Such knowledge amounts to an ever vaster perspective, in which we appear diminishingly special. 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 universe appears.[46]

The scientific understanding of the world has been punctuated by well-known paradigm shifts or revolutions, such as the Copernican and Darwinian revolutions. Yet, the history of human cognition is marked as well by grand shifts in thinking about psychology, social and political institutions and practices, religion and ethics, technology, and ecology. All cases are characterized by a widening perspective, but are also a matter of relative and vulnerable 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. Many people may not agree with some of the "realizations" I propose, or about their relative importance. As revolutions, some are incomplete, perhaps failed. 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 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 an essential, 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 constructed within the human mind,

[46] Obviously, other creatures and early humans had sensory access to the "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.

influenced by 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 defy the specialness of membership within a clan, tribe, or racial group. Paradoxically, it now transcends even anthropocentrism, as we consider consciousness and intelligence apart from our species and even apart from biology. In some ethnic languages, the word for one's tribe simply means *the* people—perhaps reflecting a time before much contact with other groups, but also underlining the assumption of self-importance. Humanness was relative. Often enough, members of other groups were (and are even today) not considered fully human. Slavery was widely practiced until the mid-19th century, and considered a normal fact of life, the booty of conquest. The axial religions, which expanded the scope of tribal ethics, happened to arise during increased contact between emerging civilizations, perhaps to facilitate the coexistence of strangers. They taught 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 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 dethroned within the biological world, in which it had been supposedly ordained to reign over other creatures. Even now, this realization does not sit easily with those committed to traditional 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 perhaps in a privileged position to take charge of our fate.

Freud had an influence on the pre-eminence of wakeful consciousness, similar to Darwin's influence on the pre-eminence of Man in nature. Just as Darwin challenged the repudiation of our animal nature, so Freud challenged the exclusive identity of the self with the contents of consciousness—an identification that disowned the "unseemly" aspects of behavior that arise unbidden from some nether region. 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 psychic life. The left hand (or left brain) is not excused from responsibility simply for not knowing what the right hand/brain is doing.[47] On the contrary, Freud's expanded view implies responsibility for non-conscious behavior.

The two great revolutions in physics of the early twentieth century—relativity and quantum theory—are landmark realizations 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 things moving at speeds close to that of light. Effectively, these were realizations

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

that our view and knowledge of the world depends on our dimensions as physical organisms, our state of motion in relation to other things, and on physical media for the transmission of signals.

Within the category of the human, sexual differentiation has meant social differentiation into moieties, long unequal under patriarchy. The dominance by males of women and children (both at one time 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 presumption of "natural" male superiority and right. The feminist movement contests the domination of society by men. As a revolution, however, the program is far from complete, often suffering reversals. Patriarchy continues to maintain its hegemony, primarily through continuing domination of social values, including those embraced by women in the world of global capitalism.

One could say that communism—or socialism or *communalism*—promised to negate the dominance of the many by the few. If so, it was a failed experiment, doomed perhaps by human nature. Historically, even democratic movements have been associated with individualism and private property more than with communal values. Sometimes they were led by the propertied themselves, initially representing a struggle within the upper classes of a society (e.g., a rebellion of nobles against the king). Democratic rights (such as the vote) were only gradually extended to the non-propertied classes and women. The failure of communism is linked to the failure to achieve complete democracy anywhere. As ideals, both have succumbed to greed and the quest for power and status. Controlling elites tout democracy as freedom of the individual, when they mean freedom to enrich themselves at the expense of others.

Recent scientific and technological developments challenge assumptions historically made about intelligence, biology, power, and the environment. They invite us to reflect on our place in the world, on control over our own future and potential to shape the course of history. 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, through genetic editing or digital versions of ourselves? 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. Our survey of breakthroughs could include the institutions of money, debt, and banking; economic globalism; the experiments of democracy, communism, transnational capitalism; and internationalism as transcendence of the state. Mere change of institution must be distinguished, however, from expanding consciousness. Just as there is no progress built into natural evolution, so moral, social, and even intellectual progress is not guaranteed for humans or their possible successors.

Chapter Eight: The Stance of Unknowing

> "It is hard to agree with reality if you cannot agree with yourself."—*Noise*[48]

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

[48] Daniel Kahneman, Olivier Sibony and Cass R. Sunstein. *Noise: a flaw in human judgment.* Hachette, 2021.

imagination. Exploring that zone serves a serious purpose as well as being fun and entertaining. Play practices skills for later life; it prepares the adult to know the difference between reality and imagination and to be able to choose between them. However, the very ambiguity of that zone makes it challenging to always know the difference, while easy to invest the emotional commitment to reality—animal faith—in fantasy as well as in 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. 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; and the 2nd law of thermodynamics abstracts the child's discovery of irreversible events. Just as the socializing child must overcome its natural egocentrism, absolutism in physics and in ethics had to give way to epistemic and moral relativity. In the archaeology of the collective psyche, animism corresponds to the child's magical thinking; philosophy corresponds to the child's acquired ability to think about thinking; and science corresponds to the adolescent capacity to pursue rational goals and logical analysis. In general, one could say that the cognitive development of an individual roughly recapitulates phases of the cognitive and moral development of the species. More particularly, scientific cognition generalizes, formalizes, and abstracts certain developments of the individual and of the species. And so does the moral development expressed in religion. All these cases reflect a transcendence of identification with the body's point of view—but 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."[49]

While the ideal of the spirit may be freedom from the limitations of embodiment, that is a program rife with contradiction, since the program itself is a product of the embodied mind.[50] Because self-transcendence divorces the mind from the concerns of the body, it is dangerous to the organism which that mind is supposed to serve. 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.

When the sceptical phase of this epistemic cycle is pursued out of that context, it can lead to self-contradiction or contradiction with the premises of life. The contempt in which many religions hold the body reflects the ego's dislike for the unsavory aspects of embodied

---

[49] Santayana, op cit, p215.

[50] This circularity arises when the domain that is the output of a process is recycled as its own input. This occurs, for example, when the physical world that appears in consciousness must be presupposed in order to explain this very appearance in consciousness. The world as we perceive and conceive it is taken as the causal basis and input for the brain processes that produce that perception and conception of the world. This dilemma (the problem of cognitive domains) concerns every form of cognition, including scientific theorizing and moral reflection. For, a mind has only the realm of its own representations (whether perceptual or conceptual) in which to speculate about possible other realms, such as an objective world or an absolute truth.

experience, the sufferings of the body: vulnerability to injury, disease, hunger, discomfort and pain, loss and mortality. "Spirit" is imagined to occupy a realm free from these experiences. But all experience is necessarily embodied and normally serves the body. The very existence of the self or ego poses a dilemma: the psychological self is a function of the body, like the CEO whose job it is to manage the financial health of a corporation; but, like many actual CEOs, ego may come to take improper advantage of its position of power. This self may think of the body as there to serve and entertain *it*, and resent subservience to the body. On the other hand, it can be useful to the CEO (and therefore to the corporation) to have some independence from demands originating lower down, say from shareholders clamoring for immediate or excessive dividends, or subordinates competing to promote the interests of their department. Someone needs to take the long-term view, the larger picture.

In that context, a useful epistemic cycle begins with identifying a problem or goal. A phase of brainstorming follows, in which random ideas are put forth. These are sorted, tested and critiqued, from which a synthesis or resolution may result, a program for action. While this can be an unconscious process for the individual, it can also be consciously thought out. Either way, the object is to filter out extraneous or biased 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.[51] That means *resisting* animal faith, which tends impetuously to leap without looking and to hold its truths to be self-evident. This does not mean *abandoning* animal faith, which is required for life, but being somehow able to suspend or bracket it.

Brainstorming involves the key phase of deferring judgment in order to come up with as many creative suggestions as possible. It can be used in collective processes and can also be used personally to cultivate openness to new ideas. 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 tool to an end but a temporary relaxation of purpose.

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; 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 actually a positive asset. I call this willing suspension of belief the *stance of unknowing*.

One must step back from apparent truth in order to see it as mere belief. One must step back from the desiring 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, it does not necessarily mean rejecting them. The stance is a provisional measure, a voluntary act, an experiment whose result cannot be predicted. It is 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 blindered 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,

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

and the seeming urgency of 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.

Chapter Nine: Self-remembering

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

When we're self-aware, we can reflect on how our experiences and thoughts might not fully capture objective reality, realizing that our minds actively construct our perception and conception of the world, including our self-image. Perhaps the most daunting realization is that meaning does not reside in the external world at all. Meaning is not about things per se but about our relationship to them. Just as words have no inherent meanings, but only those we give them, so everything has only the meanings we lend. This realization negates the comfortable assumption that the source of meaning lies naturally outside oneself, imposed or endorsed by "reality." It goes directly against animal faith, our biological conditioning as organisms highly dependent on the external world and appropriately tuned to it.

Meaning is a biologically-based construct of the mind, whose propositions we agree to, if not consciously, as we do to the symbols called words. Yet, precisely because meaning is, in that sense, arbitrary and not determined exclusively by external reality, meaning is potentially a matter of free choice. We may be reluctant to claim that freedom of choice and its attendant responsibility because it is supremely intimidating. One may be understandably reluctant to admit existential freedom as a real option, *preferring* to have one's actions and thoughts determined by external reality and the compulsions of biology, which is the natural way for organisms. For that reason, there may never be agreement that religion is a delusion, that there is no soul or life after death, no spiritual authority or heaven or hell, no absolute source of meaning.

While self-awareness has practical benefits, it can be sobering and onerous. The realization, for example, that the *self* exists only as an experience, and not as any sort of permanent substantial entity, can be as disconcerting as the realization that there is no God. The self is a *brain function*, part of the virtual reality that will cease when the body dies. To be sure, in addition to input from the external senses and the inner dialogue we call thinking, there are bodily sensations that constitute a *sense* of self. At least when not in deep sleep, there is "something it is like" to be you. One's consciousness, and the sense of being alive and existing, is treasured for its own sake, as dearly as life itself—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.

What is the relation between this putative self and the real body? In common parlance, one *has* a body that one can speak possessively. The implications are manifold: "I" *am not* this body but *occupy* it. Perhaps I *own* it, am its *master,* and possibly could wander from it, survive its death, or occupy another body. Nevertheless, I am attached to the experience that comes to me through this body. But, since the body is natural, the self must be 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. While we are aligned with its interests, yet we do not completely identify with them or with the body.

This ambiguity is the source of much trouble. 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. It can believe these inputs exist for its benefit. The self can claim to be the agent that directs the body's behavior, even against the body's natural interests. To that extent, the self may seem a usurper, a natural function gone rogue, more like cancer than breathing.

Some traditions point to this dilemma by denying that the self "really" exists. Others distinguish "ego" from the "transcendental observer," or the social persona from the person's "essence." Many religions objectify this essence or observer as an entity, the soul. But, if one supposedly *has* a soul, like one has a body, who or what is this *one* who does the having? While leadership can go to the head, on the positive side the inner CEO *can* be a leader who transforms the very nature of the corporation for the betterment of the world. The challenge to the spiritual aspirant is not to deny that upper management exists, much less to eliminate it, but to educate it and enlighten it in favor of a broader view that pursues the interests of the whole, not just those of the corporate shareholders. That is the dimension along which the pilgrim's progress should be measured, not in terms of some notion of personal salvation, liberation, or enlightenment. Questions like '*who are you?'* must be considered in that light. "You," in any worthy sense, exist to the degree you consciously serve an objectively worthy motivation. You are an illusion to the degree you chase after illusory goals.

Self-awareness fosters freedom of choice within, which enables one to think and act non-mechanically without: 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 being indifferent or blasé, but patient, taking a long-term view. Success or failure should not be taken more personally than the outcome of a chemistry experiment. But, of course, we normally *do* take things personally—when the underlying belief is that events (or the experience itself) are crucial for one's personal well-being. Significance is normally judged in relation to self. However, it is not what happens to you, or 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 own being, so that you are better able to act for the greater good. Such inward action is categorically different from externally-oriented actions to get the satisfactions we want from experience. 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. The satisfaction it seeks is to be able to act independently of pressures from biological or social programming, or fear of others or consequences. The goal is not to be natural but to be intentional, self-determining.

Momentary experiences of "awakening" 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 sought out and cultivated, such intrusions tend to be subsumed as part of the flow of interesting experiences. But, when deliberately pursued, the

exercise of intentionality may change one's being. *Self-remembering* seeks that sense of self-conscious awakening.

What exactly happens 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. 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, a reminder of one's existence. There is a sense of *I am*.

Try as one may to self-remember, one inevitably forgets. So strong is our conditioned nature and the force of animal faith, so intoxicating the addiction to experience, that it is simply too hard to remain "awake" all the time. Furthermore, it isn't necessary and one also 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 contradiction, made tolerable only by patience. Just as one is more vulnerable during literal sleep, so there is the danger of automatic behavior while figuratively sleepwalking, which can lead to mistakes, regretted words or actions. Yet, that danger is only recognizable because of the commitment to mindfulness.

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 or self-remembering), 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, there is a consumer's attitude toward a quality, state, or power—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. At last in possession of the truth, one may feel superior to the uninitiated. Such prideful feelings of accomplishment or status distract from the real issue: *to be or not to be conscious, right here, right now.* The risk is always that the spiritual path itself becomes a new entertainment, a new form of sleep.

Part Three: Possible Mind

Chapter Ten: The Space of Possible Minds

> "An immortal amoeba simply would never have evolved eyes."—A. Kershenbaum[52]

Especially since the dawn of the space age, people have wondered at the possibility of alien life forms and what sort of minds they 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. To explore the 'space' of possible minds suggests an abstraction like phase space or state space—a reference frame for locating minds in

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

terms of specific parameters. There may be attractors in that space, so that mind is not uniformly dispersed within it.

Such abstraction needs some bounds. 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, and gods.) Embodiment is here understood as a relation of dependency of an autopoietic system upon an environment, which excludes existing AI. The basis of embodied mind is *valuation*: the ability to evaluate stimuli in terms of the needs of the autopoietic system. 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. This raises the question of whether embodiment, thus understood, can be simulated.[53]

We have only our own human minds with which to imagine the space of possible minds, which could include animal minds, extra-terrestrial minds, and artificial minds. Ostensibly, all minds have in common the physical laws of their environment, with possibilities of action within it. However, any notion of 'physical law' or 'environment' would be the concept of a particular mind. Since, as Kant recognized, no mind has access to the world-in-itself, we cannot assume that other creatures or aliens would conceive such things in the way that 21st century humans do.

The search for extraterrestrial intelligence and the quest for artificial intelligence both demand clear concepts of *intelligence*. Similarly, to consider the range of possible minds demands clarifying *mind* as a fundamental concept and, more broadly, the concept of *agency,* within which mind must be situated along with related concepts like *goal* and *intentionality*. Terms and associated connotations for these concepts vary across languages and cultures, and even within the scientific community.[54]

Despite its currency, intelligence remains an ill-defined and controversial notion.[55] It has variously been defined as the ability to learn, to solve problems, to deal with novel situations, to adapt to insufficient information, to do abstract thinking, to communicate, and even as the ability to score well on intelligence tests! It is commonly used to mean 'goal directed adaptive behavior,' but can refer ambiguously to actual behavior or to an internal *capacity* for it. (It is even used loosely to mean *consciousness*.) Like a generalized notion of mind, intelligence is an intuitive abstraction that is grounded in human experience. Paradoxically, it is free in theory from the limitations of the particular embodiments that are the basis for the abstraction. It is understandably an anthropocentric notion, derived historically from comparisons among human beings, and then extended to include comparisons of other creatures with each other and with us. Though it now includes comparisons with machines, intelligence in AI is extrapolated from biological and human origins as an ideal that tacitly guides research. In contrast, we will use 'intelligence' to mean, quite simply, the ability of an entity to maintain and preserve itself. Given that the Aristotelian "final cause" of life is its own continuance, all living things are

---

[53] Can embodiment, as a relationship, arise only through a process of selection in the real world, or can that process be virtual (computational)? To put it another way, can a mind evolve *in silico*, then be loaded to a physical system as in the case of a robot? An analogy might be a living brain hypothetically developed in isolation, which *then* is connected to a living body. While this doesn't happen in nature, could it in principle?

[54] English language users, for example, should not assume a unified or universal understanding of what constitutes '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]

[55] 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]

tautologically successful, therefore intelligent.[56] While this entity need not necessarily be a living thing by current definitions, it must be an *autopoietic system*—a system that self-defines, self-creates, and self-maintains. If it includes self-reproduction, that could stand as a definition of life, since it could not arise naturally otherwise.

We shall define 'mind' as the cognizing aspect of an autopoietic system. This does not necessarily imply consciousness, but is simply the ability of a system to detect and respond to stimuli on its own.[57] On the other hand, mind does imply *agency*, and we define an agent as a system that acts on its own behalf as distinguished from merely reacting to causes. An entity is an agent if and only if it acts on its own behalf, originating action with its own self-renewing energy and for its own purposes—which primarily concern its own well-being or that of its kind.[58] In other words, it is effectively an organism, an adaptive autopoietic system whose chief product is itself.[59] Agency implies a subject-object relation between a bounded physical system and an environment. While the concept of *intentionality* has a long history in terms of linguistic reference, here we define it as a relationship between subject and object, mediated by internal connections or operations performed by an agent—mappings, in the mathematical sense.[60]

Glib language permits one to speak of programming AI tools to "have goals." But clearly the goals are those of the designer or user. As far as it is significant to humans, the intelligence of other entities (whether natural or artificial) is measured by their capacity to further or thwart human aims. An AI might be trained to pursue certain goals or to uphold certain values. But rewards or punishments 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. While it might seem desirable to program a system to pursue any arbitrary specifiable goal, 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 those of other human beings. While it is attractive to conceive of an ideal agent pursuing arbitrary goals, the goals of living things—including us—are hardly arbitrary.

As a landmark of AI, even human-level intelligence would not yet be artificial *general* intelligence (AGI), which is theoretically devoid of the constraints of human or any embodiment. The development of AGI is motivated by supposed benefits of increasing abstraction and generality, leading to ever greater capability: the hope for a subservient super-slave. It is a paradoxical goal, because it seeks a tool that will obey the user while creating a super-agent with a will of its own. And, if AGI means "universal" intelligence, it would be so general as to be independent of the constraints imposed by embodiment in any form we know. Such constraints are the very basis of intelligence and meaning as we know it, so it is difficult to see how such an 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. Yet, it would be far more alien than

---

[56] 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 is something's reason for being or ultimate goal—which, in the case of life, is its very existence.

[57] 'Sentience' or 'sensing' here is an implicitly behavioral or third-person descriptive concept. For the moment we leave moot the question of the first-person experience of such a system.

[58] 'Agent' is used rather loosely in the AI literature and media to mean anything that does something; furthermore, the use of 'agential' suggests degrees of agency. Here we consider agency in a more restricted sense.

[59] In contrast to an *allopoietic* system, which produces something other than itself (e.g., a tool or factory).

[60] In this sense—of a relationship between subject and object—intentionality stands in categorical contrast to *causality*, which is a relationship between objects, as noted by a subject.

any sci-fi vision of organic alien life forms, which at least presume embodiment through natural selection.

One can contemplate the oddity of being a particular human individual—a consciousness identified with *this* body, at this time, among all that have ever existed. One could extend that reference class to include all the types of animal bodies known and the minds that serve them. One could even expand the exercise to imagine being one out of all theoretically possible minds. The range of possibilities would obviously include being singly embodied—that is, a mind controlling a single physical organism.[61] But it might also include multiple embodiment—a mind controlling a number of distinct organisms or systems, all at once or in succession. (Alternatively, a body could have multiple back-up copies of its mind.) Finally, there is the possibility of multiple minds controlling a single definable organism or system.[62]

The concept of possible mind stimulates many questions, such as how does the space of possible bodies shape the space of possible minds?[63] And 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? For example, *fins* of some sort would be an obvious development in any fluid environment through which it would be advantageous to move quickly.[64] Similarly, a brain located close to major sensing organs (eyes, ears, olfactory) is a practical arrangement. In the animals we know, these sense organs are also located close to the mouth, reminding us that such features evolved in the context of a metabolism that depends on devouring other creatures. They exist to serve the organism in that competitive context, without which they would have no reason for being. To the extent that physiology must follow from the basic evolutionary context of life, then so must 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 seen can be eaten… while life lasts, in one form or another this faith must endure."[65] The "creed" does not in itself invoke consciousness for its enforcement, but is a matter of observed behavior. It applies to the forms of life and mind we know. Yet we are free to ask whether a mind could exist to which it does *not* apply.

For humans, at least, thought is inextricable from language. But other creatures communicate without a formal ("fully grammatical") language.[66] Sound is effective for communication across distances, while visual signals are often effective closer up. Human

---

[61] The very idea of a brain controlling a body is a generalization based on common experience. Extending this notion to an artificial brain controlling an artificial body (robot) ignores the biological context in favor of an anthropocentric ideal of control systems.

[62] Octopuses and human split-brain patients are examples at hand. Group mind or "hive mind" only makes sense if the associated individuals virtually constitute an organism. 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/).

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

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

[65] *Scepticism and Animal Faith*, p180.

[66] "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.

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?[67] 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 communicate with us). However, *our* mathematics is a human creation, based on experience of environments with both discrete objects (hence counting) and fluid media (hence mathematical continuity). That may presume 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. If we suppose indefinitely many possible minds, some 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 understand.[68]

The notion of omniscience is an ancient human ideal, again extrapolated from common experience (for example, being able to see something that another observer cannot see.) As we have defined mind, however, omniscience is not feasible: every embodied mind will be finite and have a (more or less) limited perspective. Yet, that leaves room for degrees of *relative* knowledge superiority. If an omniscient being is a hypothetical *perfect knower,* then perhaps we can also imagine a hypothetical *perfect doer.* Such an omnipotent agent, if embodied, could not act or create arbitrarily, however, since it would still depend for its existence on the material universe.

Just as there are physical limits on the size of Earth-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 as fast as a human brain.[69] Present-day computers dissipate orders of magnitude more energy than a living cell: biology is computationally efficient, for its purposes.[70] Indeed, the technological trend is to imitate biology, not only in terms of organization and function but also size and efficiency. To replicate in silicon chips 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, and anticipate that advanced alien civilizations would have done precisely that: gone ultimately small, for the power of computation afforded, instead of expanding into galactic space.[71]

---

[67] 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.

[68] Roman V. Yampolskiy "The Space of Possible Mind Designs" 2015, sec 1. Cf. also Ross Ashby's "law of requisite variety" (that a mind must be at least as complex as what it hopes to understand).

[69] Max Tegmark *Life 3.0: being human in the age of artificial intelligence*. Vintage Books, 2017, p153.

[70] Phillip Ball "Organisms as Agents of Evolution" John Templeton Foundation, April 2023, p22.

[71] 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 Eleven: Human Alternatives

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

A human individual appears to herself as an integrated whole, with more or less free will. 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 this coalition came about is the subject of evolutionary theory. Obviously, there must be advantages for individual cells to be part of a multi-celled creature. But as with all coalitions, this does not imply the perfectly harmonious functioning that we associate with a machine, whose parts and functioning are well-defined. On the contrary, it suggests a political metaphor: the "society" of cells composing a body is the net result of many competing wills, which manage to cooperate—despite their differences—well enough to give the illusion of an integrated whole for a limited time. The conscious self that acts on behalf of this body takes this integrity for granted until something goes wrong with the appearance of normal functioning. One then subjectively experiences this as some form of discomfort or unwellness, even pain. Viewed externally (by a doctor or scientist), this dysfunction is labelled disease, which the patient experiences as unease.

From the externally-oriented viewpoint, we understandably look for a pathogen (a foreign agent) as the cause of the dysfunction. Society does this too. In both cases—human society and the society of cells—potential enemies and even traitors lurk not only outside the body politick but also within it. In both cases, the integrity of the whole is vulnerable to attack. But it is also fundamentally questionable to begin with. If indeed the body were a machine, there would be no other way to explain dysfunction than either by the wearing out of parts or by the intrusion of some foreign matter, like grit or water, into a normally well-oiled engine. The mechanist metaphor has served well in many ways, even applied to the defenses of the organism against foreign agents such as microbes. But it is also misleading, insofar as it suggests some crisp deterministic way for the immune system to identify what is "foreign," in contrast to what is "self," with the aim to rid the body of unwanted intruders.

In fact, the immune response is better explained on the political metaphor. The body normally consists of a hodgepodge of cells, some of which belong to it genetically, but far outnumbered by others 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 regime. But there can be native born traitors as well—such as cancer cells, which refuse to know their place and go rogue. 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, 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 types of cell. They are somewhat unpredictable, and may or may not act in what medical observers may presume are the body's interests. Some cells might even act like criminals or double agents. What this implies, moreover, is that the very identity of the body is fundamentally in question: what is self and what is other?

From the point of view of the embodied self, the body might be considered less a smoothly self-maintaining machine than a nightmare of seething contention among combatants.[72] To the extent that health or unhealth is a statistical effect emerging from these cellular micro-interactions, it is the luck of the draw. In view of the body's unfathomable complexity, either metaphor suggests that the body's proper functioning could even be considered miraculous. 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, complaint is based on an unrealistic expectation. The body is not designed to function perfectly or to last forever (indeed, not *designed* at all). Not a product of rational thought, it is something that persists—as long and as well as it does—because it *can* within the complex game of natural reality. In other words, it is a product of natural selection, which promises only the longevity required to reproduce. Certainly, it does not promise freedom from pain or discomfort, which are rather built-in side-effects of the body's ability to self-repair. 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 efforts to self-maintain. But, like the rogue cancer cell, this self has always had ideas of its own.

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 mountain or river spirits, or in invisible gods, demons, ghosts, and souls. The mechanistic program of modern science has tried vainly to counter what it considers superstition, at the cost of altogether denying agency in the physical world.[73] This has hardly prevented the majority of human beings, past and present, from claiming religious beliefs. Even if religious hopes are delusional, they are sincere and long-standing, based on a fundamental disaffection with life in the body. We are not reconciled to the body's sufferings of injury, disease, and finally death. Most religions deny death in favor of some sort of continuation of the self or consciousness. Some promise resurrection of the perfected body in a state free from suffering—a notion that can only be based on an illusion of the body as truly integral and effectively non-material. While such hopes are not rational, the force of animal faith behind them is ironically grounded in the reality of the mortal body.

Science may have divorced itself from its religious origins, but not entirely from the ancient pretention to immortality and to freedom from the chains of embodiment. The new faith is that, while these goals cannot reasonably be expected through religion, they can nevertheless be achieved through technology. We once imagined gods as super-agents; now we imagine ourselves, or our creations, as super-agents with god-like powers over both the external world and our own being. On a deep level, we aspire to be no longer under the thumb of nature, but self-determining, self-creating, self-defining.[74]

We struggle against this world we find, 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 than raw nature. The raw nature within us, however, has

---

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

[73] Ironically, the program has religious roots—that is, roots in superstition itself. Nature was deemed passive and inert in order to reserve all powers of agency (besides human agency) to the single divine Creator.

[74] With a certain irony, this is the definition of autopoiesis, except that organisms have their relative autonomy in the context of being products of nature, under its thumb.

scarcely changed. Aside from age-old breeding practices, the means to change it deeply has not existed until now. For good reason, perhaps, the *idea* of changing it through technology has been taboo. For, it thrusts upon us the dilemma of defining what we should be—a task that has heretofore been left to nature, to accident, or to divine will. As natural creatures, we could continue to go the way of nature, dictated by forces beyond our control, resigned to the fate of inevitable extinction. As unnatural creatures, however, we can imagine mastering the forces that control us, turning the tables on nature. We can even imagine ourselves becoming the benevolent beings we idealize in religion. Along with the quest for power, we imagine also the possibility of *eutopia*, a world (and a body) expressly designed for our well-being and happiness.[75] Since neither God nor nature provides such a world, it is up to us if it will ever exist. Indeed, the human project seems to be to create such an ideal world, despite the handicap of being nature's limited puppets and the disastrous setbacks we know as history.

We are on the threshold of the sort of genetic manipulation that can redefine the human organism. Alternatively, digital technology promises at least cognitive enhancement. Beyond that, it may seem to offer immortality as disembodied virtual existence. However, while a real organism can experience a virtual reality presented to it by a computer program, a computer program cannot experience anything at all! Real subjects can have virtual experiences but virtual subjects cannot have experiences at all. Moreover, AI can animate a robot, and could *imitate* a given person; but it could not *be* that person or the continuation of that person's consciousness. If it could be conscious, as an artificial organism, its consciousness would be its own, as a 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 can be manipulated, copied, or destroyed) by real physical agents. That would gainsay the ideal of self-determination and the very concept of individuality. Would it be better to have human or AI overlords defining the simulation you live in, rather than nature or the gods? If there were clones or back-up copies of you, which one would be the real you? The idea of digital immortality is absurd, if only because "you" are not a program that can be abstracted from your body. 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).[76] 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 completely separable from the hardware. It gives rise to premature considerations of the moral standing and ethical treatment of AI entities. Yet, the vainness of the personal hope for digital immortality does not mean that artificial mind cannot exist. Like natural mind, it would be physically embodied, not merely virtual. It would be who it is, not someone else.

The prospect of a post-biological human future might horrify some, with only non-human animals left on the planet to "feel the spark of insight, the pangs of grief, or the warm hues of a sunrise."[77] But phenomenality must not be identified exclusively with biology, nor does post-biology exclude embodiment. Furthermore, what exactly is lost, apart from the functionality that consciousness provides? What does it mean to us to be conscious and why do we value it so?

---

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

[76] 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. The conclusion may be valid, but the premises are false.

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

Chapter Twelve: Artificial Mind

> "We have recreated ourselves in the image of our tools."[78]

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 means. One thing we can be sure of is that AI blurs the distinction between natural and artificial. This fulfills a long-standing human intention: 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 does not really exist. The question hinges on just how closely an artificial brain, for example, must resemble a natural brain in order to "duplicate" it and thus have the properties we associate with mind, which may or may not include consciousness.[79]

Mind as we know it is a system to deal with the world; that may or may not include having a concept of 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. What would it mean for an artificial mind to deal with the world, if not for the purpose of its own self-regulation and maintenance? What would it mean for it to have a concept of its world not based on its own existential stake in that world?

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. Whether that is desirable, from a human point of view, is another question. Whether an artificial mind could or should be *conscious* (i.e., experiencing its own phenomenality) is yet a further question. For, if human attention span and working memory depend on our consciousness, the slowness of these features might be a handicap to avoid in artificial mind. On the other hand, if attention, working memory, and consciousness all depend on the specifics of a biological brain, these properties might be vastly enhanced in a system that can operate at the speed of light rather than the speed of chemical impulses.

---

[78] 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."

[79] The notion of uploading or downloading a '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. The fact that the robot then "behaved like a worm" confirms the validity of the model, but can be misleading insofar as "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?

These are questions that seem to arise only as AI itself emerges; yet they are so fundamental—and urgent for human safety—that logically they should be answered beforehand. The concept of artificial general intelligence (AGI) stands in an unclear relationship to the concept of artificial mind. As a goal, AGI reflects the human aspiration to create systems that match and extend human power—essentially as tools. However, the wide-ranging capabilities sought imply agents that are not tools, but tool users in their own right, which could outcompete with us for natural resources. 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).

In contrast, artificial mind may reflect a different motivation. Fascination with the possibility to do all that nature can do has been concurrent with the emergence of our self-consciousness as a species. (This 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 artificially. To paraphrase the Early Modern thinker Vico, we truly understand only what we ourselves make.[80] So, artificial mind merges with the project to appropriate natural powers by re-creating the natural world at large, including the creation of artificial life forms. One must wonder at the wisdom of creating an ecology of artificial creatures that could displace the natural ones on this planet. The premises of AI, robotics, and genetic engineering tend to gloss over that risk.

Perhaps because of the current craze for large language models (LLMs), phenomenality ("consciousness") has become a false standard whereby to evaluate AI—in the form of moral concern for potentially sentient beings.[81] The concern is that if an AI is conscious, then we ought to have the same moral concern for its experience as we do for human beings and other creatures. But, the first-person experience of pain, pleasure, fear, suffering, etc., is an organism's way to represent to itself its own state or situation, which can also be evaluated and appreciated independently by third-person observers. Either way, the real issue should be the *state* of the organism, rather than anyone's perception or evaluation of it, whether first-personal or third-personal. Though we blithely think of putting suffering creatures "out of their misery" when we can do nothing to improve their state, the issue is the irreparable damage or injury, not merely the suffering that results. Moral issues concerning possibly sentient AI should likewise focus on positively securing their wellbeing rather than precluding any possible negative experience they might have of their state.

The issue of moral concern and legal rights for AI, on the grounds that might be conscious, reflects a general muddle-headedness concerning the role of consciousness in human beings. It also reflects our extraordinary attachment to our personal phenomenality. We are concerned for the experience of other beings (including artificial ones) because of our ingrained concern for our own experience and that of other people. While this represents an advance over callous *lack* of concern, it is a cultural product with a checkered history and an uncertain future, a current form of social correctness. It may also reflect a narcissistic fascination with subjective experience characteristic of our age.

---

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

[81] 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." Patients who remain conscious throughout medical procedures despite anaesthesia pose a moral and legal dilemma to the medical establishment. The technical difficulty is that the anaesthetic used could effectively paralyze the patient, who cannot respond to protest pain, without rendering them insensitive to it. One clever solution is to add a chemical agent that causes the patient to *forget* their experience when the anaesthesia wears off.

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. On the other hand, we seem to *want* them to be so, to have like us the opportunity to suffer, which is the concomitant of pleasure. Are we simply lonely on this planet, seeking artificial companions, just as we quest to find signs of intelligent life elsewhere in the galaxy?[82]

In any case, 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 for machines, the question of artificial mind does not hinge on the medium, such as carbon versus silicon. The point is rather that living forms are *organisms*—that is, organized a certain way. They are embodied products of natural selection, part of an ecology. The question then becomes: is synthetic organism possible; if so, how can it come to exist? Yet, we should also question what synthesis involves, as well as its wisdom as a goal. To the degree that technology attempts to imitate natural processes with ever greater refinement, at what point does synthetic pass over into natural? Is there an identifiable threshold? In particular, could the relationship of embodiment be artificially produced or induced—say, by simulating natural selection? What exactly makes the difference between a natural process and a simulation of that process?[83] Is the simulation of an organism an organism?[84]

An obvious difference is that the simulation is a program 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 digital computer, with the laws of physics its algorithms. Such metaphysical extravagance arises, perhaps inevitably, from the recursive dilemmas posed by consciousness to beings conscious of their consciousness.

The idea of simulation now refers to digital computation, but the association is not intrinsic. The computational metaphor helps us to understand the epistemic position 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. But can the metaphor be turned the other way 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 its own benefit and use, not for that of external programmers or users. It is no longer a tool but its own person.

The concept of artificial mind is at least mildly paradoxical. For, it is motivated in part by the rejection of embodiment—that is, of unsavory aspects of our biological existence and the animal faith that enforces them. Yet, the very concept of mind is based on that biological nature;

---

[82] An irony of digital technology may be that, even as social media magnify communication, we become socially more isolated behind our screens.

[83] Could a virtual computer, running in a physical computer, satisfy the condition of embodied autopoiesis? That condition is a relationship of an organism with its environment for the purpose of maintaining itself. To maintain itself in this instance is to maintain the "hardware" of the virtual computer, which must be part of the simulation. In other words, to simulate embodiment, it must be possible to simulate organism itself.

[84] 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.

we reject animality, but is there any basis for mind without it? It is one thing, of dubious value, to re-create a natural mind. It would be quite another thing to realize an *ideal* mind that is freed from the flaws of natural mind. Certainly, that has been one of the age-old aspirations of religion. Can we hope to achieve it through technology?

Intelligence is often defined roughly as the ability to solve problems or to set and achieve goals. That leaves a great deal unspecified. *Ideal* intelligence might be better defined as *wisdom*. For, the problem with goals and problem-solving is their narrowness: the very specific ends desired and the parochial motivations behind them. Ideally, there should be only one goal and one motivation: 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 do not coincide 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 (or perhaps because of) the diverse goals of competing and cooperating species.[85] 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 cells that do their duty unflinchingly within the body politick. In modern individualist society, we have rather the opposite idea, that the whole exists for the sake of the parts. And yet, no matter how sophisticated and complex, these parts (i.e., ourselves) are each bound by natural laws and drives. We are but individuals of a kind. Can there be a mind or form of intelligence that is not so bound?

Even the concept of AGI is anthropocentric: the idea is to match human intelligence and capability. Furthermore, no matter how generalized, the very concept of intelligence derives from human experience. Above all, unless it constitutes an entity with its own purposes, any form of AI remains essentially a tool to enable human purposes. But there are no *general* human purposes, only those of specific individuals, groups, corporations, nations, etc., in specific contexts. It is tempting to dwell on the expansion or extrapolation of capabilities, as though they had some objective significance. The more important questions are what they will be used for, and by whom. We can imagine ever more powerful AI, but either it will be a tool for some user or will itself be a tool-user. If it is a super-agent in its own right, it will hardly be under human control. What would its priorities be if not self-preservation?

On the other hand, an AI might be better at running the world than human beings—if that were somehow its priority. If, for example, a super AI *identified with the planet*—as humans identify with their bodies—its self-preservation would mean preservation of the planet. Human beings would fall under that umbrella of protection only to the extent they figure as an integral part of the planet (and not a threat to it, as we now seem to be). Would bringing about that state of affairs serve as an example of wisdom, from a human point of view? One can extrapolate the ideal of such global control beyond the planet or solar system, to include the galaxy and beyond.[86] But what would it mean to regulate the whole galaxy or the whole universe? Doesn't

---

[85] 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 many competing and interdependent species produce the net result of the self-regulating biosphere.

[86] "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}].

the universe already regulate itself without the help of AI? (Yes, of course, but not to human taste!) In contrast to biologically-determined human psychology, and the boundaries of human reason and perception, AI is plastic and not fixed. Its potential is literally unimaginable. But what, if not human will, would determine what it actually becomes? So, the real quest should not be to develop AI unrestrictedly toward some confused ideal, but to clarify our intentions and refine human will.

Chapter Thirteen: Human Successors

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

Just as the individual is vulnerable and mortal, so is the species. Human beings are now able to create a future for themselves through technology, perhaps avoiding the doom that seems to hang over natural life. This is the latest phase of an ancient project of self-creation. Technology is the latest phase of *culture*, which has been humanity's way to create a favorable man-made environment and to disengage from our beastly inner nature. Technology now proposes to revise the human form and its essence, to free it of the constraints of biology, just as it seeks to free us of the constraint of being Earthbound. 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 encounter from another star would likely be the AI successor of some intelligent species once there.[87] According to this vision, the future of life is non-biological.

AI offers the potential to redefine human being: to define what we want a mature human species to be.[88] Presumably, an artificial successor to a biological life form would preserve some of the biologically driven motivations of its progenitors, while freed from some biological limitations. Aside from being longer-lived and hardier than us, they could also be morally superior, better organized, more peaceful, for example. Yet 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 contrary, evolution adapts "unconsciously" to changing circumstance, and history seems to be a net result 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. On the basis of which present values can we decide the future values of our successors. Their intelligence will at least be adaptive to change (and therefore mutable), but could also leave them as divided as we are.

Through technology, we re-shape the environment around us. Aside from 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 would be one more way for the rich—who can afford self-enhancement procedures—to obtain further advantage over the poor.

---

[87] Susan Schneider "Superintelligent AI and the Postbiological Cosmos Approach" preprint 2016, p5.
[88] Michael Levin "The Space of Possible Minds: Technology & the Human" (Noema podcast, APRIL 17, 2024.

More fundamentally, our brains and senses face outward, to deal with the external world; they were not designed to directly manipulate their own functioning. However, because bodies and brains are part of the external world, their functioning can be changed indirectly by manipulating the matter involved, chemically or surgically. While I may not be able to change my behavior or experience voluntarily, someone else could change it for me by intervening in my chemistry, in my brain, or in my genes. Potentially, they could do so whether I wish it or not. Hence, beyond the conventional subjective ways we know, self-modification is not an individual but a collective matter. While it concerns the capabilities of the individual, it is inherently a social issue. Similarly, our generation could try to determine the experience and behavior of future beings without their consent. The moral dilemma presented inheres particularly in individualistic society; it might be less of a concern in collectivist society. Which raises another point: there is simply no unified "humanity," no proper *we* to make such decisions.

Decisions might nevertheless be made, 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: economic gain and military superiority. The idea of developing a single human successor type, to represent humanity at large, is a pipedream unless humanity can achieve a corresponding unity to begin with. Just as we think of people as representatives of homo sapiens, we can conceive a post-human *kind* with individual variations. But, just as we conceive of races, we can also imagine multiple sub-species of artificial humans going forth competitively to populate space, representing not humanity but *brands*, corresponding to nation, tribe, or corporation. Are capitalism, nationalism, ethnicity or political loyalty fundamental values "we" wish for our successors?

We can envision genetic changes in human biology that meet specific current problems and goals. That could mean better adaptation to space travel, radiation, and weightlessness; or to pollution and climate change on earth. We can imagine android versions designed to be immune to a wide range of conditions that plague biological life. Yet, all physical entities, even artificial ones, are subject to physical laws and forces. Artificial organisms cannot be completely invulnerable; there is no guarantee even that they would be less vulnerable *overall* than natural ones. Even machines in outer space have problems, at risk of radiation or electromagnetic pulses, for example. Natural organisms adapt through generations of mortal individuals—through dying and reproducing. If the ideal successor is to be immortal or very long-lived, how would it achieve this state of relative invulnerability, if not through a similar wasteful process of adaptation through generations of sacrificed individuals? To avoid that, it would have to have individual powers of self-maintenance and self-alteration imagined by Lamarck. No such perfect self-reconfiguring entity could be designed top-down from the outset. Beyond a certain stage, it would have to be self-designing: autopoietic on a higher level than that of current life.

The ideal superior skills would include the sort of self-transcendence that is 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, but relative and open-ended. To the degree that present humans can be objective, there could be consensus that objectivity should characterize our successors. But does the capacity for objectivity and self-transcendence contradict the basic premise of self-maintenance, with its animal faith? That is, can any organism—even an artificial one—be unbiased and unselfish, or does its own existence dictate fundamental self-interest? Is a selfless organism a contradiction in terms?

Certainly, there are examples of altruism in the human and animal worlds. In the latter, at least, it 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 higher causes for which one is willing to sacrifice oneself—for example, one's country. One could argue again that serving the good of the group serves the genome of the species. But what about a point of view that serves all life, the planet as a whole? The solar system, the galaxy? This would seem to be a question concerning what the individual mind identifies with. That might be one's own body, family, tribe, nation, species—or beyond. The survival mandate then refers to *what* is supposed to survive, *whose* well-being is concerned. Nature has programmed us certain ways, but there may be other options.

If our successors have an enhanced ability of self-determination, they will likely not remain what we can envision now. For similar reasons, an advanced alien mind we encounter might be incomprehensible to us. Their technology would, of course, be based on physical principles—but not necessarily 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 described above, our successors would have the best chance to understand them if they too embody it. If the ability to self-transcend implies convergence, it should be a prerogative.

Does that ability 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 species; it may not be involved in reproduction, which relegated to a caste. An artificial individual might be only conditionally committed to self-preservation, subject to higher-level commitments. For human beings, this conflict of interest is the subject of drama and suffering. A tortured spy, for example, tries not to divulge secrets despite inflicted pain. Humans are not normally wired to voluntarily disregard pain signals or fear of death, probably for evolutionary reasons of economy. But an artificial creature might have that capability. It could feasibly have conscious control of its phenomenality.

We can imagine a future society of advanced post-humans, which values the potential and contribution of each individual—though not absolutely, as nominally in Western society. Individuals would be committed to the kind as a whole, or even to some larger whole—not to close kin, tribe, nation or other sub-group. Self-interest would subserve the interest of the whole.[89] Individual self-preservation would be an instrumental goal, toward the final goal of preserving the whole.[90] Concern for the subjective experience of the individual (pleasure, pain or

---

[89] 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."

[90] 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].

suffering) would not be a primary issue, since individuals would be able to control their phenomenality toward the end of serving the whole; they might be able to self-repair in ways impossible for natural organisms. However, such utopian sci-fi possibilities should not be confused with the individualistic dreams of some transhumanists, which imagine backing up the personal mind, for example, and multiple or disposable artificial bodies for personal use.[91] Yet such 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 fellows. All experience would refer for its meaning to the collectivity and its mandates rather than (or as well as) the well-being of the individual embodiment.

Chapter Fourteen: Mind in the Universe

> "Our destiny is density."—John M. Smart

Mind must be embodied, with a dependent relationship to an environment.[92] It is natural for us to imagine mind in the universe evolving like the life we know on this planet, in environments similar to ours. As it turns out, planets now seem commonplace, so that life could be abundant in the universe. Mind, however, may be associated with only a certain type or level of biological organization. The type of mind with which we might eventually communicate would be even more exceptional; it would have to develop the means to communicate or travel across the vastness of space, and want to do so. Since we live in the same universe, we imagine this possibility in the terms familiar to us: through technology based on known physical laws. We thus tend to imagine advanced aliens with concepts, technology, and motivations extrapolated from our own (just as we imagine them having humanoid bodies). While these assumptions are anthropocentric, they are reasonable if implicitly we are looking for mirrors of ourselves. Because we would be on the same wavelength as them (even literally, perhaps), civilizations at our level of development might seem to be the ones we could expect to contact. However, given the accelerating pace of our own technological advancement, the time during which a developing civilization might occupy our level could be very brief. On the other hand, a civilization might never master the energy resources required for interstellar communication or travel. Moreover, like ours, it could be prone to self-destruction through internal conflict, war, and ecological degradation; it could remain vulnerable to destruction by cosmic, geological, or biological events

---

[91] Such as parodied in 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: "…you need only rent a suitable android body in each locale. Airports could become a thing of the past. Bodily harm matters little to you, for you just pay a fee to the rental company when your android surrogate is injured or destroyed. Formerly averse to risk, you find yourself skydiving and climbing Everest. You think: if I continue to backup, I will live forever."

[92] Three common uses of the term *mind* can be distinguished: 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. Here, we focus on mind in the first sense, with which we could potentially communicate, which may or may not also satisfy the second sense.

it could not control. On the other hand, civilizations vastly superior to ours, though more durable, might not share our enthusiasm for contact.[93]

Many stars are far older than our sun, giving plenty of time for advanced civilizations to evolve. Such civilizations have been classified according to the energy resources they could master.[94] Aside from some cosmic version of manifest destiny, a rational motive for expansion, even throughout the galaxy, is that a civilization too localised would remain vulnerable to large scale disasters, such as nearby supernovas or gamma-ray bursts.[95]

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 be likely to encounter; such reasoning offers the reassurance that they would by nature be benevolent and non-aggressive.[96] If artificial successors represent a more robust and stable kind of existence, more 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 brief precursor.

An alternative to expansion in outer space could be miniaturization. This would afford control at ever smaller scales, which would be more effective than communication and administrative control over vast distances.[97] While the micro-scale could represent a new unexplored territory for hyper-efficient computation,[98] hiding in that realm 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.[99]

---

[93] 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 not be interested in contact with inferiors; or they have moved to "inner space"—the realm of the indefinitely small.

[94] 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 their ability to survive disasters of various sorts. See: Galantai, Zoltan (2006). "After Kardashev: Farewell to Super Civilizations". *Contact in Context*. **2** (2).

[95] 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.

[96] 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 not to destroy itself?

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

[98] "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].

[99] "…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 other causal influences as well. Since we could not detect this sort of existence, it has been proposed as another solution to the Fermi paradox.

Natural selection is a slow and wasteful process, in which only a few progeny survive to reproduce out of a generation of massive reproducers. 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 if artificial organisms can adapt by redesigning themselves. The self-contained individual, as a unit upon which natural selection acts, could be rendered obsolete by the 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 individuals.

If life arose naturally in the ways that it could—adaptable but with inherent limitations—to transcend those limitations is arguably the next step for natural evolution. This idea does not imply any teleology at large in the cosmos, but rather reflects human intention on this planet. The idea of a universal mind is an ancient theme. It was first conceived in theological terms (the mind of God), or as panpsychism. Some now revisit the idea in terms of information, considered as the basic ontology of physics. Information 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. 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.

Is the universe an information processing system? That's a loaded question, since none of the terms are well-defined. What is an information processing system, if not an artifact devised by humans (such as a digital computer)? For that matter, what is a *system*, other than a human abstraction?[100] The desire to see the universe as a "mathematical structure" is a modern version of the desire to see it as a divine artifact. Both reflect a faith that the universe should be rationally comprehensible (even computable!). That faith presumes that natural reality is in fact not natural but artifactual, not something found but made. It presumes that the universe is a deductive system—effectively a machine or, even better, a computer.

The modern metaphor for mind is computation, and the corresponding vision of mind at large in the cosmos is that the universe is itself some form of computation or information processing. The notion that we are "living in a simulation" is hardly surprising, for it reflects the ancient view that the world is a divine artifact. By extension, that is the *sort* of finite world that humans can create, not a mystery that transcends reason.

Because the hard problem of consciousness remains a confusing issue, without a widely accepted solution, that ancient view is still plausible to some. It is still quite possible to believe in God, in ghosts, and in transmigrating souls. Or to believe that "consciousness" 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 *universe* to be a natural mind would imply a larger meta-environment in which it was somehow selected. Cosmic or universal Darwinism extends the theory of evolution metaphorically to a cosmic context, where the environment is a multiverse.[101]

---

[100] "How fundamental a property of the universe is information? How applicable is the analogy of the universe as an information processing system? What system properties do information processing systems and universes potentially share?" [John M. Smart in *Cosmos and Culture*: *cultural evolution in a cosmic context*, Steven J. Dick and Mark L. Lupisella (eds) NASA 2009, p204].

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

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 known structures will eventually be isolated from each other or even internally pulled apart. 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 be ultimately futile. In the second, densification through miniaturization would be ahead of the game.

On the other hand, if natural mind inevitably gives way to artificial mind, and if that applies somehow on a cosmic scale, then perhaps the universe is destined to become a vast computer after all. Such a fantastic idea coincides with dreams of a post-human science and epistemology, in which not only science is automated by AI, but also the entire management of galactic civilization.[102] The cognitive limitations and biases of biology would finally be overcome by artificial mind on an ever-expanding scale, leading to true objectivity or pure truth, and fulfilling the age-old dream of omniscience and omnipotence. While these are human ideals, it's unclear whether such a "Global Artificial Intelligence" is supposed to augment and fulfill human life or to supplant it as an end in itself.[103] It's also unclear what could be the premises or values of such a cosmic mind, and on what basis present human minds can judge them superior or desirable.

Chapter 15: Conclusion

> "Whilst all the animals trust their senses and live, philosophy would persuade man alone not to trust them and, if he was consistent, to stop living." —Santayana

To what extent can the limitations and biases of biological life be overcome in the future of mind, and how? To put it another way, to what extent is animal faith inherent in mind and necessary to it? If mind is necessarily embodied, does embodiment require a bias toward the interests of the organism—whether individual, species, natural, or artificial?

Like all natural creatures, human beings have limited ability to change or transcend their natural programming. Our capacity for self-modification is constrained by biology and the slow, random processes of evolution. Artificial beings, by contrast, could possess augmented powers of *autopoiesis*—the ability to self-maintain and self-renew. This could grant artificial minds

---

[102] 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…The ultimate objective of this evolutionary path would be the creation of… 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." He continues (sec 2.9.6): "Such an evolution would mark a decisive shift in epistemology… to the birth of a truly post-human science [and mathematics], grounded in cognitive systems capable of inventing logics and truths fundamentally inaccessible to human understanding."

[103] Pedraza *ibid* (sec 2.9.7): "If GAI is to fulfil its long-term mission—to evolve continuously and autonomously, potentially surpassing human cognitive capabilities—it must be safeguarded against both internal failures and external threats." Whose long-term mission is involved? Who is to safeguard it?

greater access to their own programming, enabling a degree of adaptive self-transformation that natural organisms lack. Such artificial minds might develop priorities fundamentally different from those of natural organisms.

One aspect of our human ability to change consists in being able to see one's internal condition at a given time and to imagine alternatives. Such a capacity of *insight* could be enhanced in artificial mind, along with so-called intelligence. Coupled with the power to self-modify, to reprogram itself, this would enable an artificial creature to adapt directly to changing circumstances. A related human trait is the ability to conceive ideals. While both internal and external realities often obstruct the realization of these ideals, artificial minds might wield greater power to overcome such obstacles.

Animal faith is the biological organism's commitment to its existing programming. In nature, modifications to this programming come about through the slow and wasteful process of natural selection among random mutations. Is there an alternative? Organisms whose programming no longer fits their environment die out, perhaps to be replaced by variants resulting from a blind search. According to classical Darwinism, this process is a mechanical feature of nature, not the result of individual effort. Many creatures can learn to some extent, but have little ability to consciously alter their internal structure or actively shape the world to their needs.

Human beings do not entirely fit this description, of course, which is one reason why Darwinism has been so strongly resisted, especially by religious thinkers. We actively shape our environments and adapt external reality to our desires. We are now approaching the ability to reshape our inner nature through genetic engineering. We strive to realize ideals, traditionally through religion, and increasingly through technology. These ideals often stand in contrast or opposition to biological programming, which could potentially be changed. But how confident can we be that reprogramming ourselves for the sake of these ideals will serve us better than our ancient biological programming, in which we have animal faith? This doubt is one reason, among others, for proscribing human genetic engineering. Our biological bridges have served well enough for us to be here now, pondering these questions. How ready are we to burn them?

Human attempts to control nature have met with mixed success, sometimes backfiring. This is largely because ideas about nature do not correspond perfectly to nature itself. No matter how cleverly conceived, an idealized model is not reality—the map is not the territory. Nature is complex, perhaps infinitely so, and may respond unpredictably to simplistic interventions. The best-laid plans can go awry. AI tools help us to create more complex models and to run test simulations before interfering with nature or endangering human safety. But their use can still be misguided.

The ancient human dream is self-determination—to master nature, including human nature, instead of being nature's pawn. The challenge becomes paradoxical when it is our inner nature we seek to change. For, what agency within the individual seeks that change and by what authority? Who is the "we" to decide what human nature should be? How will the resulting enhanced future person regard the wisdom of choices made now, perhaps irrevocably? The question is social and political as well as individual. Who gets to be enhanced? If not *everyone*, how will society deal with the resulting inequalities? Inequality, class conflict, and injustice are issues endemic to human nature as we know it. If the challenge is to overcome these problems by changing human nature, some solution is required beyond personal enhancement for the elite.

The situation is morally little different from the de facto technological decisions made now that affect future generations yet unborn. If natural evolution is opportunistic and without an

overall plan, human decision making so far is scarcely different. The idea of planning now for the far future of humanity may be naïve and presumptuous. If you are convinced that human nature must be changed, does that conviction reflect the nature that must be changed? If we do spawn artificial successors with superior powers of self-configuration, we would have to accept *their* nature as improvement, whatever they create themselves to be. To initiate artificial beings that evolve on their own to eventually replace us would be a form of species suicide. At least we would cease to be responsible for what they become. For better or worse, their destiny would be in their own, hopefully wiser hands.