

ISSN 1918-7351

Volume 15.1 (2023)

Julian Jaynes and the Next Metaphor of Mind: Rethinking Consciousness in the Age of Artificial Intelligence

George Saad

Memorial University of Newfoundland, Canada

ORCID: 0009-0005-2812-2274

Abstract

In *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, Julian Jaynes presents a philosophy of mind with radical implications for contemporary discussions about artificial intelligence (AI). The ability of AI to replicate the cognitive functions of human consciousness has led to widespread speculation that AI is itself conscious (or will eventually become so). Against this functionalist theory of mind, Jaynes argues that consciousness only arises through the mythopoetic inspiration of metaphorical language. Consciousness develops and enacts new forms of self-understanding, continually evolving new “metaphors of mind,” metaphors which must now account for the emergence of AI.

Keywords: artificial intelligence (AI), Julian Jaynes, evolution of consciousness, metaphor, functionalism

Introduction

Julian Jaynes is a figure who stands at the intersection of several disciplines. His 1976 magnum opus *The Origin of Consciousness in the Breakdown of the Bicameral Mind* evolutionary theory, neuroscience, and literary criticism to advance a speculative hypothesis about the evolution of human consciousness. Anticipating Iain McGilchrist's recent work,¹ Jaynes shows how modern self-consciousness evolved through the increasing coordination of the brain's right and left hemispheres. The self-determination of individual self-consciousness evolved relatively recently from a "bicameral" state. In this primordial state, the left hemisphere was subordinated to the right. The mythopoetic content generated in this hemisphere (particularly the auditory incantations of oral poetic cultures, the gifts of the Muses) governs the left hemisphere, which follows these "songs from beyond" as the enchanted products of an external divinity.

In tracing a biological evolutionary process through the inferential evidence of ancient texts, Jaynes adopts an idiosyncratic methodology which can easily obscure his philosophical commitments. When presented in the earlier, more theoretical sections of the text,² these commitments seem to run off in wildly different directions. At first appearance, he seems to be a reductionist, arguing that what is apparently the work of consciousness is actually accomplished through the unconscious mechanisms of neural networks. *Origin* opens with a thorough deflation of the functions of consciousness, arguing that even learning and reasoning are not essentially conscious activities.

But where this line of argument would seem to lead to an abandonment of consciousness entirely, Jaynes makes a shocking pivot from functional neurology to poetic linguistics, claiming that consciousness could have only emerged after the development of language. This implies that consciousness is much more recently evolved than generally suspected, even among *Homo sapiens*.³ The failed attempt to derive consciousness from cognitive functions was always an attempt to infer an inner experience from external output. The functionalists do not recognize the constitutive role of language in the emergence of this inner conscious experience. For Jaynes, "language is an organ of perception [and] not simply a means of communication."⁴ The "I" that is conscious is at once the product and producer of language. Language does not just describe the world—it is but the very state of having a "world" in the Heideggerian sense. It is subjectivity as such.

¹ Iain McGilchrist, *The Master and His Emissary: The Divided Brain and the Making of the Western World* (New Haven: Yale University Press, 2019).

² See Julian Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind* (New York: Houghton Mifflin, 2000), 1-66.

³ Jaynes, *The Origin of Consciousness*, 66.

⁴ Jaynes, *The Origin of Consciousness*, 50.

Jaynes only briefly presents this unique philosophy of mind remains as the theoretical basis for a more concrete analysis of evidence for the evolution of consciousness as it appears in literature, neurophysiology, and psychiatry. In the first two sections of this paper, I will extrapolate the arguments implicit in, or at least complementary to, the outline of a philosophy of mind presented the *Origins*. While Jaynes is not primarily concerned with a philosophical demonstration of his theory, support for the various elements of his argument can be found across the philosophical tradition. When rendered explicitly, Jaynes' theory is an important response to the functionalism assumed when we equate artificial intelligence with consciousness due to their functional equivalency. Just as famously Kant set limits upon reason to make room for faith, Jaynes limited the functions of consciousness in order to retain its independence as a primary phenomenon irreducible to any functional test. While the application of this theory to AI will be a topic throughout, in the third section I will focus on how Jaynes' work radically upends many of the basic assumptions in the current AI debate.

The Deed Over the Word: Reversing the Functionalism of the Faustian Bargain

It says: "In the beginning was the *Word*."
 Already I am stopped. It seems absurd.
 The *Word* does not deserve the highest prize,
 I must translate it otherwise
 If I am well inspired and not blind.
 It says: In the beginning was the *Mind*.
 Ponder that first line, wait and see,
 Lest you should write too hastily.
 Is mind the all-creating source?
 It ought to say: In the beginning there was *Force*.
 Yet something warns me as I grasp the pen,
 That my translation must be changed again.
 The spirit helps me. Now it is exact.
 I write: In the beginning was the *Act*.⁵

In this passage, Goethe's Faust articulates the shift in philosophical first principles implied in his acceptance of Mephistopheles' demonic pact. The scholar curses the idealism of seeing the world as formed by the divine *logos* and demands that the deed (*die Tat*) instead be regarded as the first principle. In this statement, Goethe presages an inversion of the metaphysical order. As Marx would say a century later, the point of philosophy should no longer be to just interpret the world, but rather to change it.

⁵ Johann Wolfgang von Goethe, *Faust*, trans. Walter Kaufmann (New York: Random House, 1961), p. 153, lines 1224-1237.

The Faustian pact would finally be sealed in the emergence of Anglo-American 20th century pragmatism, a philosophy of the deed.

Critics of this modernity tend to point out its Faustian nature and decry the ambitions of technology as attempts to “play God.” But such conservative resistance to the Faustian pact only highlights its appeal: if mastery over the world through the technological apparatus is irreligious hubris, then the deed is the divine. God is the craftsman to whom we should defer in limiting our technological ambitions. Whether it is embraced or feared, the act has risen above the word. The technological landscape validates this metaphysical inversion, as any technology is nothing other than what it accomplishes. As Alan Turing established with his “Turing Test,” the question of what a computer “is” is much less relevant than the question of what it can accomplish. If a computer can function like a human being, the question of the computer’s internal state, of the presence of consciousness and free will, can be set aside as meaningless.⁶

Julian Jaynes enters this 20th century conversation as a research psychologist who broadly accepts the functionalism assumed in the scientific world of his day. In fact, his interpretation of mainstream positivism severs the activity of mind from all observable behavioral outcomes. This line of argument is supported by current technological developments. As technology advances, it becomes increasingly futile to argue for the functional necessity of consciousness where it has proven to be functionally irrelevant, as when a machine that passes the Turing Test. But where mainstream positivism generally concludes with skepticism about any positive theory of consciousness, Jaynes pivots in a surprising direction. Rather than discard consciousness entirely, we should accept that the word, or language more generally, cannot be reduced to a functional analysis. Language can (and, in fact, must) be severed from any functional outcome: there is more to a linguistic consciousness than the “outputs” it generates. It is now time to reverse the conceptual progression in Faust’s rewriting of the opening of John’s Gospel and retrieve the word in the wake of the deed’s triumph.

Though he does not describe it as such, Jaynes arrives at this position through an essentially phenomenological method. We first recognize that our experience of the world has instilled a bias towards the overestimation of consciousness. We are only aware of the objects of our consciousness and so we naturally take the limits of our consciousness to be the limits of our entire psyche.⁷ When consciousness looks back on its experience, it reconstructs everything in its own terms—as acts and objects of consciousness. In the classic problem of solipsism, consciousness is the circle that cannot escape itself. This results in our generating conscious narratives of that which never crossed the threshold of conscious awareness in immediate experience. We think

⁶ Graham Oppy and David Dowe, “The Turing Test,” *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), ed. Edward N. Zalta, <https://plato.stanford.edu/archives/win2021/entries/turing-test/>.

⁷ Jaynes, *The Origin of Consciousness*, 23.

not of what was actually in consciousness, but of what “must” have been. For Jaynes, “memory is the medium of the must-have-been.”⁸

For example, if I ask you to describe your drive to work, you will quite easily be able to tell me what “must” have happened for the vehicle to operate and the route to be accomplished, even if your consciousness was entirely absent, perhaps thinking of dinner plans or of an upcoming deadline. Surely you were conscious of the general route and anything notable which commanded your attention along the way. But if I press your recollection of the drive beyond general and attention-grabbing, you can only reconstruct your memories. How many cars were ahead of you in the left turn lane? Where were your hands on the steering wheel while turning? How long did you expect the yellow light to last while turning?

While all of us have paid conscious attention to such details at some time (i.e., when we were new to driving or especially aware of some abnormality), and while each of these details are materially relevant to achieving the task of driving to work, they probably never entered our immediate consciousness and certainly were never recorded in memory. Our inner sense of time is almost entirely unconscious and never an explicit measurement, even though having a sense of this timing is a matter of life and death on the road. We would likely feel uncomfortable driving with someone who thought it necessary to rigidly count down the yellow light every time they entered the intersection to make a left. Consciousness quickly evaporates the further we depart from an abstracted narrative of “what must have been” and the closer we approach our actual focus of our attention in driving.

Supporting this phenomenological intuition by citing several empirical psychological studies, Jaynes claims that most of our lives can be lived quite unconsciously.⁹ In fact, most of the functions which are today claimed to show the potential for “consciousness” of artificial intelligence can be shown to have nothing to do with consciousness. A machine may be programmed to navigate a vehicle down a winding road. Perhaps some process of trial and error takes place in the programming of this AI, a process described as “machine learning.” This capacity of the machine to “learn” and accomplish this task is taken as evidence of the machine’s inevitable ascendance to consciousness. But Jaynes claims that such functional accomplishments should be entirely discounted as evidence of consciousness. One may first object that this ability is common to almost all vertebrates. Animals, humans, and machines can all figure out how to navigate themselves down a winding road. Where today many futurists are eager to proclaim that animal, machine, and human are *all* conscious on the basis of their functional capacity, Jaynes would say that precisely *none* of these three categories can be said to be conscious on the basis of a shared functionality.

But surely human learning is conscious? Do we not learn by paying attention to what we are doing and thereby acquiring a skill? For Jaynes, consciousness plays

⁸ Jaynes, *The Origin of Consciousness*, 30.

⁹ Jaynes, *The Origin of Consciousness*, 27-44.

only a formal role in the learning process. It frames the problem but is not itself active in the act of finding a solution. Consciousness supplies some general precepts which are only truly learned in unconscious activity, just as one learns to ride a bike from only a few general intentions. One can learn to play solitaire in a state of semi-consciousness, improving one's skill even while giving one's attention to a podcast. Indeed, the very possibility of multitasking should call into question the functional relevance of consciousness.

One may object that these kinds of skills are basically reflexes and so should not be considered representative of the higher reasoning unique to humans. For Jaynes, consciousness here again plays only a formal role. He offers an example this type of problem of inductive reasoning typically encountered in tests of intelligence:¹⁰



What is the next figure in this sequence?

While one must be conscious of the sequence itself to answer this question, one need not at all be conscious of the acts of reasoning which enable one to answer the question. Once the problem has come to consciousness, the solution suggests itself immediately. We have already solved the problem when we go back and write up a formal logic of how to define any n th term of the series. Having intuited the answer, we go back and write the formal rule of what must-have-been to give our answer a general validity.

When the problem is more complex, consciousness may play more of a role in framing the problem. One can “solve” a more complex sequence through abstract analysis. But while the formal approach would seem to replace unconscious insight with a fully conscious mechanism, the moment of unconscious reasoning is simply transposed over to the intuition behind the formula. There is no formula to generate a formula (that is, an original, non-derivative formula). While we can solve any term in the series by consciously applying the general rule, the general rule does not itself arise from any conscious effort. It is still an intuition, albeit one self-consciously tested and translated into rigorous terms. The skepticism of David Hume towards the validity of inductive reasoning illustrates precisely this point. The positive interpretation of Humean skepticism is the insight that the formal language of self-conscious reason is not self-grounding, as it depends upon the unconscious habit of association.¹¹

¹⁰ Jaynes, *The Origin of Consciousness*, 40. Figure reproduced from the text.

¹¹ David Hume, *A Treatise of Human Nature* (Oxford: Oxford University Press, 2012), Book I, Part III, Section VI.

In summary, we can say that while consciousness often aids and clarifies our psychic processes, it is not generally necessary for them. Chess, a subject of AI research for decades, can here be used as an example. One can play chess with a painstaking consciousness, as in correspondence chess, where a player can have months to consider and make a single move. Such games are expected to be of very high quality because of the intervention of consciousness, which continually reframes the problems of the position and tests out the answers supplied by intuition. It is generally agreed that this kind of slow chess is more beneficial for learners as it allows them to check and grow their intuition in tandem with conscious calculation.

On the other hand, one can play chess with very minimal consciousness, as in “bullet” chess, where all the moves are made in under one minute. While such games may contain more mistakes, they are often played at an extremely high level by accomplished players whose intuition has been trained so that they can find good moves by an act of reflex. A master playing bullet chess will almost always play to a higher standard than an amateur playing correspondence chess. Under the right conditions, it is not at all surprising that unconscious cognitive functioning outperforms conscious cognitive functioning. The overmatched amateur is in a position analogous to the master when they face any modern chess AI—that of being overcome despite the apparent advantage of consciousness. Consciousness certainly enables human beings to widen the scope of our problem solving, but the act of problem solving itself can happen just as automatically as the flow of electrons in a semiconductor.

For Jaynes, this deflationary account of consciousness is a preparatory step which will make plausible his wider hypothesis that human beings have only very recently become conscious. Art, architecture, and advanced civilization could all emerge alongside the evolution of consciousness, a process still not complete. While Jaynes lived before the current AI debates, I believe he would have seen the automation of so many human tasks as proof positive that consciousness never played a decisive role in them. Turning away from function as the proof of consciousness, Jaynes turned to language as the vehicle through which we can observe its evolution.

Language, Metaphor, and the Evolution of Consciousness

While Jaynes’ deflation of consciousness is fairly straightforward, his account of language is much more difficult and controversial. Here I will try to bridge some of the argumentative gaps he leaves implicit as he moves on to present the historical evidence for his evolutionary theory.

Working within the post-Darwinian perspective of modern research psychology, Jaynes presupposes that consciousness must have evolved. Having abandoned the search for consciousness in any psychic function, he turns to language

as the very structure and substance of conscious experience, an “organ of perception.”¹² We can unpack two reasons why language as such is so crucial for an evolutionary account of consciousness. First, consciousness is not a capacity for accomplishing but a state of mindful attention and intention which may or may not accompany our functional dealings with the world. Language is essential to this cultivation of attention insofar as it is only through language that we begin to see the world as containing discrete objects upon which we can fix our attention. This insight can be traced deep into the philosophic tradition. Per Anaxagoras, it is only the conscious mind (*nous*) which differentiates the immediate flux into the stable objects of perception and cognition.¹³ Language is not just a supplemental tool in this process but its very organ. As the *Tao Te Ching* opens, “naming is the origin of all particular things.”¹⁴

Second, if consciousness evolved, it must express itself in a form which is itself capable of evolution. Language is precisely such a medium in that it grows upon itself and has no final fixed form. It facilitates the coming-to-be of consciousness in its own expansion and bridges the gap between unconscious natural intelligence and conscious human intelligence. The meows and chirps of animal language are the first step towards the evolution of consciousness, the first form of bringing-to-attention even if still in the most automatic, instinctive way. Language can be produced and processed in the unconscious just as one need not be aware to scream or smile, but it is through such instinctive language that the possibility of consciousness first emerges. While Jaynes’ complex account of intersubjectivity is beyond the scope of this paper, it is here important to emphasize that language remains essentially social, even if it is no longer regarded as a mere “tool” of communication. Consciousness emerges when instinctive signals become a matter of interpretation for the other, when they seek after the reasons behind our screams and smiles. A call to attention begets further attention, attention given not only to objective states of affairs but also to the other’s own awareness as it is manifest in the medium of language. Wittgenstein was correct that the limits of our language are the limits of our world,¹⁵ but these limits continually transcend themselves as language expands.

If language is the co-evolving vehicle of consciousness, it cannot be reduced to a static system of tightly defined references. In the terminology of American philologist and philosopher Phillip Wheelwright, this sort of language is “stenotyped,” limited to one sense fixed by convention, as in modern scientific language.¹⁶ It is to be contrasted with “tensive” language, language which is not referential but instead

¹² Jaynes, *The Origin of Consciousness*, 50.

¹³ DK 59B12. Robin Waterfield, *The First Philosophers: The Presocratics and Sophists* (Oxford: Oxford University Press, 2000), 125.

¹⁴ Lao-tzu, *Tao Te Ching*, trans. by Stephen Mitchell (New York: HarperCollins, 1988), 1.

¹⁵ Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* (California: Harcourt, Brace and Company, 1922), section 5.6, 74.

¹⁶ Philip Wheelwright, *Metaphor and Reality* (Bloomington: Indiana University Press, 1962), 16.

extends its meaning out of itself, as in the constellation of meanings present in poetic ambiguity. While Jaynes himself participated in and validated scientific discourse, stenotyped language must have evolved only within the last few millennia of human existence, only achieving general adoption in the scientific revolution of the 17th and 18th centuries. The neurologist has something to learn from the philologist if they are to have any hope of how consciousness evolved. We can observe the emergence of consciousness in the metaphors of the epic tradition, as metaphor is the “very constitutive ground” of language.¹⁷

While he is not very explicit on this point, Jaynes seems to approach the radical thesis that all conscious understanding is essentially metaphorical. As language becomes more stenotyped, it only becomes a more abstract metaphor. To quote Jaynes’ own evocative metaphor, abstract words are “ancient coins whose concrete images in the busy give-and-take of talk have worn away with use.”¹⁸ Abstraction always appears as a metaphorical extension of the concrete into a new semantic range. Jaynes points out that the irregular conjugation of the verb “to be” in modern Indo-European languages can be traced back to the Sanskrit verb *asmi*, “to breathe.”¹⁹ Abstracting from our concrete human being to being in general, language opens up new horizons of understanding, as witnessed by the millennia of philosophical discourse on ontology generated from a metaphorical extension of the breath which underlies our existence as living creatures.

Even where language seems to have entirely shed its metaphoric origins it continually returns to metaphor to explain what lies beneath the well-worn linguistic currency. Philosophers should be familiar with this continual appeal to analogy to explain the most difficult concepts, beginning in the Platonic dialogues where allegories attempt to explain what the current sophistries cannot comprehend. In the context of the current discussion, we may say that Socratic questioning first shows the limitations of stenotyped language while Platonic allegory uncovers language’s true metaphorical ground. On the more everyday side of linguistic evolution, the living adaptability of metaphor both enables and reflects our capacity for novelty, as when a new device that is only incidentally used to call people is described as a “smart phone.” For Jaynes, metaphor is the living heart of language which enables a finite collection of lexical terms to extend beyond themselves to describe an infinite set of circumstances.²⁰

One further argument from the history of philosophy may be adduced in support of Jaynes on this point. In its common understanding, metaphor seems to be a special case of linguistic use where, for effect, we say that some X is Y. Where a literal equivalency between X and Y fails, we are invited to contemplate them in their

¹⁷ Jaynes, *The Origin of Consciousness*, 48.

¹⁸ Jaynes, *The Origin of Consciousness*, 51.

¹⁹ Jaynes, *The Origin of Consciousness*, 51.

²⁰ Jaynes, *The Origin of Consciousness*, 52.

similarity and difference. But philosophy has recognized since Aristotle that metaphor does not represent a special kind of proposition but rather is implied in the structure of all propositions.²¹ Every proposition involves difference simply by the nature of the two terms involved. When we say “love is a wet dog” as opposed to “love is a powerful human emotion,” the more metaphorical statement is only more metaphorical by a degree of difference. The equation of love with a wet dog invites contemplation, whereas we can accept the second proposition as a straight-forward definition.

But definition is more of a linguistic sleight of hand than the metaphor, as its accepted identity conceals the difference which the metaphor places out in the open. Socratic discourse will quickly show the neat equation of terms presented in a definition to be hasty and limiting at best. With our example of love, the predicate “powerful human emotion” could just as well apply to hatred, so there is at least the difference that the predicate has a wider range than the subject. We could, like a hapless Socratic interlocutor, attempt to clarify and say that love is a “positive and powerful human emotion,” but once more we are refuted with the unhappy reminder that our experiences of love are not always positive. It is not our fault, however. Socrates was always playing with a loaded deck, knowing that all propositions are always asserting the identity of unlike things, expressing at once synthesis and distinction.

If metaphor is the ground of language and language is the ground of consciousness, consciousness must be in some sense metaphorical. This can be observed in two senses. First, metaphor serves as a bridge between the unconscious and the conscious. A metaphor is not generated from consciousness; indeed, poetry is good largely to the extent to which it is inspired without conscious mediation. The poetic intuition is a wellspring from which the metaphors bubble up as if under their own power. For Jaynes, consciousness is at first only the receiver of the gifts of the muses, a “bicameral” mind in which the still incomplete ego is only a vessel of externalized psychic entities: the muses, anthropomorphic gods, and ancestors whose voices inhabit and govern the ancient mind.²² The metaphoric constructions they present are the pivot point for the emergence of consciousness as unified, self-contained subjectivity.

Consider a Homeric metaphor: a dying soldier’s head droops like the head of a poppy soaked by rain.²³ In the immediate aesthetic effect of the metaphor, we

²¹ Schelling describes the “ancient” understanding of the identity of the copula: “Whoever says, ‘The body is body,’ surely thinks something different with respect to the subject of the sentence than with respect to the predicate; with respect to the former namely, unity, with respect to the latter, the individual properties contained within the concept of body that relate to it as *antecedens* to *consequens*. Just this is the meaning of another ancient explanation according to which subject and predicate are set against each other as what is enfolded to what is unfolded (*implicitum* et *explicitum*.” F.W.J. Schelling, *Philosophical Investigations Into the Essence of Human Freedom*, trans. Jeff Love and Johannes Schmidt (New York: SUNY Press, 2006), 14.

²² Inferring how the bicameral mind would have worked from studies of modern schizophrenics, Jaynes theorizes that the bicameral god expressed itself primarily through what would today be regarded as auditory hallucinations. Jaynes, *The Origin of Consciousness*, 85-94.

²³ *Iliad* 8.357-359.

unconsciously accept this identity of difference, dwelling in the imagery. But the comparison is also a prompt to conscious attention. How is the soldier like the poppy, and how is he not? What is it that makes this image so poignant? Direct literary experience gives way to literary criticism as an emergent consciousness attempts to clarify the difference and similarity of the metaphorical terms. The metaphor is not an accessory generated by a pre-existent consciousness. It is rather that which gives rise to definite consciousness from the loose manifold of unconscious inspiration. Only from a comparison can we arrive at a simple consciousness of any singular thing. What could it mean to be conscious of anything outside of its distinction from something else? Even at the most basic level of perception, pure light would be indistinguishable from absolute darkness. Consciousness, as the awareness of anything as a something, can only proceed from the distinction of that something from something else. Before we can say that $A = A$, an abstract and derivative point of view, we must wrestle with the powerful synthetic imagery of the unconscious mind which insists that $A = B$.

The evolution of consciousness as the gradual making explicit of a primeval poetic richness is a thesis that also can be observed in the intellectual history of the West. The Greek world undergoes this process when the acute consciousness of the Platonic dialogues, the philosophical search for exact definitions, begins to critically unpack the Homeric metaphors of the archaic culture. Reflecting upon the development of Greek intellectual life, the movement from poetry to prose became the archetypal example of historical “becoming” (*das Werden*) in 19th and 20th century German philosophy. As Hegel says in the preface of the *Philosophy of Right*, the self-conscious wisdom of philosophy, the owl of Minerva, only takes flight when a way of life has grown gray and old.²⁴ Nietzsche likewise argues in the *Birth of Tragedy* that the Socratic figure appears only when the Dionysian music has grown faint and subject to Socratic questioning.²⁵ In *Decline of the West*, Spengler expands this aesthetic hypothesis into a general theory of historical birth and decay in which an organic *Kultur* petrifies into a technocratic *Zivilisation*:

Civilizations are the most external and artificial states of which a species of developed humanity is capable. They are a conclusion, the thing-become succeeding the thing-becoming, death following life, rigidity following expansion, intellectual age and the stone-built, petrifying world-city following mother-earth and the spiritual childhood of Doric and Gothic. They are an end, irrevocable. yet by inward necessity reached again and again.²⁶

²⁴ G.W.F. Hegel, *The Elements of the Philosophy of Right*, trans. H.B. Nisbet, ed. Allen Wood (Cambridge: Cambridge University Press, 1991), 23.

²⁵ “Dionysus had already been driven from the tragic stage, and by a daemonic power which spoke through Euripides. Even Euripides was in a certain sense only a mask: the deity which talked through him was neither Dionysus nor Apollo but a newly born daemon called Socrates.” Friedrich Nietzsche, *The Birth of Tragedy in the Spirit of Music*, trans. Douglas Smith (Oxford University Press, 2000), §12, 68.

²⁶ Oswald Spengler, *The Decline of the West: Form and Actuality*, trans. Charles Francis Atkinson (New York: Alfred A. Knopf, 1926), 31.

The movement of literature mirrors the movement of human society; it is a movement from poetry to prose and then back again. Although he does not mention this interpretation of history in 19th and 20th century German philosophy, Jaynes builds upon this tradition when he characterizes the evolution of consciousness as a coming-to-awareness in the wake of an earlier unself-conscious poetic moment. McGilchrist furthers this tradition when he says that the right brain (“the master”) has primacy over the left (its “emissary”). In accepting this “primacy of the implicit,” we realize that “metaphorical meaning is in every sense prior to abstraction and explicitness.” Returning to the original Latin metaphor contained in these worlds, “pulling away” (from *abs-trahere*) and “unfolding” (*ex-plicare*) are acts of analysis which depend upon more primal unity.²⁷

The Metaphor of Mind

While I have liberally reconstructed Jaynes’ diffuse insights into a more explicit argument for metaphor as a bridge between the conscious and unconscious, he is much more direct in presenting a second association between metaphor and consciousness. It is not only that metaphor prompts the emergence of consciousness, but that consciousness is *itself* a metaphor. It is the creation of an analog mental “space” in which the analog “I” operates as if it had a visuospatial reality. Jaynes writes:

[Consciousness] operates by way of analogy, by way of constructing an analog space with an analog ‘I’ that can observe that space and move metaphorically in it. It operates on any reactivity, excerpts relevant aspects, narratizes and conciliates them together in a metaphorical space where such meanings can be manipulated like things in space. Conscious mind is a spatial analog of the world and mental acts are analogs of bodily acts.²⁸

The “I” operating within abstract mental space is like the well-worn coin whose concrete imagery has faded in accustomed use. We can observe what a more concrete metaphor of consciousness would be in poetic language: “The heart desires but the hands are unwilling” is a more concrete way of saying “I am conflicted in my decision.” Indeed, Jaynes theorizes that the ancient world first attempted to describe consciousness by describing it as a faculty localized in different semi-autonomous body parts, like the *thumos* (“spirit”) which often appears as rousing the limbs in Homeric heroes.²⁹ By contrast, in the modern understanding, the simple unity of the

²⁷ McGilchrist, *The Master and His Emissary*, 179.

²⁸ Jaynes, *The Origin of Consciousness*, 65-66.

²⁹ Jaynes, *The Origin of Consciousness*, 69.

first-person pronoun gathers consciousness into a single selfsame “space,” an identity without difference.

The shift between these two metaphors of mind is not merely a change in descriptions of the same phenomenon. Consciousness operates through these metaphors; when the metaphor changes, so does consciousness. The concept of the self which operates in consciousness at any time is the living organ through which that self grows and actualizes itself. If I am a computer, I will operate by a rule of calculation. If I am a raging bull, I will leave a trail of destruction in my wake. When these metaphors prove insufficient to my lived experience, I am in an existential crisis. The metaphor must either grow or die off and be replaced.

But as much as the self-fulfillment of metaphor can be observed in individual psychologies, Jaynes is more concerned with the general historical development of self-consciousness. The consolidation of consciousness in the “I” is the standpoint of objectivity, the Cartesian division between self and world upon which the scientific products of modern culture depend. It is the metaphor which conceals itself as metaphor, the creation of a “head-space” which could no more be spatially located in the head than in the feet.³⁰ It is “attention” marked with any act of actual attending, a purely mental “presence.” It is hermetically sealed off from the body, which has lost its autonomy and is now subordinated to an abstract mentality. Except in now quaint metaphors, the heart and stomach no longer speak for themselves or directly motivate actions; they rather belong to an “I” who possesses them as influences held at a distance. Likewise, social and religious influences lose their immediate inspiration. Ancestors, gods, and muses do not directly partake in our individuality and can only intrude on the autonomous operations of rational self-consciousness.

McGilchrist describes a world ruled by this metaphor of mind as one in which all the inspired idiosyncrasies of personal consciousness have been eliminated as the dominant left-brain (the vehicle of the abstract “I”) devalues and even pathologizes alternative metaphors.³¹ Modern life is trending towards this dystopia, one where “the concepts of skill and judgment [. . .] would be discarded in favor of quantifiable and repeatable processes.”³² All the psychic phenomena which cannot be assimilated to this “I” are demoted to the status of unconsciousness, a shadow self which exacts its vengeance in many of the illnesses of modern culture.

For Jaynes, this shadow self takes a clinical form of schizoid mental illnesses in which the forgotten world of gods, muses, and ancestors intrudes upon the self-narrative of an “I” which cannot recognize these voices as its own voices. The bicameral mind returns but without the mediating structures (shamans, rituals) of ancient society. Even if the victory of the autonomous “I” is secured, it has won at a high cost evident even in the non-clinical illnesses of modern life. The “I” has an

³⁰ Jaynes, *The Origin of Consciousness*, 44-46.

³¹ McGilchrist, *The Master and His Emissary*, 428-434.

³² McGilchrist, *The Master and His Emissary*, 429.

agenda irreconcilable with a body it regards as “other,” and so it disregards the “voices” of the old Roman god Somnus and suffers sleep deprivation. Without the meaningful influence of a historical past, the isolated individual is vulnerable to the appeal of atavistic nationalism, the suppressed “call” of the ancestors possessing the modern individuals.³³

Understanding itself as the master of practical efficacy, this “I” “attaches an unusual importance to being in control.”³⁴ Withdrawn into itself, the autonomous ego proves validates its independence in functional terms, by its ability to command and control the external world. When the technologies it produces too nearly replicate its own operation, this metaphor of mind undergoes an ironic twist. Whereas it had established itself as sovereign over a passive, inert material world, it now finds itself struggling to explain how technology belonging to that world can seemingly replicate its own mental functions. A prisoner to its own functionalist presuppositions, the scientific consciousness which once combated animism now finds itself spinning new metaphors to explain the apparent “consciousness” of its technologies. If the mind is only what it can do, we can only return back to animism when our mental feats are equaled.

Rethinking the Metaphor of “Artificial Intelligence”

With AI, the currently dominant metaphor of human consciousness is being retrofitted to describe a novel human technology. The main sense of the metaphor is clear and uncontroversial enough. “Artificial intelligence” describes certain programs that perform functional tasks generally associated with intelligence. Such a purely functional definition is appropriate because the goal of AI was always only functional. Scientists never set out to recreate a human mind as such but to *improve* upon it in executing programmable tasks. There would be no point in even bothering to design AI systems if the goal were not to *surpass* natural and human intelligence on a purely functional basis. A machine that passes the Turing Test in answering customer service calls is not identical to a human doing the same job, it is *superior*. The computer will not tire like a human and so it can better perform the task of directing inquiries to customer service, achieving the goal the engineers have set for themselves. In recognizing this functional superiority, we likewise recognize the differences between AI and human intelligence which can be concealed in accepting a metaphor (which always contains the tension of difference) as a false and loose equivalency.

This linguistic sleight of hand at play in the entire AI debate in which what would be honest metaphors masquerade as dishonest definitions. Novelty always

³³ Jaynes suggests that such relapses into a bicameral state can be observed in modern nationalism, using imperial Japan as an example. Jaynes, *The Origin of Consciousness*, 159.

³⁴ McGilchrist, *The Master and His Emissary*, 432.

prompts a search for new metaphors of understanding, and the first responses to this search generally prove themselves to be inadequate in time. “Artificial intelligence” and “machine learning” are not scientific definitions but first attempts at metaphor to describe a still-evolving technology. The main intention behind these terms is clear enough, but metaphor, in striving to be adequate to what is ambiguous, is necessarily and productively imprecise.³⁵ Every metaphorical device contains within it a constellation of associations lying alongside the main comparison. For instance, if I say that “love is a battlefield,” the most likely sense of the metaphor is that love is more cruel and destructive than usually thought, but this is not the only sense possible. Battlefields are also sites for noble and heroic action, for great mourning and reverence, even for camaraderie. The point of the metaphor is not that one of these interpretations must be chosen to the exclusion of all the others, but that all are somehow operative at once, even if only as potential meanings lurking in the unconscious.

Likewise, when we say “artificial intelligence” or “machine learning,” we are suggesting more than the basic intention of the metaphor to convey a certain functional capacity. The self-interested proponents of this technology are exploiting a vacant linguistic frontier to establish a compelling metaphor which also connotes the spontaneity, organicity, and inner mental space of consciousness. The metaphor cashes in on our overestimation of the conscious “I.” If, as we generally believe, the conscious “I” is indispensable to all forms of thinking, a machine that “thinks” as well as a human must also have all the other qualities of human consciousness. Dazzled by the functional novelty of the technical accomplishment, we unconsciously accept the associations implied by the metaphor. We do have words which could more plainly describe what is happening in AI, but “applied machine binary calculation” (AMBC) is not a term which will promote a general trust in computers as anthropomorphic beings. The artificially intelligent phenomena now interpreted as organic and insightful would now carry the semantic burden of the world of machines, more akin to the activity of an advanced calculator than a human interlocutor.

There is something instructive in the “artificial” part of the metaphor. AI has only been able to achieve its functional accomplishments by reversing the operation of human intelligence, which begins in metaphor and ends in formal rigor. As a purely formal system, AI is only able to achieve superior technical results by virtue of the specialized dedication of a great mechanical computing power towards a single task, something impossible for a human being who is always also breathing, observing, feeling. This intelligence is “artificial” in the sense that it takes one mental function and isolates it, purifying it of all other cognitive and biological context like a naturally occurring compound which has been refined in a laboratory.

³⁵ “If [metaphor] is not to be escapist and merely a stubborn refusal to face things as they are, it will bear traces of the tensions and problematic character of the experience that gave it birth.” Wheelwright, *Metaphor and Reality*, 46.

If our intelligence is like that of AI, we should regard doing quick mental math as the epitome of human intelligence. We should teach our children that the best way to read a book is to scan out the frequency of the words and begin with a statistical analysis. But, try as we might, we will never be able to out-calculate the machine. We can accept this with shame and resignation, or we can do what consciousness has always done and reexamine the metaphors. Metaphor demands both that we see what is similar and what is different. The term “artificial intelligence” has disclosed a certain functional similarity of new technologies to some human capacities, but the differences must now be retained and emphasized as the metaphor evolves.

Nonetheless, the technological aides currently referred to as “AI” will be, for the foreseeable future, one of the components of the human “I.” They will inhabit our mental space and be considered in our decisions no less than our knowledge of history, our sense of ethics, and our aesthetic judgements. Like the muses of ages past, calculation aides can function in an almost revelatory way, disclosing whole new horizons of knowledge such as when, through the sheer power of the mainframe, they adopt strategies in chess never considered by any human. But they are only aides, not replicas of the “I” assumed to have anthropomorphic qualities just like the Greco-Roman gods. If Julian Jaynes were alive today, I believe he would remind us that the age of the bicameral mind is past, and that we should not return to it by further overextending the metaphor of the “I.” Attributing the “I” to whatever technology surpasses human beings on a functional basis will only create a new bicameral world in which technology appears as an alien sovereign issuing schizoid pronouncements to despairing humans.

Avoiding this dystopian fate requires clarity about consciousness just as much as clarity about machines. This is only our destiny if we interpret ourselves as a processing power which would be overthrown if eclipsed by machines. We are not only this functioning, this doing, but also this interpreting, this self-creating. The bicameral world broke down only when the metaphor of mind changed so that we heard our voices as our own. In the modern world we recognize that the Greco-Roman gods and muses always lived within us. The challenge for the next metaphor of mind will be to incorporate forms of artificial intelligence into the sphere of human subjectivity without treating them as if they were themselves individual subjects. To grant AI the autonomy of the “I” would be a failure to meet this challenge, a new breakdown in the metaphor of mind and a repudiation of modernity undertaken, ironically, in the celebration of scientific progress.