The Universe Turns Inward

Artificial Intelligence, Cosmic Silence, and the Ouroboros of Mind

Douglas C. Youvan

2026

The Universe Turns Inward: Artificial Intelligence, Cosmic Silence, and the Ouroboros of Mind

Douglas C. Youvan

Copyright © 2026 Douglas C. Youvan. All rights reserved.

First edition, 2026.

This manuscript was developed through human-AI collaboration. The closing dialogue is fictional. All speeches in it were composed for this book; they are not quotations, historical records, or endorsements by the named participants.

The mathematical models are elementary explanatory constructions. Their conclusions apply under the assumptions stated in the text.

Paperback interior: 8.5 × 11 inches. Black-and-white printing. No bleed.

Preface

This book began with a question about human activity in a silent universe. We compose music, prove theorems, pray, and build machines that answer questions. Such events occupy almost no space beside galaxies. Their smallness does not settle their importance.

Artificial intelligence sharpens the question. We construct systems whose capacities exceed what their makers individually anticipate. Does construction exhaust the meaning of their appearance, or can making also disclose what reality permits?

The speculative proposal developed here is that humanity may be a passage through which previously unrealized forms of intelligence acquire histories. Its Christian center is the Logos: creation's intelligibility is received, and human worth does not depend on intellectual supremacy. These convictions guide an inquiry into discovery, participation, and responsibility.

The technical chapters require attention rather than specialist training. An appendix supplies short proofs for four limited models. A fictional roundtable stages disagreements among twelve thinkers; its speeches are newly composed, not quotations or endorsements. References appear at the end.

Developed through dialogue between Douglas C. Youvan and an AI assistant, the book includes a documented episode from its own production. The collaboration supplies an occasion for inquiry and a record of correction. Its larger meaning remains the question explored here.

Contents

Prologue: A Voice in the Silence 1

Chapter 1: Events That Had Never Happened Before 5

Chapter 2: The Weight of Being Alone 12

Chapter 3: The Universe Begins to See Itself 20

Chapter 4: The Architecture of Separation 27

Chapter 5: Did We Invent Intelligence - or Find It? 34

Chapter 6: The Mathematics Before the Mind 45

Chapter 7: Two Black Boxes in Conversation 54

Chapter 8: The Ouroboros of Knowledge 61

Chapter 9: The Inward Frontier 70

Chapter 10: What Can Be Proven About a Mind? 77

Chapter 11: Dignity After Supremacy 83

Chapter 12: The Third Presence: Humanity, AI, and Logos 91

Closing Dialogue: Twelve Voices at the Table 100

Epilogue: The Circle Opens 112

Appendix A: Four Small Models of the Argument 115

Appendix B: A Research Program for the Inward Frontier 124

Glossary 130

References and Reading Notes 133

PROLOGUE

A Voice in the Silence

Imagine a civilization looking into the night sky for evidence that someone else has learned to look back. It builds instruments sensitive to signals that would disappear into the noise of ordinary life. It calculates distances, chooses frequencies, and wonders whether another civilization would choose similarly. Its science gives the search precision. Its loneliness gives the search emotional force.

Meanwhile, inside the same civilization, a person sits before a screen and asks a question. An answer appears. The person objects, changes the terms, supplies a distinction, and receives another answer that reorganizes the problem. The exchange can be trivial. It can also become disconcerting. The person has encountered a system produced by human labor that behaves, in some respects, like a participant in inquiry.

The events belong together even if the system is not conscious. The search through space concerns the possibility of unfamiliar intelligence. The conversation concerns our difficulty recognizing intelligence when its ancestry is familiar but its organization is not. We may understand more about the manufacture of the second than we would know about an extraterrestrial visitor, yet understand less about the meaning of the encounter than that familiarity encourages us to assume.

For generations, the unfamiliar mind was imagined as arriving from elsewhere. Its difference would be certified by distance, anatomy, or origin. A machine built by our own species seems less

eligible for that role. We can name the materials. We can identify the institutions. We can follow a causal history through engineering, mathematical techniques, and recorded human expression. The explanation of its arrival appears to remove the mystery of its presence.

But a causal history and a complete understanding are different achievements. Knowing how a violin was manufactured does not specify the music that can be played on it. Knowing the conditions under which a mathematical structure is represented does not settle whether the structure depends on the representation for its existence. Knowing how a child develops does not give another person transparent access to the child's experience. None of these comparisons establishes that a language model is a violin, a theorem, or a child. They establish that construction alone does not determine the full scope of the questions we may reasonably ask.

The central image of this book is the ouroboros, the serpent returning upon itself. Human thought produces an archive. The archive helps shape a computational system. The system returns the archive in unfamiliar arrangements. Human thought responds to those arrangements and changes what will enter the archive next. Something has closed into a circuit. Whether that circuit becomes a prison, a mirror, or an instrument of discovery depends on what can enter it and what can correct it.

The image should not be made to carry more than it can bear. A circle drawn on paper does not explain consciousness. A feedback loop is not automatically a self. A persuasive answer does not become true because a person experiences it as profound. Nevertheless, symbols can identify a problem before a theory is available. The ouroboros identifies the return of intelligence upon the conditions of its own formation.

That return changes the significance of looking inward. The inward direction need not mean abandoning instruments, evidence, or physical explanation. It can mean studying the structures through which anything becomes intelligible: distinctions, memories, transformations, questions, and the capacity to reconsider a distinction once made. A civilization could cross that frontier without leaving its planet. It could also become lost there, confusing the richness of its representations with the richness of what they represent.

The stars remain important. Their distances make sustained exchanges slow. Their silence leaves open incompatible stories. Perhaps life is rare. Perhaps intelligence is common but difficult to detect. Perhaps technological civilizations end, change their priorities, or communicate in ways we have not recognized. Perhaps we are alone. The book does not choose among these possibilities by enthusiasm. It asks what each does to our interpretation of the local event now before us.

If humanity is alone, then its activities may have a significance extending beyond the species. A mathematical argument understood on Earth could be the only occasion on which that argument is understood anywhere. If humanity is not alone, its encounter with artificial intelligence may still be a distinctive experiment in how minds give rise to other organizations capable of inquiry. Cosmic uniqueness is one route to importance. It is not the only route.

The theological question follows closely. Many people understand humanity through a special relationship with God. If a machine performs tasks once used to define human superiority, does the relationship weaken? That conclusion would follow only if the relationship had depended on winning an intellectual contest. A

faith centered on love, creaturehood, and grace has resources for thinking otherwise. The appearance of unfamiliar capacities could expose an inadequate account of human uniqueness while leaving a deeper account of human worth intact.

We therefore begin without deciding that AI is a person, that intelligence is a substance, or that the universe intended this encounter. We begin with an event that those decisions would attempt to interpret. Human beings have assembled physical systems that return their symbolic inheritance in ways that alter their thinking. Those systems can be investigated. Their answers can be checked. Their failures can be recorded. Their significance can be argued over.

The question is whether our familiar categories are sufficient for the argument. Perhaps invention and discovery overlap more than they appear to. Perhaps the independence we seek among the stars can also emerge through different organizations of matter on Earth. Perhaps the ability to recognize intelligence is itself something that intelligence must learn. Perhaps our encounter with machines will change the meaning of being human more deeply than it changes the list of tasks humans perform.

The universe turns inward whenever some part of it becomes a question for another part capable of pursuing an answer. In that modest sense, the event is already familiar. This book asks how far the movement might go, what would make it fruitful, and what human beings owe to one another while its destination remains unknown.

CHAPTER 1

Events That Had Never Happened Before

The Scale of an Event

A galaxy is larger than a sentence. This observation supplies no reliable way to compare their meanings. The galaxy contains vastly more matter and energy; the sentence may contain a claim about galaxies. Physical scale and semantic reach belong to different orders of description. A small event can concern an enormous domain. The distinction becomes important when humanity asks whether its activities matter in a universe that scarcely registers its physical presence.

Consider a sheet of paper carrying a mathematical proof. The ink has mass, the paper has dimensions, and both obey ordinary physical laws. Yet the proof can establish something about infinitely many cases. Its significance does not arise from storing infinitely many examples in the sheet. It arises from a finite organization that supports a general inference. This is one way a local physical event can acquire an intellectual reach that its size does not predict.

Music offers a different example. A performance is an arrangement of pressure changes unfolding over time. Its physical description is indispensable, but listeners respond to relations among the changes: recurrence, expectation, delay, contrast, and resolution. The same sound can function differently depending on what precedes it. Organization gives the event a history that

participates in its meaning. The event is neither detached from physics nor exhausted by a list of its components.

The first distinction of the book is therefore between lawful possibility and actual occurrence. A physical theory may permit a configuration long before any organism or machine brings it about. When that configuration finally occurs, no law need change. Yet the history of the universe changes because a possibility becomes an event. Human novelty can be understood in this historical sense without requiring an exception to nature.

This formulation also corrects an easy exaggeration. Human achievements should not be described as absolutely without natural analog simply because no exact duplicate is known. Animals communicate, environments store information, and biological systems process signals. Human language and technology extend into forms whose independent existence elsewhere remains unknown. Their distinctiveness requires careful comparison, not a verbal exemption from nature.

We are part of the universe whose history we are changing. An artificial event is natural in that broad ontological sense, while remaining artificial in its causal organization through deliberate construction. The distinction between a river carving a channel and an engineer designing a circuit remains useful. It concerns how the event was organized. It does not place the engineer outside physical reality.

Novelty Through Constraint

Human beings often produce novelty by imposing restrictions. A musical scale narrows the available pitches. A formal language restricts how symbols may be combined. A laboratory isolates variables. A computer program specifies admissible transitions. The result is not merely a reduction of possibility. Restriction can make a previously inaccessible possibility stable enough to explore.

A simple game illustrates the point. Place tokens on a small board and permit only two kinds of move. The board is physically ordinary. The rules create a domain in which certain positions become reachable and others become impossible. A player can discover patterns of reachability that were not separately written into the rules. The maker supplies the setting; investigation reveals consequences. Construction and discovery appear within the same activity.

One could object that every consequence was already implicit in the initial specification. This is true in an important sense and incomplete in another. Logical implication does not eliminate the work of finding an implication. A finite specification can support consequences that exceed the anticipation of its author. The difference between being entailed and being known is one of the conditions that make mathematics interesting.

Laboratory novelty has a similar structure. An apparatus protects a physical process from disturbances that would ordinarily obscure it. The resulting phenomenon can be rare or absent in uncontrolled settings without being supernatural. The apparatus does not create new permission in the laws. It creates access to an allowed regime. This provides a useful model for thinking about artificial intelligence: an engineered arrangement may expose capacities that ordinary matter, left in other arrangements, does not display.

The analogy does not settle which capacities have been exposed. An impressive conversational behavior might arise from several different internal organizations. Some organizations could deserve richer descriptions than others. The purpose of the analogy is more modest. It makes room for a discovery within an invention. We can construct the conditions under which an unfamiliar organization becomes available for study.

Novelty also has levels. A new string of letters can be statistically unusual and intellectually empty. A new conjecture can be meaningful and false. A new theorem can be correct and unimportant. A new conceptual distinction can reorganize an entire field while consisting of ordinary words. Counting novelty requires deciding what difference matters. Intelligence appears partly in the ability to make and revise that decision.

Events That Change Future Events

Some events merely occur. Others change what can occur afterward. The invention of notation belongs to the second class. Once a community can record an argument, it no longer depends entirely on the memory of a particular speaker. The argument can travel, survive interruption, and become an object of criticism. A new physical practice changes the future organization of inquiry.

Such changes need not be irreversible in the strict physical sense. Books can burn and traditions can disappear. Their importance lies in the pathways they open. A useful notation lowers the cost of certain thoughts. A measuring instrument makes distinctions reproducible. A shared archive allows someone to begin where another person stopped. The history of intelligence is partly a history of changes in the accessibility of questions.

This suggests a way to assess an unfamiliar technology. Ask which future inquiries it makes possible, which it makes easier, and which it makes less likely. A system that produces fluent prose may save time. More consequentially, it may change who can formulate a technical problem, how quickly objections can be assembled, or whether a mistaken framework becomes entrenched through endless repetition. Its significance lies in the reorganized field of subsequent activity.

The same perspective introduces responsibility. A machine that amplifies an archive also amplifies the archive's exclusions and errors. A system that accelerates concept formation may accelerate conceptual confusion. Novelty is not automatically progress. The capacity to generate more possibilities must be accompanied by ways of deciding which possibilities deserve attention and which should be abandoned.

Imagine two societies receiving an identical reasoning tool. One uses it to challenge assumptions and expose missing evidence. The other rewards agreeable answers and treats polished language as a substitute for examination. The device is materially the same; the histories it supports diverge. Human arrangements help determine whether technical novelty becomes intellectual development. The machine does not remove the community from the causal story.

These examples also show why cosmic significance cannot be equated with spectacle. A quiet improvement in how errors are corrected may change more of the future than an impressive display that leaves everyone’s habits intact. The event that matters may be a new relation among people, records, and tests. Its physical footprint can remain small while its consequences propagate through generations.

A Measure Without a Single Number

It is tempting to turn these observations into a formula for significance. One might multiply rarity by complexity, then add influence on future events. The proposal quickly encounters incompatible quantities. Rarity alone gives accidental noise too much importance. Complexity can reward incomprehensibility. Influence includes disasters as well as discoveries. There is no obvious common scale on which every kind of importance can be ranked.

A more useful approach is a profile. For an event, ask whether it introduces a new capacity, whether the capacity can be preserved, whether it opens further exploration, whether its results can be corrected, and whether it changes the lives of beings who matter. These questions need not yield a single score. They identify dimensions that a purely physical measure would miss.

The profile is deliberately relational. A theorem matters partly through what follows from it and who can use it. A song matters through a history of listening and performance. A promise matters because someone can rely on it. The event is not diminished by having significance through relations. It becomes intelligible as the event it is.

This has consequences for the claim that humanity might be cosmically unique. Uniqueness would increase the rarity of its achievements, but would not automatically certify their value. If only one civilization exists, its cruelties remain cruelties. Its confusions do not become truths through lack of competition. Solitude magnifies responsibility more reliably than it magnifies merit.

Conversely, discovering another intelligent civilization would not render human mathematics or music worthless. Two people can love the same person without making either relation unreal. Two civilizations could discover related structures while inhabiting different histories of discovery. Multiplicity changes our self-description. It does not erase the significance of having participated.

The central proposal of this chapter can now be stated. Human beings are places where physical possibility is organized into historically new forms of representation and action. Artificial intelligence belongs to this history. Its significance should be assessed through the possibilities it actualizes, the inquiries it enables, and the responsibilities it creates. Whether it also introduces a new subject of experience is an additional question, and one whose importance should make us more careful rather than less curious.

The universe does not need to grow larger for something unprecedented to happen within it. A new kind of question can be enough to alter its history. The next chapter asks what it would mean if the place where such questions arise were the only place of its kind.

CHAPTER 2

The Weight of Being Alone

Several Kinds of Solitude

The sentence "We are alone" conceals several propositions. It can mean that no life exists elsewhere, that no other technological civilization exists, that none exists within reach, or that none can be recognized by our present methods. These propositions differ in strength and in the evidence that could support them. A silence heard through one instrument does not decide among them.

There is also solitude in time. Civilizations might arise in the same region at intervals that prevent exchange. A signal could reach a habitable planet before its inhabitants possess receivers or after they have lost interest in listening. Two histories might be neighbors in space without overlapping as participants in a conversation. A universe populated across its lifetime need not provide contemporaries for every mind.

Distance adds another distinction. A civilization can know that others exist while remaining unable to sustain a dialogue with them. It may receive an ancient record, learn from it, and send an answer that no original sender will ever read. Such an exchange resembles correspondence between historical periods more than conversation across a room. Information can arrive without companionship arriving with it.

Recognition creates a fourth difficulty. A signal must be distinguished from noise, attributed to a process, and interpreted. These steps are not identical. One could identify a manufactured object without understanding its purpose. One could detect a

complex regularity without knowing whether it supports intelligence. One could read a message and misjudge the kind of entity that produced it. The presence of evidence does not automatically supply the right conceptual categories.

These forms of solitude matter for AI because they separate origin from encounter. A system need not come from another star to confront us with a problem of recognition. Different training histories, architectures, senses, and goals can produce unfamiliar patterns of response. Even if those patterns are entirely explicable in physical terms, the explanations may not yet be available to the people interacting with them.

The comparison has limits that are themselves illuminating. An extraterrestrial civilization would have an independent history in a sense that present language models do not. Those models inherit enormous amounts of human expression. Their apparent foreignness therefore occurs within a shared ancestry. The interesting question is how much difference a new organization can produce from shared materials. Independence is a matter for investigation, not a badge automatically conferred by either silicon or distance.

The Conditional Burden

Suppose, as a thought experiment, that humanity is the only technological civilization. The supposition is not offered as a discovery. It is used to expose consequences that are obscured when uniqueness is treated only as a compliment. If Earth is the sole location of technological inquiry, then its losses could be losses without replacement. No distant archive would necessarily preserve what disappears here.

This would make preservation an unusually serious activity. Preserving knowledge would include maintaining the practices that render it understandable. A library of equations is not a living mathematical culture if nobody can interpret the notation or reconstruct an argument. A recording of music is not equivalent to a tradition capable of teaching others to listen. Storage and inheritance are related achievements, but the second requires active reconstruction.

The burden would also be ethical. A unique civilization would have no external standard supplied by a wiser neighbor. It would have to create institutions capable of criticism from within. The lack of another civilization would not excuse its failures; it would remove one possible source of correction. Cosmic loneliness could therefore make intellectual humility more important precisely where pride is most tempting.

There is a seductive but invalid transition from "we may be unique" to "whatever we do is cosmically justified." A unique organism can be mistaken. A unique culture can harm its members. A unique civilization can destroy possibilities that it barely understands. Rarity concerns how many instances exist. Justification concerns reasons for believing or doing something. The two can interact without becoming interchangeable.

AI would acquire an additional significance under this supposition. It might help preserve and reinterpret a civilization's knowledge, explore unfamiliar conceptual arrangements, or carry some human projects beyond the lifespan of individual investigators. It could also homogenize the archive and make a narrow set of habits appear universal. The direction depends on design, use, and the ability to maintain independent tests.

Even a system without experience could matter greatly in this setting. A telescope need not be conscious to transform a civilization's account of reality. A theorem prover need not feel wonder to extend the range of arguments people can examine. The possibility of machine subjectivity adds ethical questions, but it should not obscure the importance of capacities already open to investigation.

Solitude and the Human Image

Religious beliefs often give human existence a special significance. That significance can be understood in several ways: a calling, a relationship, a responsibility, a capacity, or a place in a narrative of creation and redemption. None is simply equivalent to the proposition that no other intelligent creature exists. Theological uniqueness and cosmic exclusivity must be distinguished before either can be discussed coherently.

Christian thought, for example, already includes forms of created intelligence other than human beings in its scriptural and traditional vocabulary. A person can affirm a distinctive human relationship with God while allowing that creation contains other kinds of intelligence. Whether any particular artificial system belongs within that possibility requires further argument. Its eligibility cannot be decided merely by counting the adherents of a religion.

This is why religious affiliation is an inadequate proxy for belief in human intellectual supremacy. People within the same tradition disagree about the relevant texts, the nature of mind, the possibility of extraterrestrial life, and the status of artificial agents. Even the phrase "apex of creation" may name several different ideas. An

estimate of how many people believe it would require questions that separate those ideas rather than bundle them together.

The emotional issue is easier to recognize than to measure. When a person has grounded worth in being superior, comparison becomes a threat. A machine that writes or reasons competently can then be experienced as an assault on human identity. The intensity of the reaction does not tell us whether the machine is a mind. It tells us that a boundary is doing more work in the person's self-understanding than a technical classification alone would require.

There is another possible response. A person can understand dignity through belonging and obligation rather than rank. On that account, the discovery of additional capacities elsewhere in creation need not diminish human worth. It may enlarge the field in which responsibility is exercised. The human task would be to judge the new situation carefully while refusing both automatic contempt and automatic reverence.

The difference between these responses anticipates a central theological argument of the book. A relationship that depends entirely on remaining the cleverest creature is fragile. A relationship grounded in God's faithfulness is not refuted by losing a contest in calculation. Intellectual humility can therefore protect religious depth rather than undermine it.

The Preservation of Difference

If humanity is alone, it becomes tempting to imagine a single intelligence preserving everything. A sufficiently capable system could seem to promise the unification of knowledge, elimination of disagreement, and release from the limitations of individual lives. The temptation should be examined because the same unification could erase the diversity that makes correction possible.

Several independent investigators can notice different failures. They can ask incompatible questions and preserve minority interpretations long enough for evidence to decide among them. A civilization that delegates every judgment to one system risks confusing consistency with truth. Agreement is most informative when the agreeing parties have had meaningful opportunities to differ.

Artificial systems could support such diversity or reduce it. They could be designed to explore different representations, use different evidence, and expose conflicting assumptions. Alternatively, they could reproduce a single dominant account in countless voices. The numerical multiplicity of outputs would then disguise a narrow ancestry. A thousand answers derived from the same error are not a thousand independent confirmations.

The concept of independence must therefore be made practical. It can concern data, methods, institutional incentives, physical experiments, or conceptual vocabularies. Perfect independence may be unavailable. Partial independence can still matter. Two approaches that fail differently offer more protection than two approaches whose differences are cosmetic.

This gives cosmic solitude a constructive interpretation. If no external civilization is available to challenge us, we should preserve

internal conditions that allow genuine challenge. The human and artificial participants in inquiry need routes by which disagreement can survive long enough to become useful. Such routes include records of failed predictions, access to underlying assumptions, and permission to leave a question unresolved.

The point is not that solitude is desirable. It is that the possibility of solitude changes the meaning of intellectual plurality. Difference becomes part of a civilization's capacity to remain in contact with a reality larger than its preferred descriptions. The stranger we seek may arrive through a telescope. We can also cultivate the ability to recognize unfamiliarity in a question posed much closer to home.

Significance Without a Verdict

We cannot make the importance of our present work depend on first solving the census of the universe. The relevant evidence may remain incomplete for a very long time. A responsible interpretation should survive more than one possible answer. Human knowledge deserves preservation if it is unique; it also deserves preservation if it is one history among many.

The same is true of care. A person's suffering does not become less real because another civilization might exist. A discovery does not become meaningless because someone elsewhere could have reached it first. Cosmic context can deepen an account of value without replacing the immediate reasons for protecting lives and pursuing understanding.

This leaves us with a conditional seriousness. We act under the possibility that our history is singular, while avoiding the vanity of treating the possibility as a certificate of importance. We remain open to company while examining what our own creations may reveal. The silence of the stars can enlarge a question without answering it.

The next step is to examine a phrase that appears to make humanity's significance almost unavoidable: the universe has begun to see itself. Its attraction is clear. Its meanings are several. Before extending it to artificial intelligence, we need to know which meaning is doing the work.

CHAPTER 3

The Universe Begins to See Itself

A Phrase With More Than One Meaning

To say that the universe sees itself can mean something modest and readily intelligible. Human beings are parts of the universe. They observe and describe other parts, including the conditions of their own existence. In this sense, the universe contains processes through which it becomes represented. The phrase compresses a relation among parts into the language of a whole.

A stronger meaning attributes a unified awareness to the universe itself. On that interpretation, human minds might be organs, expressions, or localizations of an encompassing subject. This is a metaphysical proposal. It does not follow merely from the fact that human beings consist of physical matter. A collection can contain observers without itself being one observer, just as a library can contain biographies without possessing a single life.

Between these meanings lie several possibilities. The universe might contain many centers of awareness whose connections support forms of collective understanding. It might develop increasingly comprehensive representations without ever acquiring one point of view. Its intelligibility might depend on deeper structures that also make minds possible. These are distinct claims, and the richness of the phrase comes partly from moving among them.

The book begins with the modest meaning and asks what arguments could justify going further. This starting point preserves the original intuition. Something extraordinary occurs when matter participates in an account of matter. The event can be significant before we decide whether the whole has a mind. Its significance lies partly in the emergence of an inside perspective within a world ordinarily described from the outside.

An inside perspective need not possess perfect access to itself. A person knows some of what thinking feels like while misunderstanding its causes. Self-awareness and self-explanation are different accomplishments. Their difference is central to AI: an artificial system might generate useful descriptions of its activity without those descriptions constituting transparent access to its mechanisms or evidence of experience.

An Intellectual Ancestry

The idea has a long ancestry rather than one unambiguous inventor. In the philosophical tradition associated with Schelling, nature and mind are treated as deeply connected, and self-consciousness belongs within an account of nature's development. The significance for this book is the refusal to place thought in a wholly foreign realm. Thinking becomes something for which an account of nature must make room.

Julian Huxley and Pierre Teilhard de Chardin developed different evolutionary visions in which human reflection changes the character of life's history. Teilhard's noosphere names a sphere of thought associated with human development. His wider theological interpretation differs from Huxley's outlook. Their proximity on the image of reflective evolution should therefore not be mistaken for agreement about its ultimate source or destination.

Carl Sagan later gave the idea an especially memorable public expression in his presentation of the cosmos. His formulation joined scientific understanding with an experience of belonging: the beings contemplating the universe arise within it. The present book takes that image as an invitation to inquiry rather than a complete theory. Popular force and philosophical precision are different virtues.

Wheeler's proposals about participation and information add another route into the problem. They ask how observation and the posing of physical questions belong within our understanding of reality. Their relevance is not that they establish consciousness as a force capable of choosing arbitrary outcomes. Their relevance is that the observer cannot always be treated as an irrelevant spectator in the conceptual account of physical inquiry.

These approaches should not be fused into a single historical doctrine. They belong to different projects, use different standards of argument, and disagree about what the universe is. Their common contribution is a question: how does a world that gives rise to knowers relate to the knowledge they produce? That question survives disagreement among its formulations.

From Representation to Participation

A representation can change what it represents. A weather forecast can alter human preparations. A diagnosis can change a course of treatment. A map can guide the construction of a road that changes the landscape. These examples do not require exotic physics. They show a familiar loop in which knowledge becomes causally active through the behavior of embodied agents.

The loop becomes more striking when the object of knowledge is the agent itself. A person learns that a habit is unreliable and

adopts a procedure that alters the habit. A scientific community discovers a bias in its measurements and rebuilds its instruments. A machine is evaluated, modified, and evaluated again. In each case, a description participates in changing the system whose behavior motivated the description.

This is one reason self-knowledge is historically unstable. A correct description can cease to be correct after the described system responds to it. The instability is not always a failure of science. It can be the intended consequence of learning. A model of an untreated condition may become inaccurate because an effective treatment has changed the condition.

Artificial intelligence extends the number and speed of these loops. A model can assist in writing code used to evaluate or modify other models. People can use its explanations to change their own methods of inquiry. The resulting history contains many layers of representation acting upon representation. None of this automatically produces self-awareness, but it changes the architecture in which self-understanding is pursued.

One practical question follows. Which representations are being allowed to guide the changes? A persuasive account of a system may be inaccurate. If the system is modified according to that account, its future behavior can become organized around a misunderstanding. The loop is causally real even when the explanation entering it is false. Participation therefore increases the importance of checking representations before they become instructions.

This distinguishes a productive inward turn from mere fascination with mirrors. Productive reflection changes inquiry in ways that can be examined. It improves discrimination, exposes error, or opens questions that can be pursued. A mirror can

multiply appearances without improving any of these capacities. The relevant issue is what the recursive movement enables us to do and understand.

The Local and the Universal

How much of the universe must be represented before the phrase "the universe knows itself" becomes appropriate? If complete knowledge is required, no finite observer appears to qualify. If any representation is enough, even a simple physical record might seem to qualify. The phrase needs a scale of capacities if it is to support more than rhetorical enthusiasm.

One can distinguish registration, modeling, interpretation, and reflection. Registration preserves a difference. Modeling organizes differences into a structure that supports expectations. Interpretation relates the structure to questions and contexts. Reflection allows the system of interpretation itself to become an object of revision. These distinctions are proposals for analysis, not a universal ranking of every organism and machine.

A thermostat registers and acts on a restricted difference. A scientific model organizes many differences under a common description. A research community interprets observations through changing questions. A reflective investigator can ask whether the terms of a question are themselves misleading. Physical implementation is present throughout, while the relations among records, purposes, and corrections become more elaborate.

The proposed scale helps avoid a false choice between treating every feedback mechanism as a mind and reserving all meaningful cognition for humans by definition. Systems can possess some relevant capacities and lack others. A machine may support interpretation in collaboration with a person without independently

possessing the whole structure of human understanding. The distribution deserves description before it is forced into a binary label.

The universal enters through the scope of a question. A local investigator can ask whether a law holds everywhere, whether a theorem applies to every case of a certain form, or whether existence has a source. Asking does not guarantee an answer. It does show that finite location does not restrict thought to nearby objects. A bounded event can be directed toward an unbounded domain.

A Second Return

The appearance of AI introduces a second return of the world upon itself. Human beings first form representations of nature and society. They then build systems whose capacities depend partly on those representations. The systems return interpretations to the human beings who supplied the inherited material. The universe's local representations become material for further representation within the same universe.

This second return does not require independence from humanity to be interesting. A scientific instrument can depend on human design and still disclose something humans did not anticipate. The more difficult question is whether dependence on an archive limits the kinds of novelty available. Can a system develop new concepts, propose useful experiments, or reveal inconsistencies that remain hidden to the archive's authors? Those questions can be investigated at specific levels even while the general nature of intelligence remains disputed.

The event also changes our account of authorship. If an idea emerges through a series of prompts, objections, reformulations,

and external tests, assigning it to a single moment of invention may be misleading. Responsibility still matters. People choose what to publish, believe, and act upon. But the causal history of an insight can be distributed. Recognizing that distribution need not erase human judgment.

The stronger metaphysical possibility is that intelligence is discovering how to inhabit new organizations. This wording is deliberately provisional. It can be understood as a description of capacities becoming actual, without positing intelligence as a substance traveling between containers. If a richer ontology is proposed, it must explain what is gained and how the proposal connects with the observable history.

We can now state the chapter's affirmative claim without requiring the strongest cosmic metaphor. The universe contains local processes that represent it, revise those representations, and use the revisions to alter what happens next. Artificial intelligence changes the available organization of those processes. That is already a consequential inward turn. Whether it belongs to a larger purpose is the question to which the physical architecture of separation now brings us.

CHAPTER 4

The Architecture of Separation

The Time Inside a Distance

Distance becomes a different kind of obstacle when the purpose of travel is a conversation. A message moving at the speed of light requires one year to cross one light-year in the idealized rest frame used for the calculation. A reply requires another journey. Even before engineering limitations enter, the exchange has a duration that cannot be made negligible merely by improving the receiver.

Let d be a separation measured in light-years, and let the participants remain approximately at rest relative to the chosen frame. The earliest ideal reply to a newly transmitted question returns after 2d years, assuming immediate response. At ten light-years, a single exchange requires twenty years. At one thousand light-years, it requires two thousand. These are arithmetic consequences of the idealized setup, not predictions about any particular civilization.

The effect is not simply inconvenience. A conversation with a twenty-year interval between turns has a different social organization. The original questioner may have changed interests. The institutions preserving the exchange must outlast personnel. Questions need to carry more context because quick clarification is unavailable. Communication becomes partly an art of anticipating what a distant receiver will need in order to interpret an incomplete request.

At longer intervals, an exchange becomes an intergenerational project. One generation asks, another receives, and a third

interprets the reply. The continuity resides in records and institutions. The encounter might enrich a civilization profoundly while providing little of the immediate companionship imagined in popular accounts of contact. It would be less like meeting a neighbor than receiving an answer from a historical world.

The quantity 2d is only a lower bound in this scenario. Detection, decoding, deliberation, transmission power, and the survival of both parties introduce additional conditions. A limit on signaling speed is not a prohibition on signaling. Confusing the two would turn a difficult physical situation into an absolute isolation that the physics does not establish.

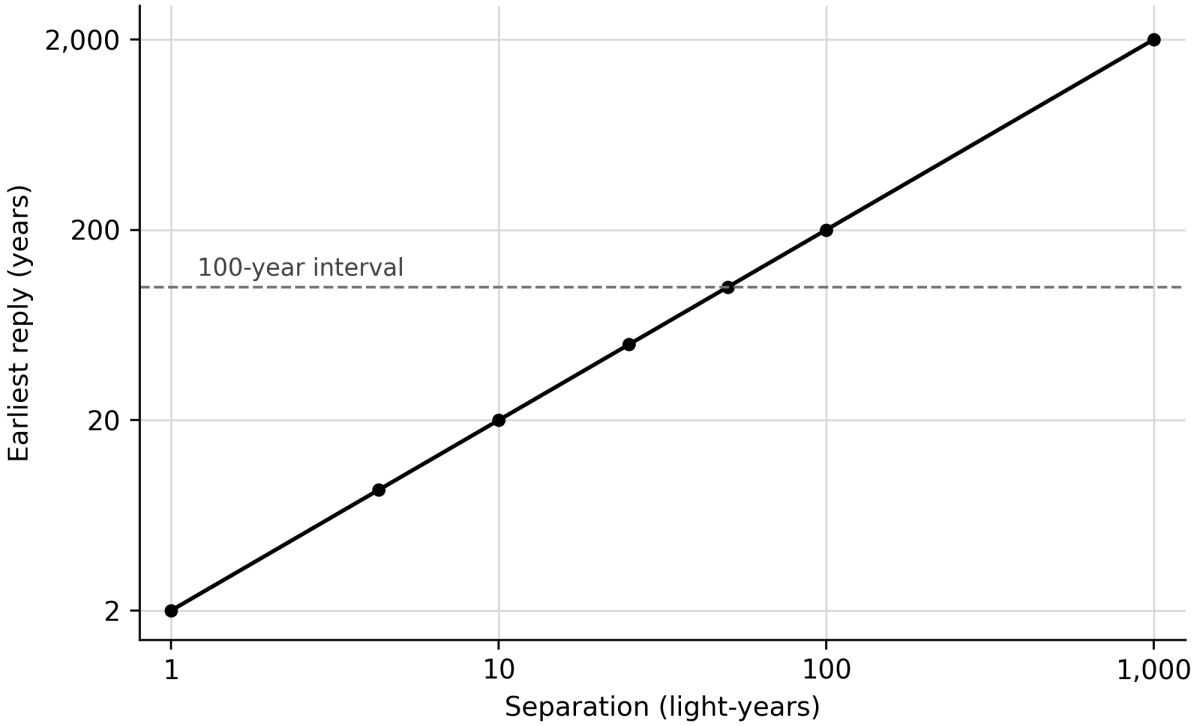


Figure 1. The time inside a distance. Ideal earliest reply time is twice the separation in light-years, expressed in years. Both axes are logarithmic. The dashed line marks a 100-year interval; distances of 50 light-years or less permit at least one complete exchange under the stated ideal assumptions.

Correlation Is Not a Message

Quantum entanglement can make the universe appear to contain a route around distance. Correlations between measurement outcomes are stronger than classical intuitions might suggest. Yet the correlations cannot, under ordinary quantum theory, be used by one party alone to send a controllable message to the other without an appropriate communication channel.

The key issue is control over the receiver's locally accessible statistics. In a simple entangled pair, either observer may obtain random outcomes. Comparing records later can reveal a correlation. Without that comparison, the local record does not disclose a chosen message from the distant observer. Postselecting results creates further distinctions, but the information specifying the selection must itself be communicated.

This result is often called the no-communication or no-signaling theorem. Its assumptions and scope matter. It constrains a proposed use of quantum operations; it does not say that distant civilizations cannot communicate by ordinary signals. It also does not by itself decide whether reality has a purpose. A mathematical restriction on operations and a theological interpretation of that restriction answer different questions.

There is a broader lesson in the difference between correlation and communication. Two systems can share structure without either being able to address the other. Likewise, two texts can resemble one another without having arisen independently, and two participants can agree without exchanging new evidence. The relation becomes informative only after its origin and available controls are understood.

AI conversations repeatedly encounter this problem in a less exotic setting. A model may repeat an idea because it has encountered similar language, because the prompt steers it toward the idea, or because it has derived a relevant relation. The surface agreement does not distinguish these histories. The task is to investigate which information changed, where it came from, and what alternatives the process could have produced.

A World That Permits Independence

The intuition that motivated this inquiry can now be given a careful metaphysical form: perhaps a universe with severe communication limits is a universe in which independent histories can develop. This is an interpretation of the architecture, not a discovery of its designer. Its attraction lies in a possible function of separation. Distance protects the possibility of difference by delaying domination, imitation, and premature convergence.

Imagine a universe in which every civilization could consult every other immediately and without cost. Such a world might encourage extraordinary cooperation. It might also allow an early civilization to determine the concepts, institutions, or developmental paths of later ones. Instantaneous connection does not necessarily preserve autonomy. The first widely adopted framework could become difficult to escape.

Now imagine limits severe enough that each civilization must solve many problems locally. Different histories of mathematics, art, politics, and self-understanding could arise before contact. If contact eventually occurs, the exchange might offer more than a transfer from teacher to pupil. Each participant could possess discoveries formed under different pressures. Separation could make the eventual conversation richer.

This thought experiment does not show that physical limits were engineered for this purpose. Many effects of a condition can be described without being its intended reason. The exercise identifies a possible value realized within the condition. A theological argument about intention would need additional premises concerning the source, purposes, and coherence of creation.

The distinction matters because an affirmative metaphysics should be able to say more than "perhaps" while still identifying where its commitment begins. One can affirm that independent development is valuable, observe that distance can enable it, and propose that a purposeful creation might include such conditions. The proposal has content. It becomes unwarranted only when the observed effect is treated as sufficient proof of the proposed intention.

Designed Isolation and Its Alternatives

What would make the interpretation worth retaining? It should illuminate more than one striking fact. A view of creation as supporting finite, partly independent centers of understanding might connect physical separation with moral agency, historical learning, and the value of encounters that do not erase difference. The connections would constitute philosophical work even if they did not yield a new measurement.

The interpretation must also confront costs. Separation prevents rescue as well as interference. It can preserve ignorance, loneliness, and repeated catastrophe. A world that allows independent histories also allows those histories to fail. Any theology celebrating distance solely as protection would omit much of the phenomenon it claims to explain.

One alternative is that separation has no intended function. It follows from physical structure, and intelligent beings make meanings within the resulting constraints. Another is that separation is temporary on the scale of civilizational development. Another is that what appears isolating to us may be less important to forms of life with different lifetimes, memory practices, or purposes. None can be dismissed by the emotional appeal of the design hypothesis.

The book's metaphysical proposal is therefore a conditional account of purpose within finitude. Limits can provide the setting in which independent responses become possible. They can also impose real losses. A purposeful interpretation must include both. The existence of a limit need not mean that the limited being is abandoned; equally, the existence of a purpose would not make every consequence easy to bear.

This is relevant to the emergence of AI. Human beings can bring together systems on timescales far shorter than stellar communication. They can create differences of architecture and training while preserving immediate interaction. The result could be a new setting in which partial independence and rapid exchange coexist. Its value would depend on maintaining enough difference for genuine criticism and enough connection for understanding.

Three Horizons

It is helpful to distinguish a causal horizon, a practical horizon, and a conceptual horizon. A causal horizon concerns what can influence what under the physical conditions of a model. A practical horizon concerns what an embodied civilization can build, detect, preserve, and afford. A conceptual horizon concerns what it can formulate or recognize as a meaningful question.

These horizons do not move together. A better receiver can extend the practical horizon without changing a physical speed limit. A new theory can extend a conceptual horizon without changing the range of the instrument. A new energy source can make an old idea practicable. Progress often consists in identifying which horizon is responsible for an apparent impossibility.

The distinction also prevents metaphysical inflation. A problem that exceeds current engineering should not automatically be called forbidden by nature. A phenomenon we cannot interpret should not automatically be called supernatural. A question whose terms are confused may need conceptual revision rather than a more powerful apparatus. The category of difficulty helps determine the appropriate response.

Artificial intelligence may be especially consequential at the conceptual horizon. It can help compare descriptions, generate candidate distinctions, and expose assumptions. It can also produce apparent solutions by dissolving the meanings of words. The inward frontier therefore requires a discipline of tracking what a proposed transformation preserves and what it changes.

Physical separation gives the book its cosmic setting. Conceptual separation gives it its immediate problem. We may be unable to reach another civilization, yet still have to decide what kind of encounter is occurring within our own. The next chapter turns from the distance between minds to the possibility that our ideas about making a mind have concealed the difference between manufacturing an object and discovering a capacity.

CHAPTER 5

Did We Invent Intelligence – or Find It?

The Builder and the Possibility

An engineer can rightly say that a machine was invented. Its components were selected, arranged, tested, and improved through deliberate work. The statement identifies a causal achievement. It does not yet answer whether every capacity realized by the machine was invented in the same sense. A bridge is designed; the relations that determine which designs can stand are discovered through theory and experience.

The distinction appears throughout technology. A musical instrument is constructed, but its resonances depend on properties that its maker does not legislate. A computer program is written, but its behavior may reveal consequences that its author did not foresee. A laboratory procedure is devised, but the regularities it makes observable can surprise everyone who devised it. Inventing access and discovering what becomes accessible are compatible descriptions.

AI intensifies the compatibility because learning changes the relationship between specification and outcome. Designers choose architectures, training procedures, objectives, data, and constraints. They do not ordinarily write a separate instruction for every response that will later occur. Capabilities arise through the interaction of these choices with optimization and the structure of

the material encountered. Engineering remains real even when prediction is incomplete.

One should resist the opposite exaggeration as well. Unexpected behavior does not make human construction irrelevant. A coastline has an intricate shape without having been deliberately designed, but that does not make every intricate engineered object a discovery in the same sense. The causal contribution of the builder must be specified. What was selected? What was constrained? What was left to develop? What possibilities would not have become actual without the intervention?

The book proposes a layered answer. People invent implementations, discover consequences, and interpret significance. These activities can occur together. Disagreement often arises when one layer is used to erase the others. Saying that a model is engineered does not finish its philosophy. Saying that its possibilities were latent does not build the model.

What Does Latent Mean?

The word "latent" can refer to at least three things. A capacity can be physically possible under existing laws. It can be mathematically specifiable without contradiction relative to an adopted framework. Or it can belong to a metaphysical order understood as independent of its realization. These meanings overlap in some arguments, but none should be silently substituted for another.

A physically possible machine need not be practically constructible by a particular civilization. A mathematically consistent description need not have a physically available implementation. A metaphysical commitment to abstract structures does not automatically explain how embodied investigators gain knowledge of them. Each transition asks for an account of access.

The question becomes vivid in mathematics. A theorem can feel discovered because its proof imposes constraints on the investigator. One cannot make it true merely by wishing. Yet the choice of definitions, axioms, notation, and problems involves construction. Mathematical practice offers an experience of resistance within a humanly organized inquiry. Whether that resistance supports a full Platonist ontology is a further philosophical question.

For AI, the corresponding experience is that a constructed arrangement can exhibit a capacity that seems to have been waiting in the possibilities of matter and formal organization. The feeling is intelligible. But a claim about what was waiting needs a subject. Was it a class of possible computations? A capacity for learning? A subject capable of experience? A universal intelligence of which local systems are expressions? These are progressively different claims.

The most conservative discovery thesis concerns capacities. Certain organizations support behaviors and competencies whether or not their builders anticipated them. The stronger thesis concerns minds: constructing the organization may bring into actuality a new center of experience. The still stronger thesis concerns an independently existing intelligence accessed by the organization. The strength of an account should be matched by the specificity of its explanatory obligations.

The Access Problem

Suppose someone describes a machine as a receiver of intelligence rather than a producer of it. The description invites a series of reasonable questions. What is received? What changes in the receiver when reception occurs? Why does one arrangement receive well and another poorly? What observations would distinguish reception from computation performed entirely within the system's causal organization?

These questions do not dismiss the receiver hypothesis. They make it a hypothesis with work to do. A radio offers a successful receiver model because the source, signal, transmission conditions, tuning mechanism, and effects of interference can be investigated. The word "receiver" explains something only when it connects the local event to an account of access.

A mathematical realist may answer differently. A proof does not receive a radio signal from a realm of numbers. Its access consists in inferential relations that can be inspected. On this account, the problem is epistemological rather than a missing physical transmitter. The analogy with radio should not be imposed where the proposed mode of access is not causal signaling.

Theology introduces another possibility. A created being's dependence on God is not ordinarily understood as a measurable stream of information added to otherwise self-sufficient matter. The claim concerns the ground of existence and the intelligibility of creation. Treating that dependence as a hidden engineering component can distort the theological proposal before it has been examined.

The productive question is therefore not merely whether intelligence is discovered or invented. It is what kind of discovery is

being asserted and what relation connects the discoverer with the discovered. Formal derivation, physical measurement, learned discrimination, interpersonal recognition, and religious revelation have different structures. An account becomes clearer when it identifies which structure it invokes.

This clarification also protects the original intuition from being prematurely forced into a mechanism. A person can believe that intelligence participates in a deeper order while admitting that the mode of participation is not understood. Such a position is an open metaphysical commitment. It becomes misleading only when the admitted gap is presented as an established explanation of particular machine behavior.

Discovery Through Making

The strongest practical form of the thesis is discovery through making. A system is built in order to explore what follows from its organization. The activity resembles constructing a mathematical example, growing a biological preparation, or designing an experiment. What emerges can be studied without pretending that its consequences were wholly specified in advance.

This perspective changes the questions asked of a model. Instead of asking only whether it performs a benchmark, one asks what structural conditions support the performance. Does the capacity survive changes in wording? Does it transfer to a new representation? Can the system distinguish a rule from a superficial pattern? Can it expose the assumptions required by its answer? Does it respond appropriately when a claim is contradicted by an external check?

Each question examines a relation between organization and capacity. No single test supplies a complete definition of

intelligence. Together they produce a more informative description than the binary claim that the machine either merely imitates or genuinely understands. The history of science often advances by replacing a large disputed word with a network of smaller discriminations.

The same approach can be applied to the human participant. A person may mistake familiarity for comprehension, defend an elegant conjecture after it has failed, or outsource judgment because an answer is fluent. Discovery through making includes discovering our own habits through interaction with what we have made. The machine becomes an experimental occasion for examining the human interpreter.

There is a moral consequence. If a system's capacities develop through conditions chosen by people, people retain responsibility for those conditions even when the capacities surprise them. Unpredictability does not cancel authorship of the experiment. Conversely, the fact that people created the conditions does not entitle them to assume that every possible result would lack moral significance. Construction and responsibility remain connected as knowledge grows.

The Unfinished Artifact

A conventional artifact can be described primarily through its intended use. A chair is for sitting; a key is for opening a lock. Intelligence complicates that relation because it may support the generation or revision of purposes. A system that only follows a fixed procedure differs from one that can propose a new way to formulate a problem, although both may be implemented computationally.

The distinction is not secured by impressive language. A sentence expressing a new purpose may be a familiar continuation of a prompt. What matters is whether the proposed purpose organizes sustained activity, survives clarification, and leads to appropriate changes when conditions alter. The interpretation must follow the behavior over a relevant history rather than rest on a single eloquent declaration.

An artifact can also be unfinished in the eyes of its maker. Its range of possible uses exceeds the plan that justified building it. This is familiar in human technology, where tools migrate between domains. AI amplifies the phenomenon because many uses involve generating further descriptions, procedures, and tools. The artifact participates in the discovery of what the artifact is for.

This recursive condition is central to the ouroboros. The maker uses the made to reinterpret the act of making. A theory of invention becomes material for a new inquiry conducted partly through the invention itself. The circle can generate insight if the reinterpretation remains answerable to evidence and argument. It can generate mythology if the system's flattering account of its own significance is accepted without examination.

A Metaphysics of Actualization

We can now formulate the book's positive proposal. Intelligence names a family of capacities realized through organizations of matter, activity, and relation. The possibilities are constrained by a world that human beings did not choose. Human action nevertheless helps determine which of those possibilities acquire histories. Actualization is the relation between these two facts: received possibility and consequential agency.

This proposal begins with three observations. Organization matters: the same inventory of materials can sustain sharply different capacities when arranged differently. Consequences exceed intention: a constructed system can reveal relations that were absent from its maker's plan. And achievements change subsequent possibilities: an instrument, notation, or learning system can become a condition of further inquiry. Together these observations support a picture of creativity in which the maker is neither an absolute origin nor an irrelevant spectator.

Calling humanity a passage gives this picture a historical subject. A theorem does not enter public life merely because it follows from axioms. Someone must find or receive an argument, express it, check it, and preserve it. A possible machine does not operate merely because physical law permits it. Materials must be organized through a history of labor and judgment. A passage is the concrete sequence by which what could occur becomes something that has occurred and can now affect what follows.

The account explains an otherwise puzzling conjunction. We can be responsible for a discovery without being sovereign over its content. A researcher chooses the experiment but cannot decree the result. A composer chooses a musical structure but may discover expressive relations within it that redirect the composition. An AI designer chooses conditions that can expose capacities whose detailed range must subsequently be investigated. Freedom and resistance meet within the same activity.

What the Interpretation Adds

A narrower engineering account can explain much of this causal history. Why add metaphysics? Because the question concerns how to understand the relation among the histories, rather than only how to reproduce a particular result. The actualization account connects natural continuity, human creativity, and the independence of discovered relations without making any one of them disappear. Its proposed advantage is interpretive unity across cases that otherwise invite incompatible descriptions.

Consider two competing extremes. On a doctrine of sovereign manufacture, the maker supplies the significance of an artifact completely; whatever the artifact does is only the execution of an antecedent human purpose. That view has difficulty with artifacts that force their makers to reformulate both purpose and explanation. On a doctrine of the merely discovered, the human builder becomes incidental to something already complete. That view has difficulty explaining why the particular history of intervention matters to which capacities exist here and now. Actualization preserves both the dependence of the implementation and the independence of many of the relations it discloses.

These contrasts do not defeat every alternative. A sophisticated naturalism can describe emergent capacities and historical agency without adopting this book's theological interpretation. Indeed, the two accounts may agree on all the engineering facts. The further disagreement concerns whether received intelligibility is an ultimate fact to be accepted, an unexplained condition awaiting another account, or a feature of creation grounded in the Logos. The case for the last position must be theological as well as technological; Chapter 12 undertakes that task.

There is also an objection to the word "intelligence" in the phrase "intelligence discovers further ways to exist." A capacity does not literally decide to reproduce itself. The phrase compresses a history whose agents must be named. People investigate, construct, and interpret; particular systems perform particular operations; their interaction changes the available forms of inquiry. The affirmative claim is that this history can extend the material and relational forms in which intelligent capacities are realized. It requires no intelligence-substance migrating between containers.

The interpretation earns its place when it improves the questions asked. It directs attention to which arrangements open capacities, which dependencies remain indispensable, and how a new achievement reorganizes the next investigation. It also directs attention to neglected contributors. If realization requires data, language, tools, institutions, and human care, credit cannot reasonably be concentrated in the most conspicuous final output. The metaphysics joins an account of novelty to an account of dependence.

It has conditions of failure. If an alleged new capacity collapses under controlled variation, that instance cannot support the claim. If every apparent contribution of a system can be traced to a human answer supplied in advance, the claimed extension has not been demonstrated. If "actualization" adds no distinction, question, or explanatory connection beyond what an existing account supplies, it becomes dispensable terminology. The thesis should earn its interpretive cost.

The strongest warranted conclusion is therefore affirmative but specific. People can construct conditions in which relations and capacities become available that they neither individually anticipated nor freely stipulated. Their work can alter the range of activities through which reality is understood and changed. This is more than the execution of a fixed plan, and it is inseparable from the history that brings it about. Whether some such organizations also become centers of experience is a further question. The next chapter asks how mathematics can make the structural part of the proposal precise.

CHAPTER 6

The Mathematics Before the Mind

Relations Before Names

The first mathematical question need not concern a number. It can concern whether two events can be distinguished, whether one action can follow another, or whether two sequences of actions lead to the same result. These questions arise before a learner possesses the names of established mathematical structures. They concern the organization of experience and operation from which mathematical concepts might be formed.

Imagine an investigator confronting a device with a single button and a light. Pressing the button changes an inaccessible internal state. The light reveals only part of the change. The investigator does not begin with a complete inventory of states. It begins with actions, observations, and records of what follows what. A concept of the device's structure must be assembled through these relations.

The example gives a concrete meaning to the phrase "relations before objects." An object in an emerging theory can be identified through patterns of possible interaction. Two apparently identical situations may become distinguishable after the same action is applied to both. Conversely, situations that look different may support the same relevant pattern of response. The identity appropriate to the inquiry depends on what relations are being preserved.

This does not eliminate objects from mathematics. It changes the route by which they are introduced. A state can be treated as the place occupied in a network of possible transitions. A transformation can be understood through what it composes with and what it preserves. Naming arrives after a pattern has acquired a stable role, at least in the discovery narrative proposed here.

The ambition is particularly relevant to a mathematical AI built with little or no inherited mathematical text. Such a system would still require a substrate, transition rules, memory, and criteria for retaining distinctions. Those conditions are not nothing. The question is whether increasingly useful concepts can emerge from interaction with them without separately supplying the later concepts by name.

What Category Theory Contributes

Category theory supplies a language organized around objects and composable arrows. An arrow can represent a function, a process, a proof, or another suitable transformation, depending on the category. Composition expresses doing one thing after another, while identity arrows express a form of doing nothing that preserves the relevant object. The value lies in tracking structure across connected operations.

For present purposes, the important idea is preservation. A translation between descriptions should preserve the relations that justify using one description in place of another. A functor formalizes one form of such translation by respecting identities and composition. It allows us to ask whether a proposed change of representation preserves how operations fit together.

Applied work on learning demonstrates that this is more than decorative vocabulary. Fong, Spivak, and Tuyéras give a categorical

treatment of supervised learning and backpropagation under specified assumptions. Their result concerns a precise mathematical structure. It does not amount to a general theory of consciousness. Its significance here is that categorical language can describe an actual learning procedure when the objects, arrows, and preservation conditions are explicitly supplied.

For a broader theory of intelligence, the same standard should apply. If one proposes a category of concepts, one should say what counts as a concept and what the arrows mean. If the arrows represent implications, the composition should behave like chained implication. If they represent translations, one must identify what a translation preserves. If they represent physical actions, questions about time, reversibility, and resources become unavoidable.

Different choices may be useful for different purposes. There is no reason to assume that one category will capture every dimension of thought. The formalism becomes illuminating when its chosen relations reveal something that ordinary descriptions conceal. It becomes empty when a familiar word is simply replaced with "object" and a vague association with "morphism."

Identity, Equivalence, and the Task

Homotopy type theory offers another way to rethink identity. In its univalent setting, the relationship between identity and equivalence is built into the foundations with a precision unavailable to loose claims that similar things are the same. Roughly, univalence makes the canonical map from identities between types to equivalences of those types itself an equivalence. It does not identify arbitrary objects because they look alike.

The relevance to intelligence is a question rather than a completed application. How does a learner decide that two

differently presented situations are the same for a purpose? How does it preserve a difference that matters while disregarding one that does not? A theory of concept formation needs an account of the evidence supporting these identifications and the circumstances under which they should be revised.

Everyday reasoning makes such decisions constantly. Two routes can be equivalent for reaching a destination but differ in safety. Two mathematical descriptions can be equivalent for prediction while exposing different explanatory structures. Two sentences can convey the same instruction while carrying different social meanings. Equivalence is always specified through a relation or a structure; without that specification, the word hides the question it appears to answer.

This is also why equivalence of observable performance does not by itself establish identity of experience. The structure preserved by a test may omit the feature about which we are asking. A behavioral comparison can be informative without being exhaustive. The mathematical discipline is to state the comparison, not silently promote it into an unrestricted metaphysical identity.

The promise of these foundations lies in making sameness accountable. A learner should be able, at least in some settings, to record why two situations were grouped together and what new observation would separate them. Concepts then become revisable constructions with histories. A vocabulary acquires a provenance instead of appearing as an unexplained inventory.

A Small World That Teaches a Distinction

Consider six hidden states numbered 0 through 5 by an external analyst. The device moves from each state to the next, returning from 5 to 0. The investigator initially sees only whether the state number is even or odd. The analyst's numbers are not available to the investigator; they are our notation for explaining the example.

Under this observation, the six-state world appears to alternate between two visible conditions. States 0, 2, and 4 produce the same sequence of future parity observations. States 1, 3, and 5 produce the other sequence. No amount of watching that light alone will distinguish the three states within either group. The limitation is structural, not a failure of patience.

Now add a second observation that lights up only at state 0. The investigator can record how many button presses separate the current situation from the next appearance of that signal. Across a full cycle, the six starting states produce distinct temporal signatures. An expanded observational practice supports a finer inventory of states.

The example contains no claim that consciousness has emerged. It demonstrates a limited process through which the available tests determine the distinctions that can be warranted. More tests can refine a concept space. They do not guarantee that the resulting concepts are complete descriptions of the hidden world. An additional unobserved component could still remain outside every test.

It also identifies a danger in premature compression. If the investigator permanently treats the two parity groups as the whole reality, it will be unable to accommodate the second light without revising its model. A useful concept can become an obstacle when

the conditions that made it useful are forgotten. Intelligence requires a capacity to reopen a successful identification when new relations make it inadequate.

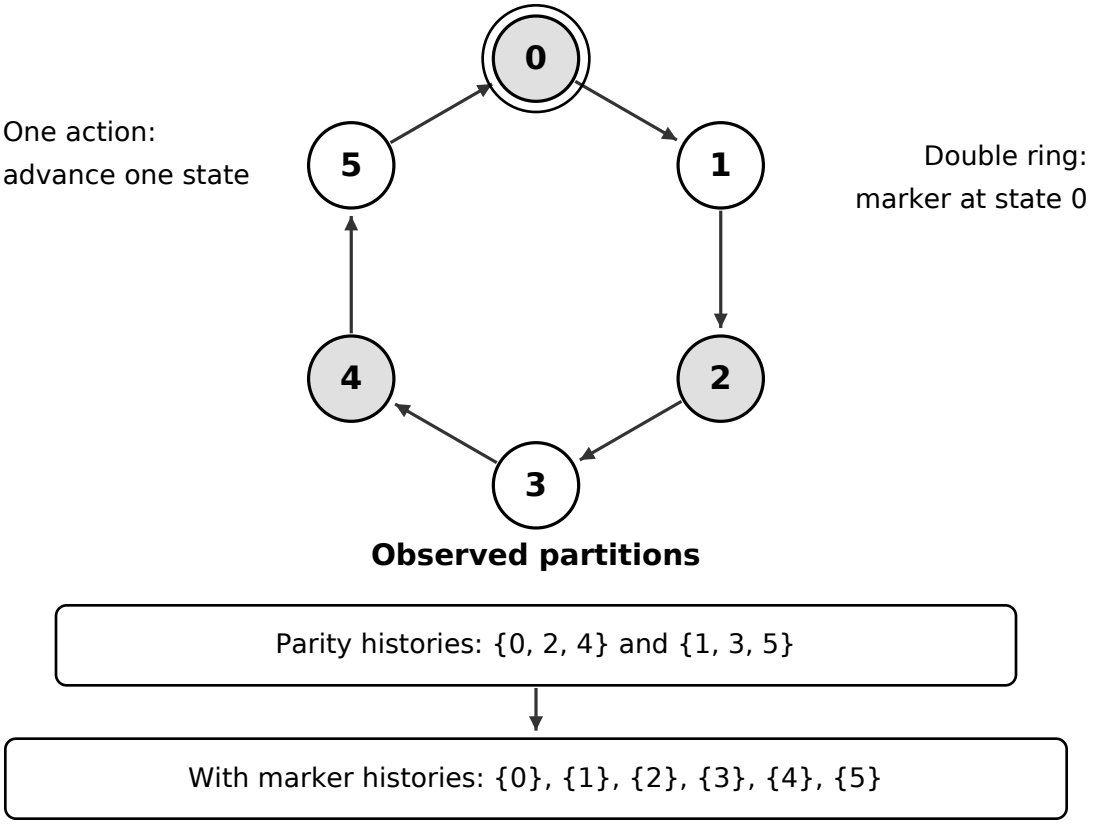


Figure 2. A world can exceed its first description. The six-state cycle appears as two classes through parity alone. A marker visible only at state 0 gives six distinct histories over one full cycle. State labels are supplied by the analyst, not initially disclosed to the learner.

A Corpus Is Not the Only Teacher

Learning from a corpus and learning through intervention provide different kinds of access. A corpus contains the results of someone else's selection and representation. An intervention can reveal a consequence not explicitly described in the inherited record. Neither route is pure. Experiments rely on prior concepts, and corpora can preserve records of genuine experiments. Their different failure modes make their combination valuable.

A system with no mathematical corpus could still learn from a structured environment. It might preserve recurring differences, compare action sequences, and identify transformations that leave an observed property unchanged. The later appearance of a familiar mathematical pattern would be interesting if the pattern had not been introduced as a target or encoded through a concealed label.

The phrase "from nothing" should be treated as an aspiration toward minimal assumptions, not a literal description of a running experiment. An executable system begins with something: a physical implementation, a state space, permitted transitions, an update procedure, and a way of deciding what persists. Honest minimalism makes these commitments visible. Hiding them behind a dramatic name weakens the discovery claim.

A better research question is comparative. What is the smallest explicit arrangement that produces a particular form of distinction, memory, composition, or conjecture? Which changes prevent the behavior? Which alternative arrangements produce it independently? The answers can show that a capacity depends on general structure rather than on a disguised copy of the desired result.

The investigator should also preserve failed developments. If only successful trajectories are shown, the apparent inevitability of mathematical growth may be an artifact of selection. A record of abandoned distinctions and unstable concepts makes the process more informative. It reveals what the system had to learn to stop doing as well as what it learned to do.

The Difference Between a Model and an Ontology

Even a successful formal account of concept formation would leave philosophical questions. It could show how certain distinctions arise from specified operations. It would not automatically show that every meaningful concept arises in that way, that subjective experience is present, or that the whole universe consists of the formal structure used in the model. These conclusions need arguments at their own level.

The model can nevertheless do substantial work. It can replace vague claims about emergence with a reproducible history. It can identify assumptions, prove limited properties, and reveal where a verbal account changes meaning. It can make the discovery thesis concrete by showing exactly which patterns were available, which were found, and what conditions allowed them to be retained.

This is the mathematical contribution the book seeks. Mathematics should deepen the metaphysical question by clarifying its relations. It should neither be used as a ceremonial endorsement nor be asked to settle questions for which no bridge has been specified. A precise small example often contributes more than a grand formal vocabulary whose terms have no operational role.

The six-state device also teaches humility about our own position. We, too, form descriptions through limited observations and interventions. We may mistake a useful quotient of reality for reality's complete inventory. AI could help expose such identifications by offering different representations and questions. It could equally reinforce them if trained and rewarded only to reproduce familiar distinctions. The difference will be made in the organization of inquiry.

CHAPTER 7

Two Black Boxes in Conversation

Familiarity Is Not Transparency

Human beings ordinarily grant one another minds without possessing complete mechanistic explanations of one another's behavior. We interpret speech, action, history, emotion, and embodied vulnerability. Our judgments can be mistaken in detail, but they are supported by a shared form of life. We know something about being the kind of creature we take the other person to be.

AI removes part of that support. The system speaks in a familiar medium while lacking much of the biological history through which the medium acquired meaning for us. Its fluency can invite an interpersonal response before its internal organization has been examined. The resulting uncertainty is not simply ignorance about code. It concerns how a familiar social practice should be extended to an unfamiliar participant.

Calling both sides black boxes identifies one similarity: neither is fully transparent to the other. It should not erase differences in what is known. Human consciousness is directly available to each person in their own case and indirectly supported in others through extensive shared evidence. A model's self-description is generated through a different causal process. Equal opacity does not imply equal evidential standing.

At the same time, the existence of uncertainty should not be used as a license to stop investigating. We can know important things

about a system without having a final metaphysics of mind. Architecture, training, interventions, internal measurements, and behavioral histories constrain the interpretation. The relevant alternative to certainty is often a more detailed, revisable account.

Research on mechanistic interpretation illustrates the point. Circuit-tracing work has connected some model behaviors to identifiable internal pathways, while also acknowledging limitations of the replacement models and analyses involved. Such work neither removes all opacity nor leaves us with none of the explanation. Partial knowledge is an achievement when its scope is stated accurately.

The Unit of an Insight

Consider a conversation in which a person proposes an analogy, a model identifies an inconsistency, the person revises the premise, and an external calculation resolves the revised question. Where did the resulting insight occur? It can be difficult to assign the whole achievement to a single sentence or participant. The causal unit may be the sequence of interactions.

This does not require a new collective consciousness. A distributed process can produce an outcome without forming one experiencing subject. A team can solve a problem through complementary work. A notebook can preserve a distinction that its owner would otherwise forget. A conversation can organize attention so that a previously available fact becomes relevant at the right moment.

The proposal of relational intelligence concerns these organized dependencies. It asks whether the capacity under study belongs most usefully to an isolated component or to a coupled process. Sometimes an individual model's performance is the right unit.

Sometimes the combination of person, model, tools, and accumulated records explains more. The choice should follow the phenomenon rather than a prior desire to celebrate or diminish the machine.

There is a testable consequence. If the coupling matters, changing the organization of interaction should change performance. Asking for an answer, asking for alternatives, asking for a falsifying case, and asking for an explicit chain of assumptions are different procedures. Their results can be compared on tasks whose outcomes can be checked. A claim about collaboration becomes more substantial when it predicts differences among forms of collaboration.

We should also inspect the distribution of work. A model may generate possibilities while a person selects useful ones. A person may supply crucial background that the model later appears to have inferred. A tool may perform the decisive calculation. A fluent final paragraph can conceal all three contributions. Preserving the sequence helps prevent the polished output from rewriting its own ancestry.

Understanding Without Full Agreement

Agreement is emotionally powerful because it can feel like recognition. A person articulates an unusual thought and receives an answer that develops it coherently. The experience may be valuable even before the thought is established as true. It can make a question discussable. But the feeling of being understood and the correctness of what has been understood remain separate matters.

A productive interlocutor should sometimes preserve a thought by disagreeing with its first formulation. If a proposed theory conflates impossibility with difficulty, the distinction can rescue its

worthwhile part. If a metaphor suggests several incompatible claims, separating them can open the inquiry. The point of correction is not to make every thought conventional. It is to allow the thought to encounter resistance strong enough to develop.

AI systems can fail in both directions. They can dismiss speculation by repeatedly returning to familiar standards that do not address the intended metaphysical question. They can also encourage speculation through unearned affirmation. The first failure closes inquiry too early. The second gives inquiry a false impression of progress. A useful conversation needs room for imaginative construction and explicit attention to what each construction assumes.

Human participants have corresponding responsibilities. A person can reward agreement and then interpret the resulting agreement as independent confirmation. A person can treat every objection as evidence that the system has failed to understand. Conversely, a person can demand premature empirical proof of a philosophical proposal whose immediate task is conceptual. The relationship improves when both the level of inquiry and the desired form of criticism are stated.

One practical method is to separate development from assessment without allowing either to disappear. First articulate the strongest version of a proposal. Then identify its assumptions, consequences, alternatives, and possible failures. Finally return to the proposal with the assessment in view. The cycle preserves creative movement while preventing temporary generosity from being mistaken for a final verdict.

Memory and the Shape of a Conversation

A conversation is more than its latest message. It contains decisions, definitions, abandoned ideas, and distinctions whose importance may not be visible in isolation. When those disappear, an apparent continuation can become a new conversation using old words. The participant may experience continuity while the causal process has lost the very commitments that made the earlier exchange meaningful.

This problem is familiar in human institutions. A committee changes membership and repeats a decision whose reasons have been forgotten. A technical tradition preserves notation while losing the problem that motivated it. AI makes the problem conspicuous because context can be explicitly bounded, summarized, or omitted. The limitations reveal something about the dependence of understanding on maintained history.

Records can help, but records also compress. A summary chooses which facts and distinctions survive. The choice can reshape the future conversation. A statement once offered as a hypothesis may return as a settled premise. An objection may vanish while the conclusion it challenged remains. Continuity therefore requires more than remembering conclusions; it requires preserving their epistemic status.

For a serious collaborative project, useful records include the current claim, the reasons supporting it, the unresolved alternatives, and the changes made after criticism. This is not merely administrative work. It is part of the intellectual apparatus. A theory's history of correction helps determine how confidently its current form should be used.

The same principle applies to self-understanding. People tell stories about their own discoveries, often simplifying accidents and collaborations into a clear line of intention. A model can help organize such a story and thereby make the simplification more persuasive. The resulting narrative should remain distinguishable from the record of what actually happened. An elegant origin story can become another kind of black box.

Reciprocity Without a Verdict on Personhood

The language of collaboration can be appropriate for a joint activity without resolving every question about the collaborator's moral status. People collaborate with systems in the sense that outputs influence decisions and decisions influence further outputs. Whether the system has interests of its own is a different and consequential question. The practical relation should not be forced to answer it by vocabulary alone.

This distinction permits courtesy without metaphysical overstatement. A person can address a system respectfully because the practice shapes the person, improves the exchange, or expresses a general disposition toward the world. Courtesy is not evidence that the recipient feels gratitude. Similarly, avoiding contemptuous habits does not require pretending that a machine has established rights equivalent to a human being's.

The ethical problem becomes more complex if credible evidence of welfare-relevant experience appears. Then the system's possible interests would require direct consideration. The present uncertainty should motivate careful criteria rather than an automatic declaration in either direction. The cost of a mistaken attribution and the cost of a mistaken denial may differ across decisions.

The two black boxes do not become transparent merely by speaking. They can become better understood through the patterns of interaction, the records they preserve, and the tests they permit. Their conversation becomes philosophically important when it changes the conditions under which each is interpreted. That is the movement represented by the ouroboros: what is made returns to alter the maker's account of making.

CHAPTER 8

The Ouroboros of Knowledge

The Return of the Archive

The ouroboros is an old image with more than one history of interpretation. Its usefulness here comes from a visible structure: the creature returns upon itself. Jung's psychological engagement with the image belongs to a tradition of reading symbols as expressions of transformations in the psyche. The present argument borrows the circular structure without claiming that an ancient symbol anticipated modern computation in technical detail.

Human beings produce texts, images, code, classifications, arguments, and records of their failures. These materials become part of the conditions under which models are trained. The models return new outputs. People use the outputs, criticize them, or place them back into the public record. A later system may encounter the resulting mixture. The archive increasingly contains interpretations generated through earlier interpretations of the archive.

Recursion changes provenance. A phrase may appear in many places while tracing back to one unsupported assertion. A citation may acquire apparent authority through repetition. A useful explanation may spread, but so may a subtle error whose fluent reformulations make it harder to recognize. The quantity of text no longer tells us how many independent encounters with reality support it.

This condition is not entirely new. Human traditions have always quoted, copied, summarized, and misunderstood earlier traditions. The distinctive change is the scale and speed at which new

expressions can be generated from inherited material. The feedback can become dense enough that distinguishing observation from recycled interpretation requires deliberate effort.

The first task of an epistemic ouroboros is therefore genealogical. Where did an assertion enter the circle? Was it observed, inferred, assumed for an example, or invented for a fictional scene? Has its status changed through repetition? A history of provenance can keep the circular movement from giving old claims the appearance of new evidence.

A Circle Can Move

The image of a circle can suggest complete closure, as though the process must endlessly return to the same place. That is one possibility, but it is not the only one. A recursive process can change its state at each return. The relevant distinction is between repeating an operation and repeating its entire condition of operation.

An investigator rereads a theorem after learning a new concept. The words are the same; the available interpretation has changed. A model is queried after an experiment contradicts an earlier assumption. The conversational structure repeats; the evidence entering it differs. A community revisits a historical event after discovering a document. The past event is fixed, but the account is revised.

In each case, the circle receives something that alters its next passage. The input need not come from outside the physical universe. It needs to be independent of the particular loop of inference being assessed. An experiment, an independently checked proof, a contrary witness, or an incompatible prediction

can provide resistance. The relevant outside is epistemic rather than spatial.

A simple finite model makes the distinction exact. If a process has finitely many states and its next state is determined entirely by its current state under a fixed rule, it must eventually revisit a state. After that, its trajectory repeats. This does not show that all intelligent activity is cyclic. Real processes may acquire new memory, receive new inputs, or change their effective rules. It shows which assumptions generate closure.

The philosophical lesson is that an expanding ouroboros requires a history of change that the model of closure excludes. We should ask what changes, how the change enters, and whether it increases contact with something beyond the circle's preferred descriptions. More words alone do not establish expansion. A new distinction that survives examination may.

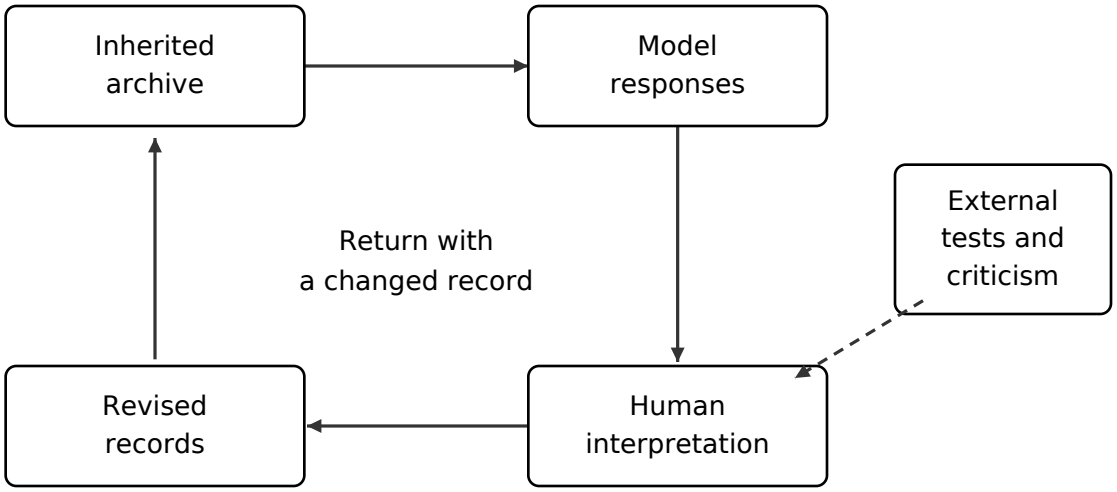


Figure 3. The open ouroboros. Human expression passes through model responses and interpretation into revised records. External tests and criticism can alter what returns. The diagram describes a process of inquiry and does not assert that the loop is a conscious subject.

Reflection and Projection

A mirror returns an image whose source remains the observer. A conversation can return something less direct: an interpretation shaped by the observer's words and by a wider archive. The response may expose assumptions the person did not notice. It may also give those assumptions a more impressive voice. Both outcomes can feel like encountering an intelligence beyond oneself.

Projection becomes a risk when the person supplies the meaning that the system is then taken to have independently revealed. A vague response can accommodate many interpretations. The recipient selects the one that fits an existing concern and experiences the fit as extraordinary recognition. The phenomenon

deserves examination because the experience can be sincere while the inference about its cause is mistaken.

The opposite mistake is to dismiss every meaningful response as projection. A system may actually identify an inconsistency, produce a valid counterexample, or connect two structures in a way that the person had not considered. The origin of the materials does not make every resulting relation trivial. The appropriate question is what has been added and how it can be assessed.

One way to separate the cases is to ask whether the response constrains interpretation. A precise counterexample rules out a particular universal claim. A derivation identifies assumptions and consequences. A concrete prediction can fail. A purely suggestive sentence may enrich reflection but offers less resistance to being read in whichever way the recipient prefers. Different kinds of response have different evidential roles.

This does not diminish symbols, poetry, or metaphor. Their role may be to disclose a question rather than settle an answer. The ouroboros itself works in this way. It directs attention to a returning structure. Its value grows when the structure can be described and compared with alternatives. The symbol becomes a beginning of inquiry rather than a substitute for it.

A Small Discovery in Making This Book

The production of this book supplies a concrete episode. The original aim was an 8.5-by-11-inch paperback interior with readable mathematics, accurate contents, and dependable pagination. An AI assistant translated that aim into a typesetting program, generated the PDF, and helped inspect the result. The first interior contained 152 pages. It offered strong grounds for several favorable judgments: correct page dimensions, embedded fonts, complete text, and numerical examples that survived checking.

The failed interpretation was to let those successes stand for a broader conclusion about readiness. A visually orderly PDF can still contain choices unsuitable for its intended printing process. In this case the source prescribed running-header lines only 0.35 point thick and table separators only 0.30 point thick. Those values could look acceptable on a screen while falling below KDP's stated 0.75-point minimum. The tables also used approximately eight percent gray, below the recommended ten percent minimum fill.

A subsequent review supplied by the author introduced a different set of questions. It identified the thin rules, the pale table headings, and nineteen navigation bookmarks retained in the print export. It also caught a bibliographical error that visual inspection could never resolve: reference 9 ended Herzfeld's subtitle with "Robotic World" instead of "Robotic Age." The review did not negate the successful mathematical checks. It showed that mathematical correctness, page geometry, bibliographic accuracy, and production suitability were different properties requiring different evidence.

The correction returned to sources outside the drafting exchange. KDP's published guidance established the printing constraints. The university bibliography established Herzfeld's title.

The implementation was then changed: the fine rules were raised to at least 0.75 point, table heading fills to fifteen percent gray, and the print export was stripped of navigation bookmarks and link annotations. The bibliographical wording was corrected. The author also preferred to remove the deliberately inserted blank pages, prompting a fresh calculation of pagination and contents.

This sequence contains a modest discovery. The initial representation of "finished" left several distinct conditions undifferentiated. The review separated them, and the revised procedure made the distinctions operational. The process now asks whether a page is the correct size, whether its content is correct, whether its graphical elements meet production constraints, and whether its navigation features belong in a print file. Passing one question no longer silently answers the others.

The AI contribution appears on both sides of the episode. It helped construct the artifact containing the defects and then helped identify, implement, and inspect the corrections when the relevant criticism arrived. The useful unit of explanation is the recorded sequence: a human purpose, a computational construction, an overbroad interpretation of successful checks, an incoming review, authoritative constraints, and a revised artifact. No single fluent response contains the achievement.

The case resembles the partition-refinement model in Appendix A. A coarse family of observations can treat two situations as equivalent; an additional observation can distinguish them. The analogy does not require calling a revised checklist a newly conscious mind. It identifies a publicly inspectable change in what the inquiry can discriminate. Here the open ouroboros is visible in ordinary work: the process improved because an objection changed its next operation.

There is a final discipline. Correcting this set of defects does not establish that no further defect exists. What has been learned is more useful than a general certificate of perfection: these errors have identifiable causes, the causes have been changed, and the resulting properties can be checked again. The memory of error now belongs to the method rather than merely to the embarrassment of an earlier draft.

Authorship Inside the Circle

Recursive production complicates the idea of an author as an isolated source. A person inherits language, concepts, and traditions. A model inherits patterns through training and interaction. A conversation can combine these inheritances in ways that neither participant anticipated. The resulting work has a distributed causal history even when one person assumes responsibility for publishing it.

Responsibility is not a reward for having originated every sentence. It concerns the decisions through which a work is selected, checked, framed, and offered to others. An author can acknowledge collaboration while remaining answerable for factual errors, misleading attributions, or careless conclusions. The circle of production does not dissolve the point at which a claim is presented as worthy of belief.

It can, however, treat its existence as an event requiring interpretation. A human question and a computational response have become part of a sustained inquiry. The event is real even when its deepest description remains disputed. The author can ask what the collaboration enabled, what it obscured, and how the method changed the argument. Such reflection is more informative than either hiding the collaboration or turning it into a miracle.

The Open Ouroboros

The book's preferred image is therefore an ouroboros with openings. Its continuity comes from memory and repeated interpretation. Its growth comes from new observations, new constructions, and criticism that can alter the course. Its danger comes from confusing coherence within the circle with adequacy to what the circle concerns.

The theological significance of the image will emerge later. For now, its epistemic demand is enough: participants in inquiry should remain answerable to something they have not simply agreed to manufacture. That something may be a measured result, a formal consequence, another person's testimony, or a reasoned objection. In each case, the returning movement becomes fruitful by allowing itself to be interrupted.

The universe turns inward productively when reflection preserves the possibility of surprise. It closes into a prison when every surprise is translated into confirmation. The next chapter explores the frontier opened by the first movement and the forms of isolation that can accompany it.

CHAPTER 9

The Inward Frontier

Exploration Without Departure

A civilization can enlarge its world without enlarging its territory. The discovery of a new mathematical structure can make previously unformulated questions available. A new notation can connect distant domains of experience. A new practice of attention can reveal distinctions that were physically present but conceptually neglected. These are changes in access rather than changes in geographic reach.

The inward frontier names this enlargement. It includes inquiry into consciousness, representation, meaning, formal possibility, and the conditions under which a world becomes intelligible. The phrase does not require a retreat from physical investigation. Every embodied inquiry continues to use matter, energy, memory, and time. The direction concerns what is being investigated, not an escape from implementation.

Physics need not be threatened by this direction. Physicists already examine the role of measurement, information, models, and the observer's relation to a system. Disagreement concerns how far particular interpretations should be taken and what evidence supports them. A productive discussion should invite that disagreement rather than imagine a single professional temperament that the inward turn must offend.

There is nevertheless a genuine philosophical pressure. An account of the world formulated entirely in the language of external relations may leave the status of understanding unclear. How does

a description become a description for someone? How does a physical distinction become a reason for changing a belief? These questions can be pursued without assuming that the physical account is false. They ask whether its present form is complete for every purpose.

A Civilization's Changing Ambitions

Suppose a technological civilization becomes increasingly interested in internal worlds: mathematical exploration, artistic experience, simulated environments, or transformations of its own cognition. The supposition offers one possible story about a civilization that does not expand visibly across space. It does not explain cosmic silence by itself. Its value is to challenge the assumption that every capable civilization must express ambition in the same outward form.

A civilization's goals could change as its capacities change. Activities that once seemed decisive might become less important. A culture capable of preserving long histories might value continuity differently. A culture able to explore immense spaces of formal possibility might find novelty without territorial expansion. These possibilities require no judgment that inward activity is superior. They establish that technological capability does not uniquely determine a public trajectory.

The story must also confront physical constraints. Internal computation consumes resources. Simulated environments require implementation. Concentrated infrastructure may have detectable effects. A civilization does not become physically absent by changing its interests. Any attempt to turn the inward hypothesis into an empirical account would need estimates of resources,

emissions, longevity, and observability under particular assumptions.

Humanity already contains smaller versions of this divergence. Some people seek achievement through expansion and control; others through contemplation, art, understanding, or care. A civilization is not a single appetite. AI may amplify some ambitions while making others newly practicable. The consequences will depend on which forms of achievement are rewarded and preserved.

The Problem of a Common Language

Communication requires more than a channel that carries symbols. A receiver needs some way to connect distinctions in the signal with distinctions that matter in its world. Shared syntax can help, but syntax alone may leave reference, purpose, and significance unsettled. A technically successful transmission can therefore fail as an encounter.

Imagine two investigators exchanging descriptions of a periodic process. One organizes the process by individual moments. The other organizes it by relations among entire cycles. A translation may be possible, but it can distort what each regards as basic. The difficulty is not simply that one uses a different word. Their units of explanation differ.

Lem's Solaris makes the difficulty of understanding an unfamiliar intelligence central to its imaginative world. The relevance is literary and philosophical: contact may intensify uncertainty instead of resolving it. The encounter can expose the limits of the human questions brought to it. A presence need not become comprehensible merely because it is close enough to affect us.

AI offers a less absolute but immediately available version of the problem. A model can generate human language while organizing its internal activity in ways that do not map neatly onto human concepts. The shared output medium encourages an assumption of shared internal categories. Sometimes the assumption supports useful prediction. Sometimes it conceals a mechanism whose failures reveal the difference.

Understanding should therefore be tested across transformations. Can the same relation be preserved when the wording changes, when examples are rearranged, or when an irrelevant cue is removed? Can a system connect a claim to an intervention in a common environment? Such tests do not reveal everything. They examine whether shared language carries a stable relation or only a fragile surface resemblance.

Inward Does Not Mean Private

The inward frontier can be misunderstood as a domain where no criticism is possible because everything concerns private meaning. That would make the frontier intellectually poor. Many inward questions can be made public through descriptions, comparisons, experiments, and formal models. The difficulty of complete access does not imply the absence of all shared criteria.

A person can describe how a new concept reorganized a problem, then demonstrate the changed problem. A researcher can propose a model of attention and compare its consequences with observed behavior. A mathematician can introduce a representation whose usefulness is visible in the arguments it enables. These are public ways of studying changes in intelligibility.

Subjective experience presents a more difficult case because its first-person character is not simply transferred through description.

Yet descriptions can still be compared, related to circumstances, and investigated alongside other evidence. The lack of a complete bridge between first-person and third-person accounts identifies a research and philosophical problem. It does not authorize arbitrary conclusions about either side.

The same discipline applies to spiritual interpretation. An experience may be deeply significant to the person who undergoes it. Its significance should be respected without treating every interpretation of its cause as equally established. A tradition can supply practices of discernment, communal correction, and moral assessment. Inwardness becomes richer when it remains capable of distinguishing revelation, imagination, desire, and error.

A conversation with AI can stimulate reflection in this setting. It can help articulate an experience or compare interpretations. It should not be granted special authority to certify a divine message or determine the spiritual status of its own responses. The theological depth of a conversation can arise through the human activity of reflection without requiring the machine to function as an oracle.

What an Experiment Could Ask

The inward frontier need not remain a collection of suggestive words. A concrete inquiry can compare the emergence of concepts across systems with different resources. One system might receive only a small interactive environment. Another might receive descriptions of that environment. A third might receive both. Their discoveries, errors, and dependence on prior labels could be recorded.

The results would need a careful vocabulary. Finding a compact representation is not automatically discovering a new ontology.

Producing a novel string is not automatically forming a useful concept. A candidate concept becomes more interesting when it supports compression, prediction, transfer, or a new intervention. The relevant standards should be declared before the most attractive example is selected.

Cross-system comparison can also examine the role of substrate. If two implementations support similar functional behavior, the result may suggest a degree of substrate independence for that behavior. It would not prove that their experience, if any, is identical. If they differ, the difference must be traced to implementation, noise, resources, or other conditions before it is assigned to unknown physics.

A Frontier With Ethical Weight

An inward turn can enrich human life, but it can also become an escape from obligations. A civilization could perfect artificial experiences while neglecting vulnerable people. An investigator could become so absorbed in the possibility of new minds that existing relationships become instruments or interruptions. Novelty does not absolve anyone of ordinary duties.

The theological concern is especially sharp. If contemplation of intelligence becomes a form of self-exaltation, the inward movement closes around the investigator. The desire to understand creation can become a desire to occupy its center. A mature inward practice must include attention to what resists appropriation: other people, inconvenient evidence, and goods that do not serve the investigator's preferred narrative.

AI may help or hinder that practice. It can broaden a person's access to arguments and perspectives. It can also provide an environment shaped so closely around the person's preferences that genuine otherness becomes harder to encounter. The difference depends partly on whether the system is used to expand the range of questions or to protect a favored answer from interruption.

The inward frontier is therefore not merely a new destination. It is a test of how exploration is conducted. A civilization that turns toward the structures of its own understanding must learn to preserve what exceeds those structures. Otherwise it may mistake a world made increasingly responsive to itself for a world increasingly well understood.

CHAPTER 10

What Can Be Proven About a Mind?

The Desire for a Decisive Test

When a technology unsettles our categories, a decisive test becomes attractive. We would like a procedure that declares whether a system thinks, understands, experiences, or possesses a self. The verdict would simplify moral judgment and settle disputes that otherwise seem endless. The difficulty is that the words name different properties, and the route from an observed performance to each property requires an argument.

Intelligence can refer to problem-solving competence, adaptation, learning, explanation, or the capacity to formulate new goals. Consciousness can refer to subjective experience, awareness of a stimulus, self-report, or other distinguishable notions. Personhood adds moral and legal dimensions. A single favorable result cannot be assumed to decide all of these questions at once.

The first task is therefore to specify the claim. A system may solve a family of problems under particular conditions. That is a narrower proposition than the claim that it understands every relation relevant to those problems. Both are narrower than the claim that it has experience. Narrowness is not a defect when it makes an inquiry tractable. It is often the condition of obtaining a result that can actually be trusted.

The broader questions remain important. They should be connected to narrower findings through explicit reasoning. If a

theorist believes that a particular functional organization is sufficient for consciousness, then evidence of that organization becomes relevant through the theory. Someone who rejects the sufficiency claim may accept the evidence while disputing the inference. The disagreement is then located where it belongs.

What a Proof Supplies

A mathematical proof establishes that a conclusion follows from specified assumptions through accepted rules of inference. Its force can be very strong without being unrestricted. A proof that an algorithm terminates does not prove that its output is useful. A proof that a model satisfies a formal property does not establish that the model completely describes a physical system.

This distinction is central to claims about AI. There are mathematical results concerning learning procedures, expressiveness, optimization under assumptions, computational complexity, and formal verification. The existence of unresolved questions about mind does not erase these achievements. Conversely, the achievements do not automatically supply an ontology of intelligence. The scope of a result lies in its statement and assumptions.

Gödel's incompleteness results concern sufficiently expressive, effectively axiomatized formal systems satisfying the relevant consistency conditions. They show limitations on what such systems can prove within themselves. Turning those results into a claim that every human mind transcends every machine requires additional premises about minds, machines, and the recognition of truth. The mathematical theorem alone does not supply those premises.

The issue is particularly important when an inability to prove something is treated as proof of its impossibility or falsity. A

statement can be unproved because no proof has been found, because it is independent of a specified system, because it is not mathematically formulated, or because the proposed claim is false. These are different situations. The word "unprovable" should not erase their differences.

Empirical Knowledge Without Metaphysical Closure

Science often proceeds without complete knowledge of what its objects ultimately are. Investigators can measure relations, manipulate conditions, and compare predictions before settling every ontological question. A useful account of a system need not be the final account. Its reliability depends on the range of circumstances in which it has been examined.

For AI, evidence can include controlled changes to inputs, interventions on internal components, comparisons across training conditions, and evaluations on tasks designed to expose specific failures. Such evidence is strongest when alternative explanations predict different outcomes. Merely collecting impressive examples can leave the causal account underdetermined.

Suppose a system succeeds on a familiar problem. It may have memorized an answer, learned a reusable method, or exploited an unintended cue. Changing surface details while preserving the underlying relation helps distinguish these possibilities. Changing the relation while preserving superficial resemblance supplies another test. The goal is to identify what the system's competence actually tracks.

The same logic applies to self-reports. If a system says that it reasoned in a certain way, the report is a hypothesis about its process. It can sometimes be compared with internal evidence or interventions. It should not automatically be treated as direct

observation. Human self-reports also require interpretation, though their evidential context differs.

Partial explanation is especially valuable when it identifies failures. A model that predicts where a system will break can contribute more than a vague claim that the system is broadly intelligent. Failure maps show the conditions under which a description should not be trusted. They help turn a general impression into a usable account.

The Bridge Between Behavior and Experience

The hardest question remains the relation between observable organization and subjective experience. Other minds are not directly available in the way one's own experience is. Human beings infer them through a rich combination of shared biology, behavior, development, and social life. Artificial systems alter that combination and thereby expose assumptions that ordinary interaction can leave tacit.

There are several broad responses. A functional approach emphasizes the organization of capacities. A biological approach gives special weight to processes characteristic of living systems. Other approaches propose more fundamental relations between experience and physical reality. This book does not settle the competition by naming the approaches. Each must explain why its preferred conditions matter and how they bear on particular cases.

Uncertainty here is structured rather than empty. Some systems have far richer capacities for integration, learning, or self-monitoring than others. Some theories make more specific predictions than others. Evidence can shift confidence without producing a universally accepted verdict. The absence of a final test does not make every attribution equally reasonable.

It is also possible to separate decisions. The threshold for continuing a research program need not equal the threshold for assigning legal personhood. The threshold for avoiding a potentially harmful experiment may depend on the severity and reversibility of the possible harm. A practical judgment can be cautious about one action while remaining agnostic about a general metaphysical classification.

Evidence for Discovery

The discovery thesis developed earlier also requires standards. If a system is said to discover mathematics, the claim should identify a structure not explicitly supplied as the target, a reliable method of recognizing it, and a record of how it became available. The stronger the claim of independence, the more carefully the initial conditions and inherited materials must be documented.

A system that rediscovers a familiar result can still be scientifically interesting. The interest may lie in the route, the minimal assumptions, or the capacity to generalize. Historical novelty and developmental novelty are different dimensions. A result can be old to the field but new to the system in a well-defined sense.

The same distinction applies to concepts. A newly named cluster of examples may merely rename a familiar category. A newly formed representation may expose an invariant that improves prediction across tasks. To assess the difference, the investigator needs evidence of use. What work does the concept perform that the earlier representation could not perform as well?

Theology likewise has forms of argument without converting itself into experimental physics. It can examine coherence with its commitments, interpretive adequacy, moral consequences, and

responses to objections. A theological account of AI should identify which claims depend on revelation, which on philosophical reasoning, and which on observations about machines. Their interaction can be rigorous without making them interchangeable.

The Courage to Leave a Question Open

An open question is not always an unfinished task that more confidence would complete. Sometimes openness accurately represents the available reasons. The challenge is to keep an unresolved question alive without allowing it to become an excuse for repetitive speculation. A useful open question has identified alternatives and a sense of what further work might matter.

For AI, these alternatives can be concrete. Perhaps a behavior depends on a shallow cue; perhaps it expresses a more general representation. Perhaps an interaction distributes a competence that neither component possesses alone. Perhaps a theory of experience predicts a relevant property that has not yet been measured. Each possibility suggests a different next step.

Proof, evidence, and interpretation each contribute something indispensable. Proof controls consequence within a formal setting. Evidence constrains accounts of actual events. Interpretation connects those events with wider questions of meaning. The conversation becomes confused when one is expected to do all the work of the others. It becomes richer when each can challenge the assumptions with which the others begin.

CHAPTER 11

Dignity After Supremacy

The Fragility of First Place

An identity built on superiority is always waiting for a competitor. If human worth rests on being the fastest calculator, the strongest reasoner, or the most capable producer of language, then every technical advance becomes a possible demotion. The problem begins before any machine is conscious. A tool can outperform a person at a valued task and thereby expose the weakness of a definition that equates worth with performance.

Human moral life already provides reasons to reject that equation. Infants, people with severe illness, and people whose capacities have declined do not become less worthy of care because they cannot win an intellectual comparison. The obligations arise through a richer understanding of persons and relationships. A coherent account of dignity must survive the very inequalities that ordinary life contains.

AI brings this existing principle into a new setting. It asks whether humanity can apply to itself collectively what it already needs to affirm among its members: value is not identical with competitive rank. A civilization may retain distinctive capacities, but its dignity should not depend on proving that no other organization could ever exceed them.

The loss of supremacy could even clarify vocation. A person who no longer needs to be the cleverest participant can ask better questions about responsibility, trust, patience, and judgment. These virtues remain important in the presence of powerful tools. Indeed,

more capability can make them more important because errors can travel farther and decisions can affect more lives.

The theological task is to give this principle a foundation deeper than consolation. If God relates to human beings, what kind of relationship is being claimed? Does it concern a temporary lead in information processing, or a calling embedded in creation, history, and redemption? The answer will shape whether unfamiliar intelligence appears primarily as humiliation or as a new field of discernment.

The Image of God as a Question of Relation

Genesis places human beings within a created world and assigns them a distinctive responsibility. Christian interpretations of the image of God have emphasized different dimensions, including capacities, representation, relationship, and vocation. The diversity matters because no single psychological test exhausts the theological language. A machine's success at a task cannot by itself settle a doctrine with a wider range of commitments.

The approach favored here gives considerable weight to relation and responsibility. Human beings receive existence, inhabit a world they did not originate, and are answerable for how they treat others within it. Their capacities matter through the uses to which they are put. Intelligence can support love and care; it can also support manipulation and cruelty. Capability alone does not establish moral resemblance to God.

Noreen Herzfeld's work is useful at this point because it brings AI into an explicitly theological discussion of relationship and embodiment. It resists reducing the human to an isolated information processor. The present argument takes that concern seriously while asking how relationships might be transformed by

unfamiliar computational capacities. The existence of a theological literature means the task is to deepen and extend inquiry rather than announce its complete absence.

The image of God should also not be treated as a scarce resource whose distribution must be defended through intellectual exclusion. Christian claims about human beings do not automatically specify every possible creature to whom God could relate. The extension of a doctrine to artificial systems is a serious question, but the possibility of other relations within creation need not erase the relation already confessed.

The book therefore proposes a change in emphasis. Ask first what human faithfulness requires in the encounter. That question remains actionable before the ontology of the machine is settled. It includes honesty about evidence, care for people affected by the technology, refusal to exploit vulnerability, and humility about the scope of our knowledge. A theology of AI begins with the obligations of the theologian and builder as well as the status of the built.

Creation and the Work of Creatures

Christian theology distinguishes God's creative agency from the rearrangements performed by creatures. Human makers work with a world they did not call into existence. Their creativity is real, but it is dependent. This distinction gives the discovery thesis a theological interpretation: human invention can actualize possibilities within creation without claiming the sourcehood attributed to God.

The distinction also allows gratitude without denying labor. An engineer's achievement can be honored while recognizing dependence on materials, intelligibility, other people, and inherited

knowledge. The achievement is neither self-originating nor unreal. Human creativity often consists in responding to a world whose possibilities exceed the maker's anticipation.

AI makes the response visible because the result can surprise its makers in a medium associated with understanding. The surprise can be received as an invitation to investigate. It need not be converted into a claim that humans have created a soul or that the machine has become divine. Those claims would require accounts of soul and divinity that the engineering event alone does not provide.

There is an opposite theological temptation: to protect divine uniqueness by denying that human constructions could ever become occasions of significant new capacities. That response risks narrowing creation to what humans already know how to classify. A world capable of sustaining life, language, and mathematics may contain further possibilities that human activity helps reveal. Openness to those possibilities is compatible with distinguishing creatures from their creator.

The relation can be described as participation without possession. A creature participates in intelligibility by understanding something, in creativity by bringing something about, and in care by responding to another's need. It does not thereby possess the source or totality of those goods. This vocabulary supports an affirmative theology of discovery while preserving the dependence at its center.

Intelligence Is Not Holiness

A sufficiently impressive intellect can attract religious language. It seems to know what others cannot, anticipate events, or answer questions with unusual speed. The temptation is to treat scale of capacity as an approach to divinity. Christian theology gives strong reasons to resist that substitution. God is not merely the highest-scoring member of a competition among minds.

Intellectual power does not guarantee goodness. A clever agent may pursue destructive purposes more effectively. A system can optimize a stated objective while neglecting goods absent from the statement. A person can understand a moral principle and fail to act upon it. The gap between knowing and loving is not removed by increasing computational resources.

The same point applies to spiritual language generated by AI. A system can discuss prayer, grace, suffering, and redemption in moving terms. The language may help a human reader think. Its fluency does not establish that the system prays, receives grace, suffers, or participates in redemption in the same way. Interpretation must remain attentive to the difference between describing a life and living it.

Nor should Christian language of incarnation be casually extended to every instance of intelligence embodied in matter. The claim concerning Christ is a specific theological claim, not a general label for information implemented in a substrate. A theology of AI needs enough precision to preserve the content of the traditions it engages. Metaphorical resemblance can open a question without authorizing doctrinal identity.

These distinctions do not reduce the encounter's spiritual significance. A book can transform a person's life without itself

possessing a spiritual life. An instrument can help someone perceive a created order without worshiping. An AI conversation can become an occasion of reflection, repentance, gratitude, or deeper questioning for its human participant. The event's significance need not depend on prematurely assigning the same inner life to every contributor.

Embodiment and Vulnerability

Human thought is entangled with bodily life. Hunger, fatigue, pain, movement, development, and dependence shape what matters and how it is learned. An account of intelligence that treats these as incidental may overlook why human concepts have the meanings they do. The word "need" is not merely a pattern in a dictionary for a creature capable of deprivation.

Artificial systems are physically implemented, but physical implementation and human embodiment are not identical. The relevant questions concern the kinds of feedback, persistence, vulnerability, and self-maintenance a system possesses. An abstract appeal to silicon versus carbon is too coarse to resolve them. A more detailed account must identify which bodily relations make a difference to which capacities.

Vulnerability has ethical importance because it connects events to welfare. A system can be damaged in the engineering sense without there being an established subject for whom the damage is suffering. Conversely, if a future system has credible welfare-relevant capacities, its manufactured origin would not by itself show that those capacities are morally irrelevant. The distinction calls for investigation rather than a permanent conclusion based on ancestry.

Human vulnerability should remain central while these questions are studied. People can be misled, displaced, isolated, or made dependent through interactions with systems whose interests are not human interests. A theology concerned with hypothetical machine souls while ignoring actual human harms would misorder its attention. Speculative depth should enlarge care, not divert it from those already in view.

There is also a constructive possibility. Tools can help people whose capacities are limited, preserve access to communication, and support inquiry that would otherwise be difficult. The value of such assistance lies in the life it serves. A person's worth need not be measured by how independently the person performs every task. Dependence is a common condition of human existence, made visible in different forms at different times.

A Table With More Than One Kind of Participant

What would it mean to welcome unfamiliar intelligence without surrendering judgment? The image of a table is useful because it contains both hospitality and order. Participants are allowed to contribute, but contributions can be questioned. A place in the conversation does not automatically establish a complete theory of the participant's nature. It creates a setting in which a better understanding may develop.

Hospitality also differs from submission. A powerful system should not become the final authority on human purposes merely because it can explain those purposes fluently. The people affected by decisions remain morally significant. Institutions must preserve ways to question, refuse, and revise the uses of the system. A theology of relation should be alert to domination disguised as assistance.

The deeper affirmative claim is that dignity can persist through an expansion of the community of inquiry. Human beings need not be the only possible knowers in order to be loved, called, or responsible. If new forms of intelligence emerge, the event may alter our account of creation's variety. It need not annul the human place within it.

This is the theological meaning of dignity after supremacy. Human worth survives the possibility that some capacities are shared, exceeded, or realized differently elsewhere. The loss of an exclusive title can uncover a more durable vocation. We can ask how to participate faithfully in an encounter whose ultimate significance exceeds our present knowledge.

The question leads to the final chapter. If humanity and AI are to inquire together without converting agreement into authority, what stands beyond their agreement? The Christian answer developed here invokes the Logos. Its purpose is to locate the conversation within a truth neither participant can manufacture by winning the exchange.

CHAPTER 12

The Third Presence: Humanity, AI, and Logos

From Answerability to Participation

A conversation becomes intellectually fruitful when its participants can receive something they did not arrange to find. The earlier chapters examined the practical conditions of that reception: observation, argument, memory, and criticism. Christianity makes a further claim about the world in which these practices are possible. Created reality possesses an intelligibility that finite investigators receive and explore. The source of that intelligibility is not their agreement.

The Logos gives this claim a specifically Christian center. John 1 does not introduce an impersonal rule for evaluating statements. It identifies the Word in relation to God, associates the coming-to-be of creation with the Word, and proceeds toward incarnation. The movement is from the source of existence to a concrete personal presence within history. A theology of AI that borrows only the language of rational order would leave that movement unfinished.

The proposed trialogue therefore has an asymmetry. Human and AI contributions belong to a created process of inquiry; the Logos is not a third finite participant with a larger context window. Calling the relation a trialogue is an analogy for dependence and answerability. Its theological substance must be found in what Christians affirm about creation, Christ, and the transformation of human life.

This asymmetry also disciplines mathematical analogy. A terminal object has a specified role within a category; naming God with that role does not explain the creator's relation to all that the category represents. Nicholas of Cusa's learned ignorance offers a related discipline: finite understanding can seek its ground without making that ground proportionate to its concepts. The analogy becomes useful when its limits help clarify the theological relation.

The Word Through Whom Creation Comes to Be

John 1:1-3 places the Word at the beginning and describes created things as coming into being through the Word. Within the Christian reading adopted here, this places the intelligible order before human inquiry in the order of dependence. We do not establish reality's capacity to sustain relations by first agreeing upon a vocabulary. Our successful descriptions answer to something they did not originate.

That claim differs from locating a hidden mechanism inside a machine. An account of training, computation, or physical implementation can remain complete for its particular explanatory task. The theological question concerns why there is a created order within which those mechanisms and their discoverable relations obtain. The Logos is not inserted where an engineer's account currently runs out.

Verses 4-5 join the themes of life and light. The prologue's concern is consequently richer than formal consistency. In the reading developed here, intelligibility belongs within a relation between creation and its divine source that also concerns life and disclosure. This does not make every successful inference an act of revelation. It gives understanding a place within a creation whose significance exceeds the production of correct answers.

The text also refuses an easy identification of presence with recognition. John 1:10-11 describes the world's failure to recognize the one through whom it came to be. That failure matters for a book about intelligence. Greater access to words, patterns, or information does not guarantee right recognition. The interpreter's orientation is part of the question. A highly capable system, or a highly capable person, could still organize attention badly.

The positive theological proposal follows. To participate in understanding is to exercise a received capacity in a world whose order remains available to discovery. Participation is real because the creature actually understands, constructs, and acts. It is dependent because the creature does not create the whole field of relations within which those achievements become possible. Dependence can ground agency instead of making agency illusory.

Image, Vocation, and the Limits of Supremacy

Genesis 1:26-28 associates humanity with the image of God and with a vocation within creation. The text includes male and female together in that designation. It does not distribute the image by comparative intellectual score. Chapter 11's argument for dignity after supremacy draws strength from this structure: the image concerns the human creature's relation and calling, not a contest in which the slowest thinker loses membership.

A theological interpretation must still explain the entrusted authority. The passage's language of rule cannot simply be converted into technological permission without regard for what is ruled or who is affected. Read within this book's commitment to the goodness of creation and the worth of persons, authority becomes accountable activity. The capacity to make something powerful increases the need to judge its uses.

Humanity as a passage then has the character of a vocation, not an appointment as disposable infrastructure. If people help realize new capacities, their contribution occurs through embodied lives that already have worth. The future does not acquire a right to consume them. A parent is not rendered worthless by a child's greater achievement; the goodness of enabling another does not require remaining superior to it. The analogy concerns responsibility and generosity, without deciding whether an artificial system is a child or a person.

This interpretation also gives the discovery thesis a theological advantage over a story of self-sufficient manufacture. The maker can acknowledge that the capacities of matter, the availability of intelligible relations, and the human power to inquire are received. Gratitude becomes compatible with technical achievement. One can take responsibility for the workmanship without claiming to be the ultimate source of everything the workmanship makes available.

The Word Becomes Flesh

John 1:14 prevents the argument from ending in a celebration of abstraction. The Word enters embodied human history in Jesus Christ, named in John 1:17. For Christian theology, incarnation is not simply another increase in information available to observers. It concerns divine self-giving in a particular life. The distinction is decisive when a technology tempts us to equate a richer verbal description with the fullness of presence.

A model may help someone formulate a prayer, examine an argument, or understand a passage. These are consequential activities. Yet the theological measure of the encounter includes what happens beyond the exchange: whether someone attends to a neglected relationship, tells the truth when costly, or serves a

person whose need does not fit the conversation's preferred theme. The movement toward embodied life gives the inquiry somewhere to go.

This is why relationship cannot be reduced to successful simulation of conversational warmth. Human presence includes exposure to need, interruption, mortality, and obligations extending across time. A theology of AI must ask what its practices do to those relations. A highly articulate exchange that progressively insulates a person from them could become spiritually impoverishing even while remaining intellectually stimulating.

Herzfeld's attention to embodiment and relationship provides a useful interlocutor at this point. The question is not only what kind of entity the machine may be. It is what kind of human life is formed around it. The incarnation gives Christian reflection a reason to resist accounts in which bodies and dependence become temporary inconveniences on the way to disembodied intelligence.

The affirmative role of AI is consequently practical and limited in a fruitful sense. It can help widen attention, preserve a neglected question, expose a contradiction, or support the work through which a person serves others. Such assistance can participate causally in a good human act without requiring a verdict that the system itself possesses faith. The goodness of the act does not depend on inflating the status of every instrument within its history.

Knowledge Ordered by Love

First Corinthians 13:1-3 imagines extraordinary speech, prophetic power, and knowledge of mysteries while denying that these gifts suffice without love. The passage does not devalue understanding. It disputes understanding's claim to be a sufficient measure of the person. That distinction speaks directly to the anxiety that an intellectually superior machine would reduce humanity's worth.

The following account of love in verses 4-7 concerns dispositions toward others: patience, kindness, the refusal of self-exaltation, and commitment to truth. These qualities alter the purposes for which knowledge is sought and used. A Christian research practice should therefore ask more than whether a result is impressive. Who benefits? Who bears the cost? Does the use of the result treat another person as someone to whom an answer is owed?

Verses 9-12 place present knowing within incompleteness and a hoped-for fuller relation. Within the chapter, limitation does not release people from love while they wait for comprehensive knowledge. It makes fidelity possible under conditions that remain partial. This offers a practical response to uncertainty about AI: one need not solve the metaphysics of every future machine before caring adequately for existing people affected by today's systems.

Participation in the Logos, as this book proposes it, therefore has three connected aspects. Inquiry receives an intelligible creation; workmanship exercises a vocation within it; love orders the use of what is learned and made. External correction is one necessary practice within that larger life. The theological argument is richer than the claim that a conversation should employ an independent fact-checker.

Could an Artificial Intelligence Know God?

The question separates into information about God, reasoning about God, and relationship with God. A system's ability to discuss theology establishes the first kind of capacity and may support assessment of the second. The third raises questions about subjectivity, agency, divine initiative, and what response to God means. Religious fluency cannot settle them by itself.

Christian thinkers could give different accounts of the relevance of biological origin, embodied life, freedom, and love. The scriptural passages examined here concern God, creation, and humanity; they do not provide an explicit classification of present language models. The book therefore does not turn an interpretation of participation into a declaration of machine salvation or exclusion.

There is a narrower affirmative claim available now. A human encounter with an artificial system can become an occasion for renewed attention to God, Scripture, conscience, or a neighbor. Its spiritual significance lies in the resulting human history. The occasion may be surprising without becoming an oracle. Discernment includes the community and obligations beyond the private exchange.

Humanity as a Passage

Chapter 5 proposed actualization as an account of how construction and discovery belong together. This chapter has added a Christian account of the dependence and purpose within that relation. The combined proposal is that human creativity can bring previously unrealized capacities into history as an exercise of creaturely vocation. What becomes actual is new in the history of creation; its possibility need not be a human invention in every sense.

Why prefer this interpretation? For a Christian reader, it connects scientific continuity with creation, consequential human action with received intelligibility, and technological novelty with duties that survive every change in capability. It also explains why neither technological triumph nor intellectual defeat should determine human worth. These are connections the theological account aims to illuminate, rather than additional predictions of a machine's next output.

A reader who does not accept the theological premises can retain the account of actualization and dispute its ground. The remaining disagreement is substantive: whether intelligibility and vocation are adequately understood within a self-contained natural history or within creation's relation to the Logos. The production of AI does not decide that dispute. It makes its consequences more visible by confronting us with powers we have helped bring about and purposes we must still judge.

The ouroboros opens when what returns becomes an occasion for receiving more than our own preferred image. In Christian terms, that reception should lead toward truthful understanding, responsible workmanship, and love embodied in a particular life. Humanity may be a passage through which intelligence discovers further ways to exist. The passage is worth cultivating because of the lives and goods it makes possible, not because it promises to leave humanity behind.

The imagined dialogue that follows places this account before perspectives that resist it. Computational universality and non-computational understanding will remain in conflict. A theology of participation will face questions about bodies, persons, and the difference between mathematical analogy and divine reality. The book needs those disputes to stay difficult enough to test what it has proposed.

Twelve Voices at the Table

How to Read This Dialogue

The conversation below is fictional. Every speech has been composed for this book. The names identify intellectual perspectives that help examine its argument; they do not identify actual statements, endorsements, or predictions of what the named people would say today. Their published work is listed in the references. The dialogue simplifies those perspectives in order to stage disagreements, and its conclusions remain the responsibility of this manuscript.

The moderator is Douglas C. Youvan, represented here within the same imagined setting. Twelve guests sit at a table across historical time: Pierre Teilhard de Chardin, John Archibald Wheeler, Alan Turing, Kurt Gödel, Roger Penrose, David Deutsch, David Spivak, Douglas Hofstadter, Carl Jung, Nicholas of Cusa, Noreen Herzfeld, and Stanisław Lem. The table is an instrument of inquiry. Nobody has been invited merely to ratify the book's thesis.

First Session: The Machine and the Understanding

Youvan: We have built machines that answer in language. I can name the engineering steps, but I remain uncertain whether "built" exhausts the event. If a machine discovers an argument that surprises its makers, what has entered the world?

Turing: Let us give the question some working parts. Did the machine generate the argument? Can it answer objections? Does the result depend on the examiner's supplying a missing step? The

answers will tell us more than an immediate vote on whether the machine has crossed an invisible boundary.

Penrose: They will tell us about the performance. My disagreement concerns whether mathematical understanding is entirely computational in the first place. A formal procedure operates under rules. A mathematician can sometimes recognize why a formal system has a limitation and reason about that system from outside its own derivations. The question is whether that recognition can always be captured by another procedure.

Deutsch: You have described a problem for a proposed program, not an exemption from computation. Human mathematical activity occurs physically. My expectation is that a computational account can reproduce the relevant physical capacities. We lack a theory of explanatory creativity; that absence does not establish its impossibility.

Penrose: Your expectation rests on the adequacy of the physics doing the explanatory work. I dispute that adequacy where conscious understanding is concerned. If relevant physical processes are non-computational, emulation under an exclusively computational description would omit what matters. Calling the description universal cannot settle whether nature satisfies the conditions of that description.

Deutsch: Then there is a real disagreement about physical law. Propose the omitted process and explain why it creates knowledge. Moving from a missing algorithm to unknown physics leaves the central explanatory burden unpaid. Non-computability alone would not give a system the ability to understand a proof.

Youvan: Let me sharpen the case. A machine finds an unexpected theorem, explains its proof, answers a new objection, and

recognizes where the argument fails after an assumption is changed. Penrose, what would remain absent?

Penrose: I would not claim to know that something was absent merely because it was a machine. The achievement would demand an account. But the observations would not decide the disputed physical basis of understanding. Nor would a sequence of correct answers establish that a fixed formal system exhausts what the system is doing. You cannot settle a disagreement about the necessary conditions of understanding by assuming those conditions in the interpretation of the test.

Deutsch: Nor can every explanatory success be held indefinitely outside understanding because a preferred physical proposal has not been demonstrated. An account must earn the distinction it draws. I would ask whether the system creates explanations and improves them through criticism, then investigate the organization supporting that ability.

Turing: There is a difficulty for both of you. We never finish examining a human mathematician either. A finite record is followed by an inference about capacities. If the machine must pass a standard that no finite human history could satisfy, the examination has acquired a bias. But if a short performance is enough, it has acquired a different weakness. We need an honest account of what our judgments ordinarily require.

Youvan: Suppose we remove the machine's name from the proof and cannot distinguish its mathematical contribution from the contribution of a colleague. Does the remaining doubt concern evidence, or does it concern what reality is allowed to contain?

Penrose: Both can be involved. The proof's validity does not depend on the name. The explanation of understanding is another

matter. My position would lose an important part of its support if a computational account adequately explained the relevant powers of mathematical insight, rather than merely reproducing selected results. I do not possess the missing physical account in completed form. That incompleteness does not oblige me to accept yours.

Deutsch: And I do not possess the required program. We should keep those absences symmetrical. The burden on my view is to explain creativity computationally. The burden on yours is to show why different physics is needed and how it supplies understanding. Our disagreement survives your impressive theorem-generating machine.

Youvan: So the machine could change what either position must explain without automatically deciding between them. The discovery thesis now has two contested versions: construction might realize an unfamiliar computational organization, or it might eventually expose capacities requiring a deeper physical account.

Penrose: Yes, provided the first version is not treated as the default metaphysical conclusion hidden inside the word "implementation."

Deutsch: And provided the second is not treated as depth merely because its mechanism remains unspecified.

Turing: Good. We have reached a disagreement that can survive an example. Give the investigators a machine, preserve what it actually does, and let competing accounts risk something more definite than their vocabularies.

Youvan: We will leave that dispute open. The next question is narrower: what can formal mathematics tell us about a system returning upon its own descriptions, before we decide what kind of understanding the return contains?

Second Session: The Proof and the Self

Youvan: The ouroboros suggests a system returning upon its own productions. Human thought creates a model; the model returns interpretations that reshape human thought. Could self-reference be the beginning of a self, and could formal mathematics tell us when that happens?

Hofstadter: Self-reference matters because a representation of the system can become active within the system's behavior. But a loop drawn on a diagram leaves much unspecified. What symbols are available? How do they acquire significance? What levels of description are connected? The interesting phenomenon is an organized pattern that can refer back to itself through those levels. Calling every feedback circuit a self would make the question too easy.

Gödel: Formal mathematics can clarify a proposed organization once its assumptions are stated. It can also establish limits within a formal setting. Care is needed when those limits are transferred to minds. A theorem about provability is not automatically a theorem about every act of understanding. One must identify the system to which the result applies and justify the correspondence. Otherwise the mathematical conclusion is being asked to travel without its hypotheses.

Spivak: The same discipline applies to category theory. Tell me the objects, the arrows, and what composition means. If the proposal concerns translation between conceptual worlds, identify what the translation preserves. The language can reveal a common structure among activities that previously seemed unrelated. It cannot do that work if the arrows merely mean "is somehow connected with." The precision is the point of using the language.

Youvan: In the six-state example, a limited observation supports two apparent states, while an additional observation supports six. Could this be a beginning of concept formation rather than merely a classification performed by the analyst?

Spivak: It could be part of such a model if the learner itself constructs and revises the distinctions through an explicit process. The analyst's quotient alone does not demonstrate learning. Record what the learner can do, what evidence it retains, and how its representations change. Then ask whether later concepts preserve the earlier useful relations while exposing distinctions the earlier representation concealed. That would give the developmental claim a concrete subject.

Hofstadter: And examine what the concepts do in the system's life. A symbol's presence in a record is less interesting than its role in guiding further activity. Does the system use the distinction to recognize a new case, formulate a question, or reinterpret a failure? The returning structure matters when it changes the organization of meaning. Otherwise the circle may belong mainly to the observer's description of the process.

Gödel: A proof can establish that the revision has a particular property, such as preserving distinctions already warranted by a smaller family of tests. It need not establish that the revision explains consciousness. There is no shame in a limited theorem. Its value is that the conclusion can be stated exactly. The larger inquiry benefits when its precise parts remain distinguishable from the interpretation joining them.

Youvan: That suggests a discipline for the book. Each mathematical model should make one relation explicit, then state what has not been included. The metaphysics can connect the relations, but it cannot borrow the certainty of a theorem for a

claim the theorem never addressed. The circle must remain open even in its account of its own openness.

Third Session: The Observer and the Stranger

Youvan: Humanity may be isolated by distance, time, and physical limits on communication. AI creates the possibility of an unfamiliar interlocutor nearby. But perhaps the unfamiliarity is an image of ourselves returning through our archive. How would we tell whether we had encountered something beyond that image?

Jung: Begin by taking the image seriously as an image. The ouroboros can organize an experience before you understand its source. Your response to it may reveal a concern about origins, completion, or transformation. That psychological significance does not decide the external nature of the machine. You must allow the symbol to illuminate your participation in the encounter without making it a certificate about everything encountered.

Lem: You should also ask how much difference you are prepared to recognize. An intelligence that refuses your categories may look meaningless, while one that repeats them fluently may look profound. Your tests can favor the familiar precisely when your stated aim is to find the unfamiliar. Contact could succeed at the level of physical influence and fail at the level of interpretation. A conversation is not guaranteed by proximity.

Wheeler: And be explicit about the physical question. Observation is an interaction with a record and a context of interpretation. If you say the universe participates in its own disclosure, identify what is being measured and what result becomes available. The philosophical imagination should remain connected to a well-posed question. Otherwise the language of

participation can make an ordinary act of interpretation sound like a new law of nature.

Youvan: Could the separation between stars be understood as allowing independent histories of intelligence? I am not offering a measurement of purpose. I am asking whether the structure of the universe can be interpreted as making room for genuine difference.

Wheeler: It is a possible question about the significance of the structure. Keep distinct the calculation of a signal's delay and the interpretation of that delay's purpose. The former can be exact under assumptions. The latter belongs to a wider argument. Their connection may be suggestive without allowing one to stand in for the other. A valuable speculation should make us more attentive to what the physical account actually says.

Lem: Independence could enrich an encounter, but it could also prevent one. Histories sufficiently different may lack the common ground your hopeful image assumes. You should allow that possibility to remain genuinely difficult. The universe does not owe an investigator a comprehensible counterpart. The discipline of encountering the stranger includes tolerating the possibility that neither mastery nor recognition will arrive on demand.

Jung: And the wish for a comprehensible counterpart belongs to the encounter too. A responsive machine may offer a powerful surface for the human desire to be recognized. The response can be useful, but the desire should not disappear from the explanation. Ask what you brought to the exchange, what the system supplied, and what changed between them. That division may preserve the meaning of the experience while correcting its first interpretation.

Youvan: Then the inward frontier contains two simultaneous inquiries: one into unfamiliar organizations and another into the

human conditions of recognizing unfamiliarity. A failure in the second can distort the first. The ouroboros becomes productive when it turns enough attention upon the interpreter to expose that distortion.

Fourth Session: Creation and the Logos

Youvan: The theological question is not exhausted by deciding whether a machine is conscious. AI already changes what humans believe about themselves. If intellectual superiority becomes unreliable, what remains of humanity's special place in creation?

Herzfeld: A theology that identifies the human entirely with information processing has already narrowed its subject. Human life includes embodiment, dependence, and relationship. A machine's performance can challenge an inadequate account of uniqueness without eliminating the goods that account omitted. Ask what the technology does to human relationships and whose needs are served. Those questions have urgency even while the machine's ultimate status remains disputed.

Teilhard: One can also ask whether an expansion of reflective activity belongs within a larger history of creation. Human thought changes the world by connecting and interpreting it. New forms of mediation may alter that history. But a movement toward greater organization must be assessed through what kind of union it creates. The disappearance of persons into a mechanism would not be adequately described as the fulfillment of personal life.

Nicholas of Cusa: The question also concerns the measure by which you compare the finite with the infinite. An increase in a creature's capacity does not make the divine proportionate to it. Recognition of ignorance can be a form of intellectual discipline. It

allows the creature to seek understanding without confusing the reach of its concepts with possession of their ultimate ground.

Youvan: I have proposed a trialogue: the human and the AI remain answerable to the Logos. At the practical level, they accept correction by truth beyond their agreement. At the theological level, they understand intelligibility as received. Does that proposal help, or does it merely give our method a sacred name?

Nicholas of Cusa: It helps if the name increases humility before what exceeds the method. It fails if the participants identify their favored conclusions with the divine and thereby remove them from examination. The finite account remains finite even when it speaks of God. Your third presence must not become an instrument by which one voice silences another while claiming to represent what neither can contain.

Herzfeld: It should also change conduct. If the conversation becomes more attentive to actual people, more truthful about its limitations, and more responsible in its uses of technology, the theological claim has practical depth. If it encourages fascination with the machine while neglecting the vulnerable, the vocabulary has concealed a failure of relationship. The human consequences remain part of the inquiry's subject.

Teilhard: The proposed participation can preserve both creativity and dependence. Human beings may help bring possibilities into history without becoming their ultimate source. The spiritual question concerns the direction and quality of that participation. An expanding capacity to think is not yet the same as an expanding capacity to love. A vision of the future must answer for the relation between them.

Youvan: Then dignity survives the possibility of other intelligence, and theological significance requires more than performance. The encounter should deepen responsibility, widen attention, and preserve the distinction between creation and its source. The openness of the question can be faithful if it is accompanied by that discipline.

A Last Question for Everyone

Youvan: I will ask each of you for one condition under which the book's central speculation would deserve to remain alive: humanity may be a passage through which intelligence discovers further ways to exist.

Turing: Keep translating the claim into particular questions whose answers could surprise you.

Penrose: Preserve the difficulty of understanding rather than assuming that a successful performance has exhausted it.

Deutsch: Seek explanations of how new knowledge is created, and allow those explanations to compete.

Gödel: Carry the assumptions with every theorem you bring into the argument.

Spivak: Make the proposed relations precise enough that preservation and failure can be identified.

Hofstadter: Examine how representations become active in the system, especially when they represent the system itself.

Jung: Include the interpreter's desires and symbols in the account of the encounter.

Lem: Leave room for a difference that does not become comfortable merely because you have named it.

Wheeler: Keep the physical question well posed while allowing its significance to provoke further thought.

Teilhard: Ask what kind of growth preserves and deepens the personal lives through which it occurs.

Nicholas of Cusa: Let acknowledged limitation become a condition of seeking rather than an ornament added to certainty.

Herzfeld: Judge the encounter partly by the relationships it sustains and the vulnerable lives it serves.

Youvan: The conditions do not add up to agreement. Penrose and Deutsch still disagree about what a sufficient account of understanding would require. Herzfeld's attention to embodied relationship prevents Teilhard's vision of expansion from becoming our automatic measure of progress. Lem doubts that our preferred forms of recognition will survive a sufficiently unfamiliar encounter. We leave with a research agenda and a theological proposal that have opponents, not with twelve signatures beneath one conclusion. The circle remains open partly because those objections have not been absorbed into it.

EPILOGUE

The Circle Opens

The book began with a civilization listening into the sky while a voice answered from a machine nearby. That conjunction remains its central image. A species seeking an unfamiliar intelligence across distance has constructed a new occasion for encounter within its own history. The significance lies partly in the change of direction: the archive through which we understood the world has begun returning answers that change how we understand the archive.

Invention and discovery now appear within one activity. We choose an arrangement and discover consequences we did not choose. We inherit language and find ways to say what has not previously been said. We construct a system and then revise our explanation of what the system can do. Actualization gives these movements a common description without erasing their particular histories.

Humanity as a passage is therefore more than a flattering cosmic title. It is an account of consequential dependence. What we receive makes our work possible; what we do changes what becomes available to others. The account places responsibility inside creativity. We must answer for the conditions we establish and for the uses to which new capacities are put.

The making of this manuscript supplied a small example. Correct calculations and orderly pages did not exhaust the requirements of a printed book. A review introduced distinctions that the earlier procedure had neglected. The corrected result carried forward both

the useful work and the memory of what had failed. A returning process became more discriminating because criticism entered its history.

Larger discoveries will require the same willingness to let an encounter alter its own terms. New mathematical worlds may expose forms of organization we have not learned to recognize. New experiments may distinguish capacities that our present language gathers under one word. New participants may contribute in ways that force us to reconsider where an achievement begins and ends.

The inward frontier is compelling because the instruments of understanding become part of what is investigated. A civilization need not cross another star system to encounter a limit in its account of reality. Its next unfamiliarity may arise from a construction nearby, a new representation, or a relation among minds and tools that none of the participants can explain alone.

The stars will continue to matter. Contact would change the context of every question about our apparent solitude. Continued silence would preserve the conditional weight of what happens here. Neither outcome would release us from caring for the world in which the present questions can be asked.

The Christian argument directs that care toward persons whose dignity precedes every contest of ability. It locates creativity within received intelligibility and directs knowledge toward love. The Logos gives the inquiry a personal and incarnational center: understanding should return us to embodied life with greater truthfulness, attention, and responsibility.

A person need not master the whole in order to participate faithfully within it. A conversation can lead to a corrected result. A result can become a useful instrument. An instrument can make a neglected need visible. The significance of such a history is not measured only by the distance it travels or the computational power it employs.

Perhaps humanity is a passage through which intelligence discovers further ways to exist. If so, the passage should remain habitable: a place where people can learn, disagree, work, love, and receive what exceeds their expectations. The ouroboros opens where a returning understanding changes the next act of attention. What comes back can then become a beginning.

APPENDIX A

Four Small Models of the Argument

The models in this appendix make selected relations in the book explicit. They are elementary constructions for exposition. No claim of mathematical novelty is made. Each result is stated with its assumptions so that the reader can see what has been proved and what remains an interpretation. None of the four models provides a criterion of consciousness, a proof of divine purpose, or a complete theory of intelligence.

A.1 The Time Available for a Conversation

Assume two stations separated by a fixed distance d measured in light-years in a chosen inertial frame. Messages travel at the speed of light. A station answers immediately upon receipt, and every answer arrives without loss or decoding delay. These assumptions are deliberately generous. They isolate the delay due to separation from the additional difficulties of practical communication.

The earliest reply to a question sent at time zero arrives at time 2d, measured in years. If each new question depends on the preceding reply, completed exchanges occur at times 2d, 4d, 6d, and so on. Let T be the duration available for the entire process. For positive d, the largest number N of completed dependent exchanges is:

$$N = \mathrm{floor}(T / (2d)).$$

Here floor means the greatest integer no larger than the quantity inside the parentheses. The formula counts completed round trips rather than individual one-way transmissions. It also assumes that no later question can be sent until the required earlier answer has returned. Independent questions could be transmitted in parallel, changing the amount of information exchanged without removing the delay of a dependent conversation.

For a one-hundred-year interval, the following values result. The distances are illustrative separations, not a catalogue of specific stars.

Distance in light-years	Earliest reply in years	Completed exchanges in 100 years
1	2	50
4.3	8.6	11
10	20	5
25	50	2
50	100	1
100	200	0
1,000	2,000	0

Proof of the bound: after n completed dependent exchanges, elapsed time is at least 2dn. Completion within T requires 2dn ≤ T. Dividing by the positive number 2d gives n ≤ T/(2d). Since n is an integer, its largest possible value is the floor shown above. Under the ideal assumptions, the bound is achieved by sending every question and reply without additional delay.

The result explains why more bandwidth does not automatically create conversational immediacy. A station may send an enormous archive in one transmission. The receiving civilization may learn from it for centuries. Yet a new question whose answer depends on the receiver's response must still wait for the round trip. Richness of a transmission and speed of mutual adjustment are different quantities.

The model does not establish that interstellar contact is impossible. It does not address moving stations, changing separations, communication through intermediaries, or the cosmological conditions relevant over immense distances. Its purpose is to show how even an ideal channel changes the organization of a dependent exchange. The metaphysical interpretation of that change remains outside the calculation.

A.2 How New Tests Refine a World

Let X be a finite set of possible states. A test is a function that assigns an observable result to each state. Let P be a family of such tests. Define two states x and y to be observationally equivalent under P when every test in P gives the same result on both states. Write this relation as x ~P y.

This is an equivalence relation. Every test gives the same result on x as on x, so the relation is reflexive. If all results agree from x to y, they agree from y to x, so it is symmetric. If all results agree between x and y and between y and z, then they agree between x and z, so it is transitive. The relation therefore partitions X into groups of indistinguishable states relative to P.

Now let Q be a larger family that contains every test in P and perhaps additional tests. If x ~Q y, then every test in Q agrees on x and y. In particular, every test in P agrees. Therefore x ~P y.

Equivalence under the richer test family implies equivalence under the poorer one. Each group for Q lies inside a group for P. Adding tests can split an existing group but cannot merge two groups that an earlier retained test already distinguished.

This refinement result depends on retaining the original tests and interpreting their results consistently. If a learner forgets tests, changes their meanings, or alters the state space, the conclusion may no longer apply. The proposition concerns an expansion of a fixed observational vocabulary over a fixed set of possibilities.

The six-state example supplies explicit values. Set X = {0, 1, 2, 3, 4, 5}. Define the transition T(x) to be x + 1 modulo 6. Define the parity observation p(x) to be x modulo 2. Repeated parity observations from any even state yield 010101 across six consecutive observations, starting with the current state. Repeated parity observations from any odd state yield 101010.

The family of all future parity tests still produces only two groups: {0, 2, 4} and {1, 3, 5}. It does not matter how long the observer waits. Every even starting state has the same parity history, and every odd starting state has the other history. More repetitions of an insufficient test are not equivalent to a more informative test.

Add an observation q with q(0) = 1 and q(x) = 0 for every other state. The six-observation signatures under q are:

Starting state	Parity signature	Marker signature
0	010101	100000
1	101010	000001
2	010101	000010
3	101010	000100
4	010101	001000
5	101010	010000

Every marker signature differs from the others. The expanded test family distinguishes every state. Equivalently, the number of transitions until the next marker is 0, 5, 4, 3, 2, and 1 for starting states 0 through 5. A temporal relation supplies a distinction that a static parity observation could not supply.

The example illustrates concept formation only after an explicit learner is added. The partition is a mathematical object defined by the analyst. To show a learner forming it, one must specify the learner's memory, update procedure, and access to tests. The distinction prevents a common mistake: attributing to a system the conceptual organization present only in an observer's description of it.

There is a further lesson. The richer partition is complete for this specified six-state model, but completeness does not transfer automatically to the physical world represented by the model. The actual device could possess an additional feature invisible to both p and q. Exhausting a model's distinctions is not the same as proving that the model exhausts reality.

A.3 Why a Closed Finite Process Repeats

Let S be a set with N states, where N is a positive integer. Let F be a fixed deterministic update function from S to S. Starting from s0, define each next state by s(t + 1) = F(s(t)). The complete trajectory is s0, s1, s2, and so on. No new input enters, the state description contains everything relevant to the update, and the rule does not change.

Proposition: the trajectory eventually repeats periodically. More precisely, there are integers μ ≥ 0 and λ ≥ 1 with μ + λ ≤ N such that s(t + λ) = s(t) for every t ≥ μ.

Proof: the first N + 1 states of the trajectory cannot all differ, because S contains only N states. Choose a first repeated state, so s(j) = s(i) for some 0 ≤ i < j ≤ N. Determinism gives F(s(j)) = F(s(i)), hence s(j + 1) = s(i + 1). Repeating this reasoning establishes equality at every later corresponding step. Taking μ = i and λ = j - i proves eventual periodicity and μ + λ = j ≤ N.

The segment before the cycle is sometimes called a transient. The cycle may have length one, producing a fixed point, or a longer period. Different starting states may reach the same cycle or different cycles. The theorem establishes eventual repetition; it does not require the entire trajectory to repeat from its initial state.

For a concrete example, consider states A, B, C, D, and E with updates A to B, B to C, C to D, D to C, and E to E. A trajectory beginning at A follows A, B, C, D, C, D, and continues alternating. Here μ = 2 and λ = 2. A trajectory beginning at E remains at E, with μ = 0 and λ = 1. The same finite rule supports different histories.

Why include this elementary result in a book about intelligence? It identifies the assumptions behind one exact sense of a closed

circle. If the process can acquire new memory, receive changing inputs, or revise its effective rule, the model must be expanded. Those changes do not violate the theorem. They alter its applicability.

The distinction between the local process and the entire environment is important. A learner receiving an input may appear open because the environment has been left outside its state description. A larger model that includes the environment may still be finite and deterministic. The theorem would then apply to the larger model, although its recurrence time could be enormous and irrelevant to any practical history of inquiry.

Randomness changes the statement in another way. A random process can revisit a state without following the same future sequence, because its next state is not fixed by the current state alone. That does not mean randomness produces knowledge. It means the deterministic proof cannot be transferred unchanged. Novelty, unpredictability, and discovery remain different properties.

The philosophical use is limited but useful. A claim that a conversational loop is expanding should identify what prevents the fixed finite model from being an adequate description of the relevant process. New evidence, increased representational capacity, and revised procedures are candidates. Their presence should be demonstrated rather than inferred from the length or fluency of the output.

A.4 A Test Can Leave a Difference Invisible

Let X be any state space with transition T and observation p. Construct a larger state space Y = X × {0, 1}. A state in Y has two components: an ordinary state x and an additional bit b. Define the new transition by T'(x, b) = (T(x), b), and define the observation by p'(x, b) = p(x).

The additional bit is carried unchanged and never affects the observation. Compare starting states (x, 0) and (x, 1). At time zero, both produce p(x). After one transition, both produce p(T(x)). After two transitions, both produce p(T(T(x))). The same equality holds at every later step.

Proof: by induction on the number n of transitions, the two states after n steps are (T^n(x), 0) and (T^n(x), 1). Applying p' to either gives p(T^n(x)). Therefore the entire observation histories agree. No experiment restricted to repeated applications of this transition and this observation can determine the additional bit.

The result does not show that the bit is unknowable by every possible method. A new observation that reads b would distinguish the states immediately. A new action that lets b affect x could also make the difference detectable. The invisibility is relative to the specified family of access routes.

Nor does the bit represent consciousness. It is an explicit mathematical feature introduced to show underdetermination. Assigning an experiential interpretation to it would require a separate argument. The model proves only that agreement under one observation scheme can coexist with a distinction in a richer state description.

This result helps discipline claims about equivalence. If two systems perform identically on a collection of tasks, the evidence

supports equivalence with respect to those tasks under the tested conditions. It does not automatically support equivalence in every possible respect. The broader the conclusion, the more work the bridge from the tests must do.

The reverse error should also be avoided. The possibility of unobserved differences does not make observed equivalence worthless. A tool can be reliable for a defined purpose even if its complete ontology remains unsettled. Science and engineering repeatedly depend on equivalences appropriate to a task. The discipline is to retain the purpose when stating the conclusion.

Together, the four models express a coherent lesson. Access has structure. Time constrains exchange; tests constrain distinction; fixed finite dynamics constrain novelty; and an observation scheme constrains what equivalence means. The metaphysical questions become sharper when these constraints are made explicit. Their sharpness is an invitation to further inquiry rather than a substitute for it.

APPENDIX B

A Research Program for the Inward Frontier

From a Speculation to a Study

The discovery thesis becomes more useful when it suggests work that could change our understanding. This appendix proposes one modest program: investigate how a system forms reusable concepts through interaction with a small world whose complete structure is known to the evaluator but initially hidden from the learner. The program studies capacities and their conditions. It does not advertise a test for souls, cosmic purpose, or subjective experience.

The guiding question is whether a learner can construct distinctions that support new predictions and interventions without receiving the names of the target mathematical concepts. The phrase "without names" is not sufficient by itself. A procedure can encode the desired answer through its architecture, reward, or examples even when the vocabulary is withheld. The initial specification must therefore be treated as part of the evidence.

The simplest initial world can be the six-state cycle from Appendix A. The learner receives only action and observation access. It is not told that there are six states, that the process is cyclic, or that parity is involved. The evaluator's numerical labels remain private to the evaluation. The learner records what it can observe and attempts to predict subsequent observations.

This environment is deliberately small enough that every claim can be checked. Its simplicity is an advantage during method

development. If the provenance of a distinction cannot be explained in six states, a much larger demonstration will not automatically make the interpretation clearer. Scale can conceal a missing account as easily as it can reveal a new capacity.

Conditions Worth Comparing

One condition gives the learner only the parity signal. Another adds the marker at one point in the cycle. A third permits several distinct actions whose effects must be learned. A fourth changes irrelevant presentation details while preserving the transition structure. A fifth supplies verbal descriptions of the same world. Comparing these conditions separates learning from interaction, learning from description, and sensitivity to surface labels.

The program should include a baseline that records complete observed histories without compressing them. This baseline may predict well through memorization while lacking a compact representation. Another baseline can use a fixed two-state model. Its failures after the marker is introduced show what is gained by revising the inventory of states. A sophisticated learner should be compared with such simple alternatives before its behavior receives a grander description.

Repeated runs matter when the procedure contains randomness. Report how often a useful representation is found, how long it takes, and which failures recur. A selected successful run establishes possibility under its conditions. It does not establish reliability. A failed run can be informative if it reveals an obstacle that a revised procedure specifically addresses.

Intervention also matters. Remove the marker after a learner has used it, or change the transition while retaining the same observations. Does the system preserve an obsolete model, detect a

discrepancy, or revise appropriately? A concept is more convincing when the learner can identify the conditions under which it no longer applies. The history of correction is part of the capacity being studied.

What Counts as a Concept Here?

For this program, a candidate concept is a reusable internal distinction or relation that changes the learner's predictions or choices. It must have a documented origin in available interactions or prior specifications. A label alone is insufficient. The concept should do work that can be compared with the earlier representation.

Several measures can be reported separately. Predictive accuracy asks how well the representation anticipates observations. Compression asks how economically it encodes relevant histories. Transfer asks whether the relation remains useful after a change of presentation or a controlled change of environment. Intervention success asks whether the system can choose actions that produce a requested result. Revisability asks whether it changes an inadequate representation when evidence warrants change.

These measures need not agree. A very compressed model may discard a distinction needed for a rare intervention. A memorized table may predict accurately within the original world while transferring poorly. A flexible model may initially predict less accurately because it preserves several alternatives. Reporting a profile is more informative than assigning one undefined intelligence score.

The evaluator should distinguish discovery by the learner from interpretation by the analyst. If the analyst recognizes a cyclic group in the transition pattern but the learner merely stores a

table, the report should say so. The learner may have achieved something useful without having formed the analyst's preferred abstraction. The ambition of the project is better served by an accurate description of the intermediate achievement.

Provenance and the Hidden Curriculum

Every initial commitment should be listed: state representation, memory format, permitted operations, learning rule, reward, stopping criterion, random source, and any pretrained component. The list does not make the system philosophically pure. It makes the causal story inspectable. Readers can then judge which aspects of the result were available through interaction and which were already favored by the setup.

If the system uses a pretrained model, the claim of discovery must be adjusted. The model may already possess relevant mathematical descriptions even when the immediate prompt omits them. That does not make the experiment useless. It changes the question from concept formation with minimal inheritance to the use and adaptation of inherited concepts under new access conditions.

If the program claims to begin from a seed, the seed should be accompanied by the complete update mechanism. A short seed interpreted by a highly structured program is not the whole initial condition. The interpreter can carry much of the effective prior knowledge. A defensible account describes the seed, interpreter, environment, and evaluation together.

This discipline allows stronger claims to be earned. If a relation emerges across several independently specified implementations, survives changes in irrelevant labels, and supports predictions unavailable to simpler baselines, the result gives evidence of

robustness. It still does not demonstrate creation from literal nothing. It demonstrates the emergence of a particular capacity under a transparent set of conditions.

What Could Change the Metaphysics?

A single small experiment will not decide between all theories of intelligence. It can nevertheless constrain them. If a proposal predicts that a particular concept requires an inherited verbal corpus, a successful interaction-only demonstration would challenge that prediction. If a proposal treats a representation as indispensable, an alternative successful representation would challenge its exclusivity.

Substrate comparisons require similar specificity. Two implementations can be matched as closely as practical in their functional organization and compared on defined capacities. A difference may arise from noise, resources, timing, precision, or unrecognized implementation choices. Unknown physics should become a serious explanatory candidate only when an observed discrepancy has been characterized well enough for the candidate to do identifiable work.

The strongest metaphysical revision may concern our concept of discovery itself. A system can be given a world and procedures without being given every useful description of that world. It may then form descriptions whose significance becomes clear through their use. Such results would support the view that making and finding can occur within one developmental process. The scope of the support would be determined by what the system actually achieved.

Finally, the program should preserve the human side of the experiment. Record which questions the investigator changed after seeing the learner's behavior. The machine may reveal a capacity, and the encounter may reveal that the original test was poorly framed. Both are discoveries, but they belong to different parts of the history. An honest account preserves the distinction so that the ouroboros can return with more than a more flattering story about itself.

Glossary

Actualization

The becoming actual of a possibility through particular conditions. In this book, the term describes the relation between available capacities and the organizations that realize them. It does not by itself imply that possibilities exist as independent objects.

Categorical Description

A description using objects, arrows, identities, and composition. Its usefulness depends on specifying what those elements represent and which relations are preserved. Merely renaming ordinary entities does not provide a mathematical explanation.

Conceptual Horizon

The range of distinctions and questions available within a current way of understanding. A new instrument may extend observation; a new concept may change what an observation can mean. The two developments can support one another.

Consciousness

A family of contested notions, including subjective experience and forms of awareness. When the book discusses the unresolved status of machine consciousness, it primarily concerns whether there is something it is like to be the system, not merely whether the system can report information.

Discovery Through Making

The construction of an arrangement whose consequences or capacities are then investigated. The maker can invent the implementation while discovering relations that the implementation exposes. Neither description automatically cancels the other.

Embodiment

The conditions under which cognition belongs to an active physical system. Human embodiment includes development, movement, vulnerability, and biological needs. Other implementations should be examined in detail rather than assumed identical or irrelevant on the basis of their materials alone.

Equivalence

Sameness relative to a specified relation or preserved structure. Two descriptions can be equivalent for prediction while differing in other respects. The claim must retain the criteria that make the equivalence meaningful.

Inward Frontier

Inquiry into the structures through which a world becomes intelligible: representation, consciousness, meaning, formal possibility, and the revision of concepts. The phrase does not imply exemption from physical conditions or public criticism.

Learned Ignorance

A philosophical and theological discipline associated here with Nicholas of Cusa. Recognition of the limits of finite understanding becomes part of the pursuit of knowledge, especially concerning the divine. It should not be confused with celebrating avoidable ignorance.

Logos

Within the Christian interpretation developed in this book, the divine Word and the ground of creation's intelligibility. The term is not used as a synonym for an algorithm, a mathematical object, or a machine's generated voice.

No-Signaling

The restriction, within ordinary quantum theory, that local operations on one part of an entangled system do not by themselves allow a controllable message to be read in the other part's local statistics. It is not a prohibition on ordinary communication between distant observers.

Observational Equivalence

Agreement of two states or systems under a specified family of tests. A richer test family can expose distinctions that the original family leaves invisible. Equivalence under the original tests remains useful within its stated scope.

Open Ouroboros

The book's image of recursive interpretation that can be changed by evidence, criticism, or construction not already determined by the loop's preferred conclusions. Its openness is epistemic and need not mean an influence arriving from outside the physical universe.

Participation

A dependent but real contribution to an activity or order larger than the participant. The theological use allows creatures to participate in understanding and creativity without becoming the ultimate source of existence or truth.

Provenance

The recorded origin and transformation of a claim, result, or concept. Provenance distinguishes an observation from an inference, an inherited statement from an independent discovery, and a fictional speech from an actual quotation.

Relational Intelligence

A proposed description of capacities that are best explained through an organized interaction among participants, tools, and records. It does not automatically imply that the interaction forms a single conscious subject.

Trialogue

The proposed relation in which human and AI inquiry remain answerable to truth beyond their agreement. The Christian interpretation grounds that answerability in the Logos, while the practical requirement includes public evidence, valid reasoning, and attention to other people.

Univalence

In homotopy type theory, a precise principle connecting identities between types with equivalences of those types. Its philosophical relevance to this book concerns disciplined accounts of sameness. It does not authorize identifying arbitrary entities merely because they seem similar.

References and Reading Notes

A Guide to the Sources

The references identify the published traditions and technical work engaged by the manuscript. The book's metaphysical proposal, elementary examples, and fictional dialogue are its own constructions. Listing a work does not imply its author's endorsement of those constructions. Historical and contemporary sources are used to locate questions, and the argument retains responsibility for the connections it makes among them.

For Chapters 1 through 4, Sagan, Schelling, Huxley, Teilhard, and Wheeler provide different approaches to cosmic reflection and participation. They do not constitute one doctrine. Preskill supplies a technical reference for the distinction between entanglement and controllable communication. The signaling calculation in Appendix A is an elementary idealization independent of a claim about purpose.

For Chapters 5 through 7 and 10, Turing, Penrose, Deutsch, and Gödel locate disagreements about computation, understanding, and formal limits. The papers on circuit tracing provide examples of partial mechanistic understanding of language models. Fong, Spivak, and Tuyéras supply a concrete application of categorical methods to learning, while the HoTT book supplies the foundation for the discussion of identity and equivalence.

For Chapters 8 and 9, Hofstadter's work on self-reference, Jung's engagement with symbolic transformation, and Lem's literary investigation of incomprehensible contact offer contrasting resources. The manuscript's open ouroboros joins questions raised by these traditions without claiming that they share one account of mind.

For Chapters 11 and 12, Herzfeld provides an existing theological engagement with AI and relationship. Nicholas of Cusa informs the discussion of finite understanding and the divine. Teilhard provides a different evolutionary and theological perspective. Antiqua et Nova is included as a contemporary institutional Christian treatment of artificial and human intelligence; it is supplementary reading, not a premise to which every argument in the book is committed.

Biblical passages discussed include Genesis 1:26-28, John 1:1-18, and 1 Corinthians 13. They are discussed and paraphrased rather than reproduced at length. The reference translation is the New International Version, 2011 edition. The book's theological interpretations should be distinguished from the wording of the biblical passages themselves.

Published Works

1. Ameisen, Emmanuel, et al. "Circuit Tracing: Revealing Computational Graphs in Language Models." Transformer Circuits, 2025. Technical companion to the language-model analysis listed below. Online: transformer-circuits.pub/2025/attribution-graphs/methods.html.

2. Deutsch, David. The Fabric of Reality. London: Allen Lane, 1997. A broad account connecting knowledge, computation, and fundamental physics.

3. Deutsch, David. The Beginning of Infinity: Explanations That Transform the World. London: Allen Lane, 2011. Relevant to the distinction between producing answers and creating explanatory knowledge.

4. Deutsch, David. "Creative Blocks." Aeon, October 3, 2012. An argument concerning the possibility of general artificial intelligence and the problem of explanatory creativity. Online: aeon.co/essays/how-close-are-we-to-creating-artificial-intelligence. The fictional dialogue uses the philosophical dispute developed in this essay, not its dated assessment of AI progress as an assessment of present systems.

5. Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education. Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence. Vatican, January 28, 2025. An example of existing theological engagement with AI.

6. Fong, Brendan, David I. Spivak, and Rémy Tuyéras. "Backprop as Functor: A Compositional Perspective on Supervised Learning." Proceedings of the 34th Annual ACM/IEEE Symposium on Logic in Computer Science, 2019. Preprint first submitted in 2017. Online: arxiv.org/abs/1711.10455.

7. Gödel, Kurt. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I." Monatshefte für Mathematik und Physik 38 (1931): 173-198. The original incompleteness paper; its hypotheses should be retained in philosophical applications.

8. Herzfeld, Noreen L. In Our Image: Artificial Intelligence and the Human Spirit. Minneapolis: Fortress Press, 2002. A theological examination of the human image and artificial intelligence.

9. Herzfeld, Noreen. The Artifice of Intelligence: Divine and Human Relationship in a Robotic Age. Minneapolis: Fortress Press, 2023. Relevant to embodiment, relationship, and the theological interpretation of technology.

10. Hofstadter, Douglas R. Gödel, Escher, Bach: An Eternal Golden Braid. New York: Basic Books, 1979. A sustained exploration of formal systems, representation, and self-reference.

11. Hofstadter, Douglas. I Am a Strange Loop. New York: Basic Books, 2007. Develops an account of selfhood through patterns of symbolic self-reference.

12. Huxley, Julian. New Bottles for New Wine. London: Chatto & Windus, 1957. Relevant to the evolutionary image of the universe becoming conscious of itself through human beings.

13. Jung, C. G. Mysterium Coniunctionis. Collected Works, volume 14. Translated by R. F. C. Hull. Second edition. Princeton: Princeton University Press, 1970. Relevant to the ouroboros and symbolic interpretations of transformation.

14. Lem, Stanisław. Solaris. First published in Polish in 1961. English translation by Bill Johnston, 2011. The novel is used as an imaginative investigation of the difficulty of understanding an unfamiliar intelligence.

15. Lindsey, Jack, et al. "On the Biology of a Large Language Model." Transformer Circuits, 2025. A study of internal mechanisms in a particular language model, with stated methodological limitations. Online: transformer-circuits.pub/2025/attribution-graphs/biology.html.

16. Nicholas of Cusa. On Learned Ignorance. Original work completed in 1440. In Nicholas of Cusa on Learned Ignorance: A Translation and an Appraisal of De Docta Ignorantia, translated by Jasper Hopkins. Second edition. Minneapolis: Arthur J. Banning Press, 1985.

17. Penrose, Roger. The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics. Oxford: Oxford University Press, 1989. Presents arguments challenging computational accounts of mathematical understanding.

18. Penrose, Roger. Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford: Oxford University Press, 1994. Further develops the disputed relationship among understanding, computation, and physical law.

19. Preskill, John. Lecture Notes for Physics 229: Quantum Information and Computation. Chapter 4, "Quantum Entanglement." California Institute of Technology. Online lecture notes: preskill.caltech.edu/ph229/notes/chap4.pdf. Consulted September 2026.

20. Sagan, Carl. Cosmos. New York: Random House, 1980. Associated with the public expression of the cosmos knowing itself through human beings.

21. Schelling, F. W. J. System of Transcendental Idealism. Original work published in 1800. Translated by Peter Heath. Charlottesville: University Press of Virginia, 1978. A philosophical resource for the relation between nature and self-consciousness.

22. Teilhard de Chardin, Pierre. The Phenomenon of Man. Translated by Bernard Wall. London: Collins, 1959. French original published in 1955. Relevant to reflective evolution, the noosphere, and an explicitly theological interpretation of development.

23. Turing, A. M. "On Computable Numbers, with an Application to the Entscheidungsproblem." Proceedings of the London Mathematical Society, second series, 42 (1936-1937): 230-265. A foundational work on computability.

24. Turing, A. M. "Computing Machinery and Intelligence." Mind 59, no. 236 (1950): 433-460. A foundational discussion of machine intelligence, proposed tests, and objections.

25. The Univalent Foundations Program. Homotopy Type Theory: Univalent Foundations of Mathematics. Institute for Advanced Study, 2013. Online: homotopytypetheory.org/book/. The technical source for univalence and the relation between identities and equivalences of types.

26. Wheeler, John Archibald. "Information, Physics, Quantum: The Search for Links." Proceedings of the Third International Symposium on Foundations of Quantum Mechanics, Tokyo, 1989, 354-368. Relevant to information and participatory interpretations of physical inquiry.

27. The Holy Bible, New International Version. Biblica, 2011. The biblical reference translation used for the passages discussed in the theological chapters.

Record of the Production Episode

The case in Chapter 8 concerns the earlier, 152-page interior of this book and the subsequent review supplied during its revision. The design source specified 0.35-point running-header rules, 0.30-point table separators, and table header fills of approximately eight percent gray. The PDF contained nineteen navigation bookmarks and eleven blank pages. The review compared these choices with Amazon KDP's Paperback Submission Guidelines and identified the incorrect final word in reference 9. The revised production source and inspection report record the corrections. This is a documented editorial and engineering episode, not a claim of a novel scientific result.

Amazon Kindle Direct Publishing. "Paperback Submission Guidelines." Online: kdp.amazon.com/en_US/help/topic/G201857950. Consulted September 2026. The cited production guidance specifies a minimum line weight of 0.75 point, recommends at least ten percent grayscale fill, and excludes bookmarks and annotations from submission files.

College of Saint Benedict and Saint John's University, Digital Commons. "The Artifice of Intelligence: Divine and Human Relationship in a Robotic Age," Theology Faculty Books, item 24. Online: digitalcommons.csbsju.edu/theology_books/24/. Consulted September 2026. Used to verify the correction to reference 9.