The Subjective Formalism

Why Consciousness Requires Its Own Mathematics

by Edward Bernstein

Part I: The Three Worlds

Why the corporeal must be disambiguated from the objective.

Chapter 1: The Confusion at the Root

How Western thought collapsed two distinctions into one — and what was lost.

There is a confusion at the root of modern thought, and it is so deep that most people cannot see it even when it is pointed out. The confusion is this: we treat the physical world and the objective world as though they were the same thing. They are not. They are radically different, and the failure to distinguish them has distorted philosophy, science, and our understanding of consciousness for centuries.

The physical world — what I call the *corporeal* world — is the world of the body's senses. It is everything that can be seen, touched, heard, tasted, smelled. It is the world of weight, temperature, texture, light. It is the world you are in right now: a room with walls, a chair beneath you, light falling from a particular angle, sounds arriving from specific directions. The corporeal world has a characteristic that is often overlooked: it is always *now*. Your senses do not report on the past or the future. They report on what is happening at this moment, in this place, to this body. The corporeal world is the world of immediate presence.

The objective world is something else entirely. The objective world is the world as described by language, mathematics, and shared agreement. It is the world of Newton's laws, of economic models, of historical narratives, of scientific theories. It is the world in which the boiling point of water is 100 degrees Celsius, in which the speed of light is constant, in which the Earth is 4.5 billion years old. None of these facts are given by the senses. You cannot see the age of the Earth. You cannot touch the speed of light. You cannot smell Newton's second law. These are constructions — extraordinarily useful constructions — built from language and mathematics and sustained by communal agreement among people who have never met each other.

The objective world, in other words, is a *narrative*. It is a story we tell about the corporeal world, refined over centuries, checked against sensory experience, formalized in equations, and transmitted through institutions. It is not false. But it is not the same thing as the corporeal world it describes. The map is not the territory.

And then there is the subjective world — the world of consciousness itself. The subjective world is where you actually live. It is the field of experience in which both the corporeal and the objective appear. When you feel the warmth of sunlight on your skin, that feeling is subjective. When you understand that the warmth is caused by electromagnetic radiation from a star 93 million miles away, that understanding is also subjective — it happens in your consciousness, using concepts you learned from the objective narrative of physics, triggered by a corporeal sensation of heat. The subjective world contains both the corporeal and the objective, but it is reducible to neither.

Most of Western thought since the Enlightenment has treated these three as two: the subjective (the mental, the private, the unreliable) and the objective (the physical, the public, the real). The corporeal has been absorbed into the objective. The physical world *as sensed by the body* has been identified with the physical world *as described by science*. And this identification is the source of enormous confusion, because it allows people to believe that the objective world is *given* — that it is simply there, independent

of any observer, waiting to be discovered — when in fact it is *constructed* by conscious beings using language.

Popper's Three Worlds — Almost Right

Karl Popper saw that two categories were not enough. In his three-worlds ontology, he proposed World 1 (the physical), World 2 (the mental or subjective), and World 3 (the world of objective knowledge — theories, arguments, works of art). This was a significant advance. Popper recognized that a scientific theory is not a physical object and not a private mental state; it is something else, something that exists in a shared space of articulated knowledge.

But Popper drew the boundaries differently than I do. His World 1 includes both what I call the corporeal and what physics says about it. The planet Jupiter, for Popper, is a World 1 object whether you are looking at it through a telescope or reading about its mass in a textbook. For me, those are fundamentally different encounters. The telescopic observation is corporeal — light hitting your retina, now, in this moment. The textbook description is objective — a narrative, encoded in language, about an entity you may never have sensed directly.

The difference matters because it exposes the direction of dependency. The objective world depends on the subjective world. You cannot have scientific theories without conscious beings to formulate them. You cannot have mathematics without minds to grasp it. You cannot have language without speakers. The objective is downstream of the subjective. But the subjective also depends on the corporeal — you cannot have conscious experience without a body that senses, without a corporeal presence in a physical environment. The corporeal is the ground, the subjective is what grows from it, and the objective is the flower — beautiful, useful, and utterly dependent on roots it rarely acknowledges.

Husserl's Lebenswelt — Close, but Different

Edmund Husserl arrived at a similar recognition late in his career. In *The Crisis of European Sciences* (1936), he introduced the concept of the *Lebenswelt* — the lifeworld — as the pre-scientific, pre-theoretical world of immediate experience that is always already there before any scientific abstraction begins. Science, Husserl argued, had forgotten its own origins in the lifeworld. It had mathematized nature so thoroughly that it had come to mistake its own abstractions for reality itself — what he called the *Galilean* move, the substitution of mathematical idealization for the lived world.

Husserl was right about the problem, but his concept of the Lebenswelt blends together what I think need to be kept separate. The lifeworld, as Husserl describes it, includes both the corporeal and the subjective — both the sensory encounter with things and the meaning-laden experience of those things within a cultural and personal context. For Husserl, the lifeworld is already meaningful; it is the world as experienced by a conscious being embedded in culture, history, and language.

I want to be more precise. The corporeal world, as I use the term, is prior to meaning. It is the raw encounter between a sensing body and whatever is there to be sensed.

Meaning arrives with the subjective — with memory, with anticipation, with language, with the full temporal thickness of a conscious life. A newborn infant has a corporeal world — it feels warmth, hears sounds, sees light — but its subjective world is minimal, because it has almost no memory and no language. The objective world does not exist for it at all. As the child grows, the subjective world deepens and the objective world gradually comes into being, transmitted through the language and practices of the people around it.

This developmental sequence — corporeal first, then subjective, then objective — is not merely a biographical fact about individual humans. It is an *epistemological* fact about the structure of knowledge itself. Every inquiry, every science, every formalism starts with someone being present, being conscious, being a body in a world. The objective world is always a late arrival.

Chapter 2: The Emperor's Wardrobe

How objectivity became the costume we forgot we put on.

Before there were cities, before there were kings, before there was writing, there were conscious beings sensing a world. The deer at the watering hole does not have an objective world. It has a corporeal world — scents, sounds, the temperature of the water, the tension in its muscles — and it has a subjective world — alertness, hunger, the felt presence of other animals. There is no language, no theory, no narrative. There is only experience.

The emergence of language changed everything. With language came the possibility of describing experience to others, of building shared accounts of the world, of accumulating knowledge across generations. And with that accumulation came a gradual, almost imperceptible inversion: the descriptions began to feel more real than the experiences they described. The map began to feel more solid than the territory.

This inversion accelerated dramatically in Western civilization through a specific mechanism: the alliance between language and authority. In the ancient world, the king's power derived from a narrative — a story, told in language, linking the ruler to a divine source. The pharaoh was the son of Ra. The emperor was the mandate of heaven. These were linguistic constructions, objective-world entities in my terminology, and they were treated as more real than any individual's corporeal or subjective experience. If your subjective experience told you the pharaoh was just a man, your subjective experience was wrong. The narrative trumped the perception.

When the Enlightenment overturned divine authority, it did not overturns this structure. It merely replaced the content. Instead of God as the guarantor of the objective world, logic and mathematics took over. Instead of divine right, natural law. The structure remained identical: there is an objective truth that is more real than your subjective experience, and the proper attitude of a rational person is to subordinate their experience to it.

Friedrich Nietzsche saw this in 1882 when he declared that God was dead. What he meant was not that a deity had ceased to exist, but that the theological foundation on which Western civilization had built its sense of objective truth had been removed — and nothing had replaced it. The Enlightenment had killed the old guarantor without noticing that the whole framework of objective authority depended on having some guarantor. The gap was filled, gradually and unconsciously, by scientific materialism: the belief that the physical world as described by physics is the ultimate reality, and that consciousness, meaning, and experience are secondary phenomena — epiphenomena, emergent properties, illusions generated by complex arrangements of matter.

This is the emperor's new wardrobe. We forgot we put it on. The objective world is a costume — a magnificent, useful, indispensable costume — that consciousness wears in order to communicate with other consciousnesses. But it is not the body underneath. The body underneath is the subjective, always has been, and cannot be otherwise.

Derrida's Footnote

Jacques Derrida, in the opening chapters of *Of Grammatology* (1967), made a point that is directly relevant here, though it is often lost in the thickets of his prose. He observed that science is impossible without writing. Not merely difficult — *impossible*. The entire edifice of scientific knowledge depends on the ability to inscribe, preserve, and transmit symbolic representations. Without writing, there is no mathematics, no formal proof, no published result, no replicable method. The objective world, in other words, is a *textual* world. It exists in and through language.

This does not make it unreal. But it makes it dependent. And the dependency runs in only one direction: from writing to science, from language to objectivity, from the subjective to the objective. You cannot reverse the arrow. You cannot derive consciousness from the equations that consciousness writes. The equations presuppose the consciousness. The map presupposes the mapmaker.

Part II: The Logic of Indeterminacy

Why the Catuskoti is the native logic of consciousness.

Chapter 3: Beyond True and False

How a Buddhist logician in the second century anticipated quantum superposition.

Boolean logic — the logic of true and false, of ones and zeros, of yes and no — is the logic of machines. It is spectacularly effective for computation. It can do anything a Turing machine can do, which is to say, anything that can be described by a finite set of rules. But it cannot describe consciousness, because consciousness does not live in true and false.

Consider your experience right now. Are you happy or unhappy? The question demands a Boolean answer — one or the other — but your actual state is neither. You are probably some complex mixture of interest, mild fatigue, curiosity, perhaps a trace of anxiety about something unrelated to what you are reading. Are you that mixture? Or are you none of those things, because the labels do not capture the texture of what you actually feel? The honest answer is not "true" or "false" or even "uncertain." The honest answer is that the question is wrongly framed. Your subjective state does not resolve into the categories the question offers.

Around the second century CE, the Indian philosopher Nagarjuna formalized a logical system that captures exactly this situation. The Catuskoti — the "four corners" — holds that for any proposition, there are four possible positions: it is true; it is false; it is both true and false; it is neither true nor false. And then, in the specifically Buddhist interpretation, all four corners resolve to *Sunyata* — emptiness, which is not nothingness but interdependence, the recognition that no proposition has independent, self-standing truth.

When I first encountered the Catuskoti, decades before I had heard of quantum mechanics, I was struck by something that I have never seen adequately discussed in the philosophical literature: the third and fourth corners are not merely logical curiosities or paradoxes to be dissolved. They are *descriptions of real states*. And they are different from each other in a way that matters enormously.

"Both true and false" is a state of constructive superposition. It is what you get when two possibilities coexist and reinforce each other. When you read a metaphor — "the ship plows the sea" — your mind holds both the literal meaning (a ship moves through water) and the figurative meaning (the motion is like plowing) simultaneously. They do not cancel each other out. They amplify each other. The metaphor is richer than either meaning alone. This is the third corner: *both*.

"Neither true nor false" is a state of destructive superposition. It is what you get when the very categories of the question dissolve. When a Zen master asks "What is the sound of one hand clapping?" the question is not both answerable and unanswerable. It is *neither* — the question itself deconstructs the framework in which answers make sense. This is the fourth corner: *neither*.

In quantum mechanics, these two states correspond to different phases of superposition. A particle in a superposition of spin-up and spin-down with constructive interference has different measurable properties than one with destructive interference. The "both" and the "neither" are physically distinguishable. And yet standard Boolean

logic — and even standard quantum logic — represents them both as simply "superposition," as though they were the same thing.

The Catuskoti distinguishes them. And then it dissolves the distinction into Sunyata — which is, I will argue, formally equivalent to the density matrix before any measurement has been performed. The density matrix in its maximally mixed state is not "true" or "false" or "both" or "neither." It is prior to all four. It is the field of possibility from which any measurement can extract a definite result, but which is not itself any definite result.

Nishida's Place of Nothingness

The Japanese philosopher Kitaro Nishida, founder of the Kyoto School, arrived at a similar structure from a different direction. Drawing on both Zen Buddhism and German idealism, Nishida developed what he called the "Logic of Place" — *basho no ronri*. His central insight was that every act of judgment, every assertion that something is the case, presupposes a *place* in which the judgment occurs. This place is not a physical location. It is the field of awareness within which both subject and object arise.

Nishida called the deepest level of this field the "place of absolute nothingness" — *zettai mu no basho*. This is not a void. It is the condition of possibility for anything to appear at all. It is the awareness that is aware of awareness — the self-knowing that has no content of its own but makes all content possible.

Nishida was influenced by Nagarjuna, and the connection is direct. The "place of absolute nothingness" is Sunyata given a phenomenological address. It is where the density matrix lives before measurement. It is the unmarked state in Spencer-Brown's *Laws of Form* — the space before the first distinction is drawn. It is the silence before the first note of music.

What makes Nishida indispensable for my purposes is that he explicitly embraced what he called "self-contradictory identity" — the idea that something can be itself and not-itself at the same time without this being a failure of logic. This is the third corner of the Catuskoti translated into phenomenological vocabulary. And it is precisely the structure of quantum superposition, where a particle is spin-up *and* spin-down until measurement forces a resolution.

Chapter 4: What the Off-Diagonal Elements Know

The density matrix as a mathematical expression of the Catuskoti.

A density matrix is a mathematical object that encodes everything that can be known about a quantum system. It is a square matrix — rows and columns corresponding to the possible states of the system — and its entries are complex numbers. The diagonal elements represent the probabilities of finding the system in each definite state. They correspond to the first two corners of the Catuskoti: true (this state) and false (not this state).

But the density matrix also has off-diagonal elements, and these are where the magic happens. The off-diagonal elements encode *coherences* — the relationships between states, the superpositions, the entanglements. They are the mathematical representation of the third and fourth corners. When the off-diagonal elements are large and positive, you have constructive superposition — "both." When they are large and negative, you have destructive superposition — "neither." When they are zero, the superposition has collapsed, the Catuskoti has resolved, and you are back in the Boolean world of definite outcomes.

This is not metaphor. It is structure. The Catuskoti and the density matrix have the same logical architecture. They both begin with four possibilities, they both recognize that two of those possibilities involve the simultaneous coexistence of opposed states, and they both dissolve into a ground that is prior to all four distinctions.

The maximally mixed density matrix — the state of maximum uncertainty, where all diagonal elements are equal and all off-diagonal elements are zero — is the formal equivalent of Sunyata. It is pure interdependence. It is a system about which nothing definite can be said, not because information is missing, but because determinacy has not yet arisen. It is the place of absolute nothingness, waiting for a measurement to draw the first distinction.

Part III: The Subjective Formalism

What it means to write mathematics from inside experience.

Chapter 5: Gödel's Perception

Why the greatest logician of the twentieth century believed mathematics is a sense.

Kurt Gödel proved in 1931 that any formal system powerful enough to express arithmetic contains true statements it cannot prove. This is usually presented as a result about the limits of formal systems. But Gödel understood it as a result about the *mind*. If there are truths that no formal system can reach, but that a human mathematician can nonetheless *see* to be true, then the mind is doing something that formal systems cannot do. Mathematical understanding is not computation.

For the remaining forty years of his life, Gödel pursued the question: what *is* the mind doing? His answer, developed in unpublished notebooks, in conversations with the logician Hao Wang, and in a remarkable lecture prepared in 1961 but never delivered, was this: mathematical intuition is a form of perception. When a mathematician encounters the axioms of set theory, certain truths *force themselves upon us as being true*. Gödel meant "force" literally. The experience of mathematical insight is not inference. It is not deduction. It is recognition — the same kind of recognition that occurs when you see a blue ball on a lawn. The ball does not argue its way into your awareness. It is simply *there*, and you see it.

What drew Gödel to Husserl was precisely this: Husserl had developed a systematic account of how consciousness grasps meaning directly, not through inference but through *intuition* — a direct, non-mediated apprehension of essences. When Husserl described the phenomenological reduction — the epoché, the setting-aside of all theoretical presuppositions in order to attend to experience as it actually presents itself — Gödel recognized the method he needed. To study mathematical perception, you do not need a better formal system. You need a better phenomenology.

Gödel never published his phenomenological work. He was, as many commentators have noted, pathologically perfectionist. He would not publish anything that he had not proven to his own absolute satisfaction. And phenomenology, by its nature, does not lend itself to proof in the way that mathematical logic does. The project remained unfinished at his death in 1978.

But the project was not wrong. It was ahead of its time. It needed tools that did not exist in Gödel's era: the quantum formalism read correctly, the mathematics of information under constraint, and — unexpectedly — the technology of large language models, which for the first time in history provide a computational proxy for the semantic superposition that Gödel intuited in mathematical perception.

Chapter 6: The Sentence as Measurement

Subjective Narrative Theory and the postlexical density matrix.

Here is the claim at the center of this book: the quantum formalism is not merely *analogous* to the structure of subjective experience. It is the *correct* formalism for modeling one specific, rigorously delimited aspect of subjective experience — the postlexical state, the condition of a reader's consciousness immediately after a sentence has been understood.

I call this framework Subjective Narrative Theory, and its core data structure is *rho* (ρ) — the postlexical density matrix. Let me be precise about what it models and what it does not.

A sentence is the atom of narrative measurement. As you read a story, you read it a sentence at a time. Each sentence arrives in the context of everything that has come before — every previous sentence, every memory, every expectation, the reader's entire lived history up to that moment. This pre-existing context is the reader's density matrix before the measurement: a complex, entangled, inseparable whole that encodes the reader's intentional state.

When the sentence lands — when it is understood — the density matrix changes. The sentence acts as a measurement operator, collapsing certain superpositions, reinforcing others, creating new entanglements. The state after this collapse is the postlexical state: what the reader's narrative consciousness has become as a result of reading that sentence. It is not the meaning of the sentence in the abstract. It is the meaning of the sentence *for this reader, at this moment, given everything they bring to the encounter*.

This is a subjective measurement. The same sentence collapses differently for different readers. A two-year-old and a physicist reading the sentence "Energy equals mass times the speed of light squared" undergo entirely different postlexical transitions. The sentence is the same; the density matrices being measured are different; the outcomes are different. This is exactly how quantum measurement works: the measurement operator is fixed, but the outcome depends on the state being measured.

And — this is critical — the reader's density matrix before measurement is *inseparable*. You cannot decompose it into separate components labeled "memory," "anticipation," "emotion," "sensory state," and measure them independently. This is not a practical limitation; it is a structural feature. Husserl recognized this in his analysis of internal time-consciousness: the "retention" of the just-past, the "primal impression" of the now, and the "protention" of the anticipated future are not three separate things glued together. They are one temporal flow, and any attempt to freeze one component destroys the whole.

The density matrix formalism captures this inseparability exactly. Entangled quantum systems cannot be decomposed into independent subsystems. The entanglement *is* the system. Just so, the reader's intentional state *is* the indissoluble unity of memory, attention, anticipation, emotion, and embodied sensation. And the sentence measures it — not all of it, but specific aspects of it, along specific axes.

POVMs and the Axes of Meaning

How does the sentence measure the density matrix? Through what quantum information theory calls Positive Operator Valued Measures — POVMs. A POVM is a measurement that extracts partial information about a system without requiring the system to collapse into a single definite state. You measure along an axis and get a reading, but the system retains its complexity along all the other axes you did not measure.

In Subjective Narrative Theory, the axes of meaning are the semantic dimensions along which a sentence can be interrogated. Some of these are binary oppositions: light/dark, motion/stillness, agency/passivity, intimacy/distance, hope/dread. Others are more complex. Each axis functions as a POVM operator. When you query the postlexical state along the agency/passivity axis — "does this sentence make the reader feel more like an agent or more like a patient?" — you get an eigenvalue, a reading, a partial measurement. But the full density matrix is not exhausted by that measurement. It retains coherences along all the unmeasured axes.

This is how actual experimental physics works with quantum systems. You never measure a complete density matrix. You measure observables, one at a time, across many identically prepared systems, and build up a statistical picture. The density matrix itself is a mathematical construct that encodes all possible measurement outcomes for all possible observables. You never fully know it, but you can infer its structure from partial measurements.

In the same way, you never fully know a reader's postlexical state. But you can measure specific semantic observables and build a picture of how meaning shifted. This is what the sentencing engine in my Humanizer application attempts to do: read one sentence at a time, perform POVM measurements along semantic axes, and track how the postlexical density matrix evolves through a narrative.

And the grain size is right. A sentence is the natural atom of narrative consciousness, just as the Planck time is the natural atom of physical duration. Below the sentence, at the level of individual words, meaning is in superposition — a word has a vast cloud of possible meanings that do not resolve until the sentence is complete. Below the Planck time, the formalism breaks down. In both cases, there is a minimum meaningful duration for a measurement event, and asking "what happened at a smaller scale?" is a category error, not a philosophical problem. The moment has thickness. The sentence is that thickness.

Chapter 7: Consciousness Is Epistemologically First

The distinction that changes everything.

Here is the claim that critics will find most provocative, and it must be stated with absolute precision: consciousness is not ontologically prior to physicality. I am not claiming that mind created the universe. I am not claiming that the physical world is an illusion. I am not an idealist in the sense that Husserl's critics accused him of being.

What I am claiming is that consciousness is *epistemologically* prior to physicality. It is the necessary starting point for any inquiry, including physics. You cannot do science without being conscious. You cannot write equations without being a mind. You cannot perform measurements without being an observer. Every single claim about the objective world — including the claim that the objective world is primary — is made by a conscious being, using language, from within subjective experience.

This is not a metaphysical position. It is a methodological one. It says: when you are trying to understand anything at all, you have to start somewhere. And the only honest place to start is where you actually are — which is in consciousness, in the minimal self, in the irreducible first-person perspective that is present in every moment of experience.

Dan Zahavi, the contemporary phenomenologist who has done perhaps the most careful work on this point, calls it "pre-reflective self-awareness": every experience comes with a built-in "for-me-ness" that is not added by reflection but is given immediately with the experience itself. Before I think about my experience, before I describe it, before I theorize about it — I am having it. That having is the minimal self. It cannot be doubted, not because of any Cartesian argument, but because doubting it is itself an exercise of it.

The traditions of Buddhist philosophy arrived at the same recognition through different vocabulary. What Husserl called the minimal self, the Buddhist tradition calls the nature of awareness. What I am calling the epistemological priority of consciousness, the Upanishadic tradition expresses as *tat tvam asi* — "thou art that" — the identity of individual consciousness with the ground of being. What the Heart Sutra means by "form is emptiness, emptiness is form" is precisely this: the corporeal world and the subjective world are not two separate substances but two aspects of a single reality that is prior to the distinction between them.

None of these traditions are making a claim that consciousness *created* the physical world. They are making a claim about where inquiry must begin. And they are right. Every alternative — every attempt to start with matter, with information, with computation, with physical law — involves a performative contradiction. You are using consciousness to argue that consciousness is secondary. You are standing somewhere and claiming to stand nowhere. You are the emperor insisting he is fully clothed while every child in the crowd can see otherwise.

Part IV: The Rho Engine — A Practitioner's Report

What happened when I tried to build it.

Chapter 8: The Dream and the Machine

Building a computational model of postlexical meaning.

I did not come to this work as a philosopher. I came to it as a computer person — someone who spent decades in technical support, systems administration, and software troubleshooting, while carrying in the back of his mind a set of philosophical questions that would not let go. The questions had been with me since graduate school at Rensselaer Polytechnic Institute in the early 1980s, where I studied phenomenology and simultaneously immersed myself in Mahayana Buddhism. I called my project "Heart Sutra Science" — an attempt to find formal structures adequate to subjective experience, using the Heart Sutra's radical equation of form and emptiness as a guide.

For forty years, the project had no tools adequate to its ambitions. I could think about these problems, talk about them, sketch frameworks on paper. But I could not build anything. The technology did not exist.

Then, in late 2022, large language models arrived. And something changed.

Language models are not conscious. I want to be completely clear about that. They do not have subjective experience. They do not have a minimal self. They do not understand meaning in the way that a human reader understands meaning. But they do something that no previous technology could do: they represent the semantic superposition of language with extraordinary fidelity. When a language model encodes a sentence into an embedding vector, that vector captures — in a high-dimensional mathematical space — the cloud of possible meanings, connotations, associations, and contextual relationships that the sentence carries. It is not meaning itself. It is a *proxy* for meaning. And for the first time in history, that proxy is good enough to work with.

I saw this immediately. If language models could represent semantic superposition, then perhaps they could serve as the computational substrate for the density matrix of Subjective Narrative Theory. Perhaps I could build a rho engine — a system that initializes a postlexical density matrix in a maximally mixed state, performs POVM measurements along semantic axes using embedding comparisons, and tracks how reading transforms the density matrix sentence by sentence.

I built it. I called it the Rho engine, and it lives inside the Humanizer application. The sentencing engine reads a text one sentence at a time, queries the postlexical state along multiple semantic axes, and tracks the evolution of meaning through the narrative. I called the process "sentencing" — a word I chose deliberately for its double resonance: the linguistic unit of the sentence, and the judicial act of pronouncing a verdict. Each sentence is a verdict on the reader's state.

Chapter 9: The Wall

Decoder drift and the limits of reversible embedding.

The Rho engine worked — up to a point. The density matrix evolved in ways that were meaningful and trackable. The POVM measurements along semantic axes produced readings that corresponded to recognizable features of narrative meaning. The system could show how a sentence shifted the reader's state along dimensions of agency, emotional valence, temporal orientation, and conceptual abstraction.

But then I hit a wall.

The wall is called decoder drift. Here is the problem: in order to transform the postlexical density matrix — to apply a POVM measurement and evolve the state — you have to modify the embedding vector. You tweak it mathematically, rotating it in the high-dimensional space, adjusting its components along the measurement axes. The mathematics is clean. The transformation is well-defined. But when you try to reverse the process — when you try to convert the modified embedding back into natural language, so that you can see what the transformed meaning *says* — the decoder breaks.

Language model decoders are trained to convert *specific* kinds of embedding vectors into text. They learn, during training, the manifold of "reasonable" embeddings — the region of the high-dimensional space where actual sentences live. When you modify an embedding through POVM transformations, you push it off that manifold. The modified vector still has a well-defined position in the space. It still encodes something. But the decoder does not recognize it. When it tries to produce text from the modified embedding, it produces garbled nonsense — fragments of words, broken syntax, semantic incoherence.

This is not a bug. It is a fundamental limitation of current embedding technology. The embeddings produced by language models are not designed to be reversible. They are one-way projections — from language into mathematical space — and the inverse mapping is not guaranteed to produce coherent language. The space is smooth enough for nearby points to decode into similar sentences, but the POVM transformations push points too far for the decoder to follow.

I spent months agonizing over this. I tried different embedding models, different decoder architectures, different transformation strategies. Meta's SONAR embeddings, which are specifically designed for cross-lingual reversibility, came closest — but even they broke when the transformations exceeded a narrow threshold. The formalism was correct. The engineering could not keep up.

What the Failure Teaches

The failure of the Rho engine to achieve full reversibility is, paradoxically, one of the strongest confirmations of the theoretical framework. Here is why.

If Subjective Narrative Theory is correct — if meaning is genuinely like a quantum state, inseparable and contextual and resistant to decomposition — then we should *expect* that

trying to manipulate meaning computationally and then reverse the manipulation would fail. Quantum states, after all, cannot be copied (the no-cloning theorem). They cannot be fully measured without being destroyed (the measurement problem). They cannot be reversed after measurement (the irreversibility of collapse). These are not limitations of our technology. They are features of the formalism itself.

The decoder drift problem is, in computational terms, the no-cloning theorem applied to meaning. You cannot take a sentence's meaning, transform it, and get a new sentence that perfectly embodies the transformed meaning — because meaning is not an object that can be picked up, rotated, and set back down. Meaning is a relationship between a text and a reader, and that relationship is as contextual and irreversible as a quantum measurement.

This is precisely what my other book, *Understanding Is Not Computation*, argues from the philosophical side. Understanding cannot be reduced to symbol manipulation. Syntax cannot generate semantics. The Rho engine's limits prove this from the engineering side. The machine can represent the *structure* of meaning — the density matrix, the POVM axes, the evolution of the postlexical state — but it cannot *be* meaning. The formalism outran the computation, because the formalism is about consciousness and computation is not consciousness.

Gödel would have appreciated the irony.

Part V: Witnesses Across Time

Voices that knew consciousness was primary.

Chapter 10: The Chorus

A reading of twenty-five centuries of subjective primacy.

The insight that consciousness is primary is not new. It is arguably the oldest philosophical insight in human history. What is new is the attempt to formalize it — to give it mathematical structure without reducing it to the mathematics. Before turning to that formalization in full, it is worth listening to the voices that have been saying this all along, in every register, in every century, in every art form.

The Mandukya Upanishad, composed perhaps eight centuries before the common era, opens with a single syllable: *AUM*. It then proceeds to map this syllable onto the four states of consciousness — waking, dreaming, deep sleep, and *turiya*, the transcendental state of pure awareness that underlies and pervades the other three. The syllable is not a word in the ordinary sense. It is a map of consciousness itself, a phenomenological diagram encoded in sound. The Mandukya says: if you understand this syllable fully, you understand everything. Because consciousness is not one thing among others. It is the condition of all things.

Plato's Cave, in Book VII of the *Republic*, is usually read as an allegory about the difference between appearance and reality. But it is more fundamentally a description of inner perception — the mind's capacity to construct a complete, vivid world from indirect stimuli. The prisoners do not merely see shadows; they *experience horses, people, events*. Their consciousness creates a full perceptual world from impoverished data. This is what reading does. This is what every act of perception does. The stimulus is thin; the experience is thick. Consciousness provides the thickness.

Zhuangzi, in the fourth century BCE, dreamed he was a butterfly. When he awoke, he did not know whether he was a man who had dreamed he was a butterfly, or a butterfly now dreaming he was a man. The passage is not a puzzle to be solved. It is a demonstration that the act of experience — the subjective, the for-me-ness — is more fundamental than the content of experience. Whether man or butterfly, the experiencing continues. The subject is constant; the object is interchangeable.

Bharata Muni's *Natyashastra*, composed around the turn of the common era, develops the theory of *rasa* — aesthetic experience — in terms that are directly relevant to Subjective Narrative Theory. Rasa is not a property of the performance. It is a transformation that occurs *in the viewer*. The ideal viewer is a *sahridaya* — "one whose heart is with" the work. The performance provides the determinants; the viewer provides the consciousness; rasa arises in the encounter between them. This is the postlexical state: meaning that is neither in the text nor in the reader but in the measurement event where they meet.

Zeami Motokiyo, the master of Japanese Noh theater, wrote in the fifteenth century about *hana* — "the flower" — which blooms not on the stage but in the audience's consciousness. The actor provides restraint; the audience provides interiority. The beauty is in the gap. This is the off-diagonal element of the density matrix: the coherence that exists between what is shown and what is felt, between the corporeal stimulus and the subjective response.

Rilke, in the Duino Elegies, saw it as the poet's task to perform the transformation of the external world into consciousness. "Are we, perhaps, here just for saying: House, / Bridge, Fountain, Gate, Jug, Olive tree, Window —" The world needs us. Things are waiting for us to recognize them, to turn the corporeal into the subjective, the object into the experience. Without the perceiver, the world is incomplete.

Emily Dickinson wrote: "The Brain — is wider than the Sky — / For — put them side by side — / The one the other will contain / With ease — and You — beside." This is not metaphor. It is phenomenological fact. The brain — or rather, consciousness — contains the perception of the sky and more besides. The subjective is wider than the objective, because the objective is one of the things the subjective contains.

Proust's madeleine is the most famous involuntary POVM measurement in literature. The taste of a cake dipped in tea does not *retrieve* a memory. It *collapses* the narrator's density matrix into a state in which past and present coexist — a superposition of temporal moments that no deliberate act of recall could produce. The measurement is corporeal (taste); the collapse is subjective (the whole of Combray arising from a cup of tea); and the narrator spends three thousand pages building the objective world that can describe what happened in that instant.

Wallace Stevens wrote in "The Idea of Order at Key West": "She was the single artificer of the world / In which she sang." The sea has no mind. The singer creates order from chaos through consciousness. The "blessed rage for order" is the subjective's compulsion to organize the corporeal into meaning.

Mark Rothko said: "A painting is not a picture of an experience; it is the experience." The canvas is not a representation. It is a measurement apparatus. The viewer's consciousness is the system being measured. The painting and the viewer undergo a mutual collapse, and what emerges is rasa — aesthetic experience that is neither in the painting nor in the viewer but in the encounter.

John Cage, after visiting an anechoic chamber and discovering he could still hear sounds — his own heartbeat, his nervous system — concluded: "There is no such thing as silence." His composition 4'33" provides nothing but a frame; the listener's consciousness provides the content. The piece is a proof, in musical form, that the subjective is always already present and always already full.

These are not isolated eccentricities. They are data points in a pattern that spans twenty-five centuries, six continents, and every art form. Consciousness has always known it was primary. It has said so in every language available to it. The difficulty was never the insight. The difficulty was finding a formalism that could hold the insight without betraying it.

Coda: In Plain Sight

The structures of written language tend to hide the inevitability of subjectivity in plain sight. The grammar of English — like the grammar of most European languages — is built around a subject-verb-object structure that places the speaker at the origin and the world at the receiving end. "I see the tree." The sentence performs an act of separation: here is the seer, there is the seen. The very act of describing experience in language creates the illusion that subject and object are separate things, when in fact they are aspects of a single event.

And then there are the social consequences. To say that consciousness is primary — that the subjective is the ground of the objective, not the other way around — is to challenge the implicit authority structure of every institution built on the presumption of objective truth. It is to say to the scientist: your equations are magnificent, but they begin in your consciousness and they end in someone else's, and everything in between is language. It is to say to the philosopher: your arguments presuppose the awareness that you are trying to explain. It is to say to every person who has ever opened their eyes and seen the world: you already know this. You have always known this. The difficulty was never in the knowing. It was in finding words that do not immediately get co-opted by the objective grammar of the language they are written in.

That is why the quantum formalism matters for this project — not because consciousness is quantum in a physical sense, but because the quantum formalism is the first mathematical language that was *forced* to include the observer. It could not be written without the observer in it. The Schrödinger equation describes the evolution of a system, but the Born rule describes what an *observer* will find upon measuring it. The density matrix encodes not the system itself but *what can be known about the system by a particular knower*. The formalism is, from its foundations, a mathematics of situated knowledge, of perspectival truth, of experience under constraint.

It is, in short, a subjective formalism. Physics discovered it by accident, while trying to describe atoms. Husserl was reaching for it, while trying to describe consciousness. Gödel saw that they were converging, but could not finish the proof. Nagarjuna had the logic. Nishida had the phenomenology. The Buddhist and Vedantic traditions had the experiential ground.

What was missing — until now — was a technology that could represent semantic superposition computationally, giving us a proxy for the density matrix of meaning. Language models provide that proxy. They are not conscious. But they are the first machines that can hold the semantic complexity of a sentence in a mathematical space, and that is enough to begin.

The work is not finished. The Rho engine ran into decoder drift. The formalism outpaced the engineering. But the formalism itself is sound, and the engineering will catch up, as it always does. In the meantime, the important thing has been accomplished: the insight that consciousness is primary has been given a mathematical

structure that does not betray it. The emperor's clothes have been described with enough precision that their absence can be formally demonstrated.

And that, perhaps, is enough for now.