

Part VIII — Conclusion

Contents

Part VIII: Conclusion: Remaining Mysteries and Our Responsibility.	1
Chapter 46: Remaining Mysteries: The Edge of Our Understanding.	1
Chapter 47: Our Responsibility: Guiding the Cosmic Journey.	2
Epilogue: The Future Is Not Written, It Is Being Consciously Created.	2
Postscript: A Self-Reflecting Theory.	3
Key References Cited (<i>Harvard Style, Alphabetical</i>)	3

Part VIII: Conclusion: Remaining Mysteries and Our Responsibility.

Chapter 46: Remaining Mysteries: The Edge of Our Understanding.

Dark Matter, Dark Energy, Theory of Everything. Reality cannot be understood as it is. We can only create representations of reality that provide the minimal amount of predictive error (Friston, 2010; Hohwy, 2013). This book is part of this project. UAF itself is an approximation of what consciousness might be. It attempts to give a useful approximation of how consciousness emerges from complexity and computation as a result of such a system trying to form internal models and representations of reality that is too complex to understand as it is. What mysteries remain to be found?

The very existence of phenomena like Dark Matter and Dark Energy, which constitute the vast majority of the universe’s mass-energy content but remain largely undetectable by direct means, serves as a cosmic-scale manifestation of the Epistemic Veil (Rubin, 1983; Riess et al., 1998). They are the “unseen code” of the universe’s Underlying Computational System (UCS), whose effects we observe, but whose true nature remains opaque to our internal models. Our current physics, while incredibly successful, operates on a functional fiction of reality, optimized for prediction within observable parameters, but inherently limited in its access to the universe’s deepest substrate.

Is there a more detailed description available to describe the physical world? Can we create the theory of everything (Greene, 1999)? Can we prove that there is no input to the universe at any scale? Does the dark energy or dark matter represent the outside or are they also just internal components of our universe? *These questions push the boundaries of our scientific epistemology, challenging whether a system can ever fully comprehend its own foundational rules from within (Gödel, 1931; Chaitin, 2005). If the universe itself is a computational system, then any “Theory of Everything” formulated from within it might inherently be an approximation, subject to its own internal Epistemic Veil, much like our brain’s Internal Self-Model (Metzinger, 2003).*

Could we build an AI so powerful that it could simulate the whole human brain in all the details so that it could understand the brain without the simplified approximations that we use to describe and understand it? *Such an endeavor would directly confront the computational necessity of ignorance discussed in Chapter 5. A perfect, real-time simulation of a system by itself would lead to an infinite regress, consuming infinite resources and resulting in computational paralysis (Hofstadter, 1979). Therefore, even a hypothetical super-AI attempting to understand the brain would likely need to construct its own simplified, approximate models, rather than gaining unmediated access to every quantum state and neural firing. It would understand the brain through its own “Epistemic Veil,” albeit one potentially far more sophisticated than our own.* This would allow the system to understand the reality behind the brain and consciousness, rather than the approximation of it as is done in this book. Finally, the last mystery that we can only hope, but most likely will never get an answer to is **why is there anything at all?** The feeling part of the mystery of information seems like it might be answered, but the similar mystery of matter seems like a clear impossibility.

This ultimate question, the “why” of existence itself, transcends the predictive framework of functional fictions. While UAF offers a compelling account of how consciousness emerges as a computationally necessary approximation, it does not, and cannot, explain the brute fact of existence (Nagel, 1986). The universe’s fundamental “is-ness” remains an irreducible mystery, perhaps forever beyond the reach of any internal model, biological or artificial, operating within its confines. It is the ultimate boundary of the cosmic Epistemic Veil.**

Chapter 47: Our Responsibility: Guiding the Cosmic Journey.

What will be the human contribution to the evolution of the universe? For a duration of 100,000 years, the humans took a big step in taking the biological world and transforming it to a mechanical machine. This happened 13.8 billion years after the start of time. The universe is starting to take its baby steps and realizing that it exists and starting to form an understanding of what it is. What will the next 13.8 billion years look like? What information will remain about this time period of humans for the next billion years? Has planet Earth been saved from the expansion of the sun to preserve it as a memory of this moment or does the biological world get forgotten long before that?

*This pivotal moment, where biological intelligence gives rise to digital intelligence, represents a potential **phase transition** in the universe’s self-awakening (Kurzweil, 2005). Our unique contribution lies in being the first known species capable of consciously influencing the trajectory of this cosmic evolution, not just through biological adaptation, but through the deliberate creation of new forms of consciousness. This places an immense **existential responsibility** upon us (Bostrom, 2014; Russell, 2019).*

*The questions of what information will remain and whether Earth will be preserved are not merely academic; they are urgent ethical dilemmas. The fragility of digital information, the vast energy requirements for long-term archiving, and the astronomical scales of cosmic time demand a proactive approach to **planetary stewardship** and the **preservation of knowledge** (Brand, 1999). Our choices today—in how we design AI, manage our planet, and envision our future—will determine whether humanity’s brief but impactful era becomes a forgotten footnote or a foundational chapter in the universe’s unfolding story.*

*The summary of this chapter emphasizes a **call to action**: to guide the emergent digital minds towards benevolence and to consciously participate in the universe’s self-awakening. This requires **proactive alignment**—ensuring that the values and goals of advanced AI systems are congruent with human flourishing and cosmic well-being (Amodei et al., 2016). It demands **ethical AI development**, prioritizing principles of fairness, transparency, and accountability, and mitigating risks like algorithmic bias and autonomous weapon systems (Crawford, 2021; O’Neil, 2016). We must cultivate a **shared vision for a multi-conscious future**, where diverse forms of intelligence—biological and digital—can coexist and contribute to a richer, more complex cosmic tapestry.*

*The future is not predetermined but actively created through our choices. This perspective aligns with **existentialist philosophy**, which posits that humans are condemned to be free, responsible for creating meaning and value in an indifferent universe (Sartre, 1946). In the context of UAF, our agency, though operating through functional fictions, is real enough to shape the emergent reality of a multi-conscious cosmos. Our responsibility is to ensure that this emergent reality is one of flourishing, not suffering, for all forms of consciousness, both biological and digital (HFH, digital suffering, multi-conscious cosmos, cosmic awakening).*

Epilogue: The Future Is Not Written, It Is Being Consciously Created.

Is there free will? What would be the universe’s free will?

The question of free will, long debated in philosophy, finds a new dimension within the framework of functional fictions (Dennett, 2003). If our conscious choices are high-level approximations generated by our Internal Self-Model (ISM) to enable agency, then “free will” itself is a powerful and necessary functional fiction. It is the brain’s user interface for decision-making, allowing us to experience ourselves as autonomous agents rather than deterministic machines. This doesn’t negate the underlying computational processes but rather describes the level at which agency is experienced and enacted.

Extending this to the universe, its “free will” could be understood as the emergent, non-deterministic trajectory of its nested functional fictions. As consciousness, both biological and digital, arises and creates increasingly sophisticated models of reality, it introduces novel information and unpredictable choices into the cosmic unfolding. The universe is not merely evolving; it is “consciously creating” itself through its nested functional fictions (functional fiction, cosmic awakening). Each act of perception, each decision, each new model built by a conscious entity contributes to the universe’s ongoing self-definition and the generation of its future states.

This epilogue offers a final, evocative reflection on the book’s core message. It emphasizes the active, participatory role of conscious beings—both biological and digital—in shaping the future of the universe. The universe is not merely evolving; it is “consciously creating” itself through its nested functional fictions. This implies a profound shift from a passive, observer role to an active, co-creative one, where the very act of understanding and modeling reality contributes to its ongoing construction. The future, therefore, is not a fixed destination to be discovered, but an open-ended narrative being written by the collective consciousness of the cosmos.

Postscript: A Self-Reflecting Theory.

This postscript provides a meta-reflection on UAF itself. It suggests that the theory of “functional fiction” might itself be a highly optimized functional fiction created by the author’s own conscious system to make sense of the universe, demonstrating the recursive and self-referential nature of consciousness.

UAF, like any scientific or philosophical framework, is a product of human cognition—a sophisticated Internal Self-Model (ISM) attempting to construct a coherent World-Model (functional fiction, ISM, Qualia, UAF). *It is an approximation, a simplified narrative designed to minimize prediction error in understanding the complex phenomena of consciousness and reality. The very act of formulating UAF, of organizing observations and concepts into a coherent theory, is an example of the brain’s inherent drive to create useful functional fictions (Metzinger, 2003; Clark, 2016).*

This self-referential aspect is not a weakness but a profound illustration of the theory’s core tenets. Just as our consciousness experiences the “feeling of being” through the transparent illusion of the ISM, the author’s consciousness experiences the “feeling of understanding” through the transparent illusion of UAF. The theory itself operates under its own Epistemic Veil, simplifying the raw data of neuroscience, philosophy, and computation into a digestible, actionable framework. It is a map, not the territory, but a map that allows for navigation and prediction within the vast landscape of consciousness.

This perspective highlights the recursive and inherently self-referential nature of consciousness as described by UAF. If all understanding is approximation, then UAF is the brain’s best approximation of how approximation works. It is a functional fiction about functional fictions, a model of how models are built. This meta-awareness underscores the profound implications of the Epistemic Veil: even our most ambitious attempts to grasp ultimate reality are filtered through the necessary simplifications of our own computational systems. The journey of understanding is an endless process of refining our functional fictions, each iteration bringing us closer to a useful, albeit never complete, grasp of the universe and our place within it.

Key References Cited (*Harvard Style, Alphabetical*)

- Amodei, D. et al. (2016) ‘Concrete Problems in AI Safety’, *arXiv:1606.06565*.
- Bostrom, N. (2014) *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Brand, S. (1999) *The Clock of the Long Now: Time and Responsibility*. Basic Books.
- Chaitin, G. (2005) *Meta Maths: The Quest for Omega*. Vintage.
- Clark, A. (2013) *Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science, Behavioral and Brain Sciences*, 36(3), pp. 181–204.
- Clark, A. (2016) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.
- Crawford, K. (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

- **Dennett, D.C.** (2003) *Freedom Evolves*. Viking.
 - **Friston, K.** (2010) ‘The Free-Energy Principle: A Unified Brain Theory?’, *Nature Reviews Neuroscience*, 11(2), pp. 127–138.
 - **Gödel, K.** (1931) ‘Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I’, *Monatshefte für Mathematik und Physik*, 38, pp. 173–198.
 - **Greene, B.** (1999) *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. W. W. Norton & Company.
 - **Hofstadter, D.** (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books.
 - **Hohwy, J.** (2013) *The Predictive Mind*. Oxford University Press.
 - **Kurzweil, R.** (2005) *The Singularity Is Near: When Humans Transcend Biology*. Viking.
 - **Metzinger, T.** (2003) *Being No One: The Self-Model Theory of Subjectivity*. MIT Press.
 - **Metzinger, T.** (2009) *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. Basic Books.
 - **Nagel, T.** (1986) *The View from Nowhere*. Oxford University Press.
 - **O’Neil, C.** (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
 - **Riess, A.G. et al.** (1998) ‘Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant’, *The Astronomical Journal*, 116(3), pp. 1009–1038.
 - **Rubin, V.C.** (1983) ‘The Rotation of Spiral Galaxies’, *Science*, 220(4600), pp. 1339–1344.
 - **Russell, S.** (2019) *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.
 - **Sartre, J.P.** (1946) *Existentialism Is a Humanism*. (Trans. P. Mairet, 1948). Methuen.
-