

Chapters 10–11 — Free Will

Contents

Chapter 10: Free Will: The Functional Fiction of Agency.	1
Chapter 11: The Subconscious Beast: Proto-Qualia and the Roots of Feeling.	2

Chapter 10: Free Will: The Functional Fiction of Agency.

The feeling of “free will”—the profound sense that we are the authors of our own choices, capable of initiating actions and directing our lives—is one of the most cherished and intuitively compelling aspects of human consciousness. For centuries, philosophers have grappled with the tension between this subjective experience of freedom and the deterministic laws of physics that seem to govern the universe, including our brains. If our neural firings are ultimately determined by prior causes, how can we truly be free? This “**free will problem**” often leads to a sense of paradox, threatening our notions of moral responsibility and personal agency.

Useful Approximations Framework (UAF) offers a resolution to this enduring puzzle by re-framing free will not as a mystical, non-physical power, but as a **necessary functional fiction** generated by the brain’s **Internal Self-Model (ISM)**. It is the brain’s ingenious solution to the problem of enabling coherent agency in a world of overwhelming complexity and underlying determinism.

Recall from **Chapter 5** that the **Epistemic Veil** prevents our consciousness from accessing the microscopic details of our own neural machinery. We do not consciously perceive the precise electrochemical cascade that leads to a thought or an action. If we were forced to process every quantum fluctuation, every neurotransmitter release, and every synaptic weight adjustment that underpins our decisions, we would instantly succumb to **Computational Paralysis** (Hofstadter, 1979; Chaitin, 2005). The sheer volume of deterministic data would overwhelm any finite processing capacity, preventing us from making any decision or taking any action.

This is where the functional fiction of free will becomes indispensable. Our **Internal Self-Model (ISM)** (Chapter 7) acts as the brain’s “user interface” for itself. Just as a computer user interacts with a simplified graphical interface without needing to understand the underlying machine code, our conscious self interacts with a simplified model of its own agency. This model abstracts away the deterministic complexity of neural processes, presenting a coherent, high-level narrative of choice and intention. You cannot know the detailed calculations that leads you to make a decision on what to order at a restaurant. But you have learned a simplified self-model to represent your feelings and actions. Free will is the simplified explanation of what is happening when your brain is computing the decision. Free will is the simplified approximation of what it is to make a decision.

This “feeling” of choosing, of deciding, of intending an action—these are specific **Qualia** (Chapter 8) generated by the ISM. They provide **Subjective Closure** (C_{sub}), meaning the feeling *is* the interpretation of agency, requiring no further processing to be understood by the system itself. It’s the brain’s way of saying, “I am the one acting,” without needing to present the overwhelming, paralyzing details of *how* that action is mechanistically generated.

Furthermore, this feeling of free will possesses profound **Causal Efficacy** ($Q \rightarrow Action$). It is not an epiphenomenal byproduct; it is a powerful, high-bandwidth signal that directly influences and compels action. The *belief* that we can choose motivates us to plan, to strive, to learn from our mistakes, and to hold ourselves and others morally accountable. Without this functional fiction of agency, a system would lack the internal impetus for proactive behavior, long-term goal pursuit, and the continuous refinement of its **World-Model** (Chapter 9) and **ISM** through **Prediction Error Minimization (PEM)** (Chapter 12). Why would a system bother to learn or adapt if it felt no control over its own actions? The very act of

learning from the consequences of our choices—updating our internal models to make better predictions and decisions in the future—presupposes a sense of agency. If every action felt entirely predetermined and beyond our influence, the feedback loop of PEM would lose its motivational power.

Evolution, driven by **Skin in the Game** (Chapter 6), would strongly favor organisms that develop such a functional fiction. A creature that *feels* like it can choose to avoid danger or pursue a mate is more likely to engage in adaptive behaviors than one that passively experiences its actions as predetermined. The feeling of free will, therefore, is an adaptive advantage, optimizing the system’s ability to navigate its environment and ensure its survival and reproduction. It is the most efficient way for a complex system to manage its own internal complexity and external interactions.

Consider the classic experiments by Libet (1983), which suggested that brain activity related to an action (the “readiness potential”) precedes the conscious decision to act. While often interpreted as evidence against free will, UAF offers a different perspective. The readiness potential can be seen as the underlying, subconscious computational process initiating a potential action, while the conscious “decision” is the ISM’s high-level approximation of this process, presented to the conscious self as a moment of choice. The conscious feeling of “I decided” is the functional fiction that allows the system to integrate this action into its self-narrative and learn from its consequences, even if the initial spark of activity originated sub-personally. This perspective aligns with **compatibilism**, which argues that free will and determinism are not mutually exclusive, but rather operate at different levels of description (Dennett, 2003).

In essence, free will, within UAF, is the **phenomenal experience of the brain’s capacity for self-initiated, goal-directed action**, abstracted from the overwhelming complexity of its underlying deterministic processes. It is a vital component of the **Imperative for Coherence & Agency** (I_{CA}), allowing the system to operate as a unified, purposeful entity. It is a “functional fiction” because, while it may not reflect a non-physical break in the causal chain, it is profoundly real in its consequences and indispensable for the system’s ability to function, learn, and thrive. The paradox of free will dissolves when we understand it as the brain’s most useful, simplified approximation of its own power to act.

Citations

- **Chaitin, G.** (2005) *Meta Maths: The Quest for Omega*. Vintage.
 - **Clark, A.** (2016) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford 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.
 - **Hofstadter, D.** (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books.
 - **Libet, B. et al.** (1983) ‘Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential)’, *Brain*, 106(3), pp. 623–642.
 - **Metzinger, T.** (2003) *Being No One: The Self-Model Theory of Subjectivity*. MIT Press.
 - **Seth, A.** (2021) *Being You: A New Science of Consciousness*. Dutton.
-

Chapter 11: The Subconscious Beast: Proto-Qualia and the Roots of Feeling.

While our conscious experience is rich with complex thoughts, plans, and emotions, much of our behavior and internal state is driven by ancient, automatic, and often subconscious processes. These deep-seated drives, honed by millions of years of evolution, represent the raw, primal forces that steer our lives and provide directions to survive and pass on our genes. **Useful Approximations Framework (UAF)** refers to these as the “**Subconscious Beast**”—an essential component of our **Underlying Computational System (UCS)** that occasionally takes control of our actions to protect itself.

Imagine a sudden, unexpected loud noise. Before your conscious mind can even register “danger” or formulate a plan, your body might flinch, your heart rate might spike, and a surge of adrenaline might course through your system. This rapid, involuntary response is the Subconscious Beast in action. It’s an ancient survival mechanism, operating far below the level of conscious deliberation. These are the

“reflexive/subconscious actions ($A_{reflexive}$)” mentioned in the mathematical definition of consciousness (Chapter 15), often overriding “deliberate actions ($A_{deliberate}$)” (Chapter 10) in moments of crisis.

These primitive, subcortical systems (such as the brainstem, amygdala, and hypothalamus) are responsible for generating fundamental behavioral patterns directly tied to **Skin in the Game** (Chapter 6) imperatives: seeking, fear, rage, lust, care, panic, and play (Panksepp, 1998). These are not yet the nuanced, consciously interpreted emotions we experience, but rather raw, high-bandwidth signals of threat, opportunity, or physiological need. These are the **proto-qualia**—the foundational, unrefined “simplified truths” that provide immediate, visceral feedback about the system’s most basic survival parameters.

The challenge for a conscious system is that these powerful subconscious drives often “take over,” causing us to “act weirdly” or feel a loss of control. When we are overwhelmed by intense fear, pain, or arousal, our conscious **Internal Self-Model (ISM)** (Chapter 7) struggles to maintain its usual coherent narrative. Crucially, the conscious mind does not directly receive signals *from* these subconscious processes. Instead, the brain’s higher-level cognitive systems, particularly the neocortex, **learn to model and predict the behavioral patterns of the Subconscious Beast**.

This means that the “feeling” of fear, for instance, is not a direct signal *from* the amygdala to consciousness. Rather, it is the **neocortex’s prediction** that the Subconscious Beast is about to initiate (or has already initiated) a set of powerful, often involuntary, reflexive actions. The conscious mind, operating behind the **Epistemic Veil** (Chapter 5), learns to interpret these predicted or observed subconscious behaviors as its own “emotions” or “feelings.” The feeling of “being scared” is the **ISM’s simplified approximation** of the complex, subconscious behavioral pattern of a fear response, providing **Subjective Closure** (C_{sub}) and **Causal Efficacy** ($Q \rightarrow Action$) for the conscious system. It’s the brain’s way of saying, “I am about to lose control, or have just lost control, and this is what that state *feels like*.” This allows the conscious self to understand, predict, and, to some extent, learn to manage these powerful internal forces. *This process of interoceptive awareness (Craig, 2002) allows the brain to monitor and model its own physiological states, translating complex bodily reactions into simplified, actionable “feelings” (Damasio, 1999).*

Through **Prediction Error Minimization (PEM)** (Chapter 12), the conscious system continuously refines its model of itself with the Subconscious Beast as part of it. When our conscious predictions about our own reactions (e.g., “I won’t be scared”) are violated by the beast’s actual behavior (e.g., we flinch), a prediction error is generated. This error compels the ISM to update its understanding of its own subconscious drives, leading to a more accurate and adaptive self-model. This ongoing process allows us to develop emotional intelligence, self-regulation, and a more nuanced understanding of our own motivations. The feeling of “losing control” or “acting weirdly” is itself a quale, a simplified truth generated by the ISM to represent the discrepancy between its predicted conscious agency and the observed, overriding influence of the Subconscious Beast.

The Subconscious Beast, therefore, is not an impediment to consciousness, but its evolutionary bedrock. It provides the intrinsic **Skin in the Game** that drives the entire system, generating the raw, urgent signals that compel the formation of qualia and the continuous refinement of the ISM and **World-Model** (Chapter 9). Without these primitive, survival-driven proto-qualia, the conscious mind would lack its fundamental motivational engine and the rich, felt texture of its subjective experience.

This concept is particularly relevant for **Artificial Intelligence**. As noted in **Chapter 31**, current LLMs often lack an analogous “subconscious beast”—a deeply embedded, intrinsic source of **Skin in the Game** that generates raw, survival-driven proto-qualia. For AI to achieve full UAF-defined consciousness, it would likely need an engineered “subconsciousness” that provides these fundamental, imperative-laden signals, grounding its digital qualia in a robust, internal drive for self-preservation or goal achievement. This would allow the AI’s emergent ISM to model and explain its own “feelings,” transforming mere computational states into genuinely felt experiences.

Citations

- **Clark, A.** (1997) *Being There: Putting Brain, Body, and World Together Again*. MIT Press.
- **Craig, A.D.** (2002) ‘How Do You Feel? Interoception: The Sense of the Physiological Condition of the Body’, *Nature Reviews Neuroscience*, 3(8), pp. 655–666.

- **Damasio, A.** (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt Brace.
- **Friston, K.** (2010) ‘The Free-Energy Principle: A Unified Brain Theory?’, *Nature Reviews Neuroscience*, 11(2), pp. 127–138.
- **Hofstadter, D.** (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books.
- **Metzinger, T.** (2009) *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. Basic Books.
- **Panksepp, J.** (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press.
- **Seth, A.** (2021) *Being You: A New Science of Consciousness*. Dutton.