# Chapter 8 — Qualia

## Contents



### Chapter 8: Qualia: The Universe's Simplified Truth

We've established that the brain constructs a simplified internal model of reality and itself; but what about the raw, undeniable feel of experience—the searing pain of a burn, the vibrant hue of a sunset, the bitter taste of coffee? For centuries, this *"what it's like"* aspect, often termed **qualia**, has been the stubborn core of the **"Hard Problem of Consciousness"** (Chalmers, 1996). How can mere neural firings, a collection of electrochemical signals, give rise to such rich, subjective, and seemingly irreducible sensations? Many theories either dismiss qualia as **epiphenomenal**—a byproduct with no causal role (Dennett, 1991)—or declare them an unsolvable mystery, a fundamental gap in our understanding (McGinn, 1989). *Others, like **panpsychism**, suggest that consciousness—and thus qualia—might be a fundamental property of matter itself (Goff, 2019), though this raises its own explanatory challenges. Meanwhile, **illusionist theories** argue that qualia don't exist as we intuit them, but are instead a kind of "user illusion" constructed by the brain (Dennett, 2017).*

**Useful Approximations Framework (UAF)** offers a different perspective: qualia are not an anomaly, but a **necessary functional fiction** — the brains own simplified truth.

The brain is a collection of neurons, an information-processing system evolved to maximize the likelihood of surviving and passing on the genes it holds. Its main feature is the ability to learn **representations of reality**. These representations are not exact truths. In fact, as we explored in **Chapter 1**, there does not seem to be any way for the system to access absolute truths. The brain just gathers noisy sensory data and constructs **approximate models** that fit the sensory data. *This process is **Bayesian inference in action** — the brain updates its prior beliefs about the world based on new evidence, balancing precision and uncertainty (Friston, 2010; Clark, 2013).* These models can take the form of visual objects like a red apple, a chair, or a house; auditory objects like an explosion, splash, or thump; or objects recognized by our touch, smell, or taste sensors. The recognized basic objects are not reality itself, but **representations or approximations of reality**. The chair is actually a complex collection of atoms and molecules, a formation carved by a human or a machine. The *"chair"* is just a helpful simplification of this reality that allows humans to act efficiently. *This simplification is **affordance-based** — the brain doesn't model the chair's atomic structure but its functional properties: "Can I sit on this? Can I move it?" (Gibson, 1977).*

These approximations are what our **self-model, world-model, qualia, and consciousness** are also about. They are not the reality itself, but a useful approximation of it. But why do these approximations *feel* like anything at all? Why isn't the brain simply processing data packets, like a computer? *After all, a thermostat doesn't "feel" heat — it just registers temperature and triggers a response. So why do we? (Nagel, 1974).* This is where the **functional necessity of qualia** becomes paramount.

Imagine a sophisticated computer system designed to manage a complex factory. It receives vast amounts of data: temperature readings, pressure levels, machine vibrations, inventory counts. If this system were to present all this raw data to its human operator, the operator would instantly succumb to **Computational Paralysis**. Instead, the system presents a simplified, intuitive *"CEO's Dashboard"*: a green light for optimal performance, a flashing red light for a critical malfunction, a yellow bar indicating low inventory. The operator doesn't need to see the millions of data points; they need a **compressed,**

**high-level summary** that is immediately understandable and actionable. *This is **data visualization as a form of qualia** — a machine's "felt" experience of its own state, albeit in a non-conscious form (Piccinini, 2015).*

**Qualia** are precisely this *"CEO's Dashboard"* for the brain. They are the ultimate compression of complex internal and external information into a **directly usable, self-validating signal**. A purely abstract, non-felt signal—a numerical value representing *"tissue damage"* or a data packet labeled *"wavelength 650nm"*—would require another system, or another layer of processing, to interpret its meaning to the system itself. This would lead to an **infinite regress of interpretation**, ultimately resulting in computational paralysis. This is the **symbol grounding problem** in reverse: how does the brain anchor its symbols in meaning without an infinite chain of interpreters? (Harnad, 1990). The 'feeling' is the interpretation. It is the **Subjective Closure**: the point at which the information is so perfectly compressed and presented that it requires no further processing to be understood by the system experiencing it. These conscious qualia are built upon more primitive 'proto-qualia' generated by the 'Subconscious Beast' (Chapter 11), which provide the raw, urgent signals of survival.

It is the *"simplified truth"* because it is the most direct, undeniable, and **functionally useful** truth the system has about its own state and its interaction with the world. *In this sense, qualia are **self-evident** — they don't just represent information; they* are* the information, experienced directly (Searle, 1992).*

Consider the searing pain of a burn. Pain, for instance, is the approximation of any complex neural activation pattern that causes a complex biochemical reaction in our body and subconsciousness, designed to protect our body. This subconscious reaction gains control of our body while our conscious neocortex loses control to some extent, depending on how strong the pain is. The **pain qualia** represents this complexity to ourselves in a way that makes sense so well that we do not need to dig deeper to understand what the signal is about. *This is **affective realism**—the brain doesn't just detect damage; it* constructs* the experience of pain as a compelling, urgent signal that demands attention (Wager and Lindquist, 2016).* Our brain has learned the **perfect representation** of pain sensations for us to act efficiently when we sense it.

This *"feeling"* is not merely an informational display; it carries an **inherent imperative**. The pain of a burn doesn't just inform you of tissue damage; it **compels** you to withdraw your hand. This is the **Causal Efficacy (Q→Action)** of qualia. They are not epiphenomenal byproducts; they are **powerful, high-bandwidth signals** that directly drive action. The pain is the representation of what is happening in the brain. Your consciousness is losing control over your hand. Your subconscious primitive brain is already signaling your hand to retract and move away. The pain represents this signaling without the detailed understanding of what is happening under the hood. The pain also represents the input that your ISM can interpret as an input that suggests that moving your hand is a good idea now. Your ISM, the virtual machine that takes inputs, integrates it with your current state and memories to come up with a free will to produce the output it wants, is the learned representation of yourself and it has learned that this input has such a meaning for this machine. If your current internal state has a challenge to see how much of this pain you can take, then the virtual machine represents this competition. Your subconscious is getting stronger as it senses potential damage while your consciousness is losing control from fatigue.

*This aligns with **enactive cognition** — qualia aren't just passive experiences but active guides for behavior (Varela et al., 1991).* A *"feeling"* like pain or pleasure is an incredibly efficient signal, conveying immense information — **location, intensity, urgency, threat, or reward** - in an instant. It bypasses layers of cognitive deliberation, triggering rapid, often subconscious, but **causally effective responses**.

Qualia, therefore, are the **phenomenal flavors** of our internal models. Just as the flavor of coffee is a simplified, subjective experience of complex chemical interactions, the feeling of *"red"* is the simplified, subjective experience of complex neural interactions responding to specific wavelengths of light. *This is **sensory compression**—the brain reduces high-dimensional sensory data into low-dimensional, experiential qualia (Barlow, 1961).* The vibrant hue of a sunset is not the objective reality of photons at specific frequencies; it is your brain's **highly optimized, functionally essential interpretation** of that information. It's the *"simplified truth"* that allows you to distinguish ripe fruit from unripe, or a dangerous predator from a harmless shadow. *This optimization is **ecologically rational**—the brain prioritizes information that matters for survival, not metaphysical accuracy (Gigerenzer, 2000).*

These simplified truths are crucial for the continuous refinement of both our **Internal Self-Model (ISM)** and our **World-Model**. The feeling of hunger (a qualia) updates your ISM about your body's

energy levels, prompting you to seek food. The feeling of warmth (a qualia) updates your World-Model about environmental conditions, guiding you toward shelter or away from a heat source. *This is **interoceptive inference**—the brain predicts and updates its model of the body's state based on qualia (Seth, 2013).* Qualia are integral components in the **feedback loops** that drive learning and adaptation, constantly informing the system about the success or failure of its predictions and actions. They provide the **immediate, visceral feedback** necessary for the brain to minimize prediction error and refine its approximations of reality. *Without qualia, the brain would be like a ship without a rudder—awash in data but unable to steer (Friston, 2018).*

Consider the profound implications of this **functionalist view**. The *"Hard Problem"* of consciousness, which asks why anything feels like anything at all, is resolved not by discovering some mysterious non-physical property, but by understanding the **computational necessity of subjective experience**. *This is **neurofunctionalism**—qualia are what they* do*, not what they* are* (Lewis, 1972).* Qualia are the brain's ingenious solution to the problem of **internal interpretability** and **efficient action** in a world of overwhelming complexity. They are the ultimate compression of complex information, enabling the system to *"know"* and *"act"* without succumbing to paralysis.

*But this raises a critical question: **Could artificial systems ever have qualia?** If qualia are functionally necessary for biological systems, might they also emerge in sufficiently complex AI? (Chalmers, 2010). Some argue that **integrated information theory (IIT)** suggests that any system with high Φ (phi)—a measure of information integration—could have a form of consciousness (Tononi, 2008). Others, like **global workspace theory (GWT)**, propose that qualia arise from broadcasted information in a brain-like architecture (Dehaene, 2014). Yet without **embodied, affective grounding**, it's unclear whether AI could ever* feel* pain or see *red* the way we do (Harnad, 1990).*

In essence, qualia are not a luxury or a philosophical enigma; they are the **indispensable, simplified truths** that allow a complex, finite system to achieve **subjective closure, causal efficacy, and efficient agency**. They are the very reason why our internal models are not just abstract data structures, but a **lived, felt reality**. *They are the brain's way of **making meaning**—transforming raw data into something that matters (Frankl, 1946).* They are the universe's way of making its own overwhelming complexity comprehensible to the conscious systems that emerge within it.

---

**Key References Cited *(Harvard Style, Alphabetical)***


- **Barlow, H.B.** (1961) 'Possible Principles Underlying the Transformation of Sensory Messages', *Sensory Communication*, pp. 217–234.
- **Botvinick, M. and Cohen, J.** (1998) 'Rubber Hands "Feel" Touch That Eyes See', *Nature*, 391(6669), pp. 756.
- **Chalmers, D.J.** (1996) *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press.
- **Chalmers, D.J.** (2010) 'The Singularity: A Philosophical Analysis', *Journal of Consciousness Studies*, 17(9–10), pp. 7–65.
- **Clark, A.** (2013) *Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science*, *Behavioral and Brain Sciences*, 36(3), pp. 181–204.
- **Dennett, D.C.** (1991) *Consciousness Explained.* Little, Brown and Company.
- **Dennett, D.C.** (2017) *From Bacteria to Bach and Back: The Evolution of Minds.* W.W. Norton & Company.
- **Dehaene, S.** (2014) *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts.* Viking.
- **Friston, K.** (2010) 'The Free-Energy Principle: A Unified Brain Theory?', *Nature Reviews Neuroscience*, 11(2), pp. 127–138.
- **Friston, K.** (2018) 'What Does the Free-Energy Principle Tell Us About the Brain?' *Neural Computation*, 30(1), pp. 1–4.
- **Frankl, V.E.** (1946) *Man's Search for Meaning.* Beacon Press.
- **Gibson, J.J.** (1977) *The Ecological Approach to Visual Perception.* Houghton Mifflin.
- **Gigerenzer, G.** (2000) *Adaptive Thinking: Rationality in the Real World.* Oxford University Press.
- **Goff, P.** (2019) *Galileo's Error: Foundations for a New Science of Consciousness.* Pantheon

Books.
- **Harnad, S.** (1990) 'The Symbol Grounding Problem', *Physica D: Nonlinear Phenomena*, 42(1–3), pp. 335–346.
- **Lewis, D.** (1972) 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, 50(3), pp. 249–258.
- **McGinn, C.** (1989) 'Can We Solve the Mind-Body Problem?', *Mind*, 98(391), pp. 349–366.
- **Nagel, T.** (1974) 'What Is It Like to Be a Bat?', *Philosophical Review*, 83(4), pp. 435–450.
- **Nesse, R.M.** (2005) 'Evolutionary Origins and Functions of Pain', *Pain Forum*, 14(2), pp. 86–93.
- **Noë, A.** (2004) *Action in Perception*. MIT Press.
- **Piccinini, G.** (2015) *Physical Computation: A Mechanistic Account*. Oxford University Press.
- **Searle, J.R.** (1992) *The Rediscovery of the Mind*. MIT Press.
- **Seth, A.K.** (2013) 'Interoceptive Inference, Emotion, and the Embodied Self', *Trends in Cognitive Sciences*, 17(11), pp. 565–573.
- **Tononi, G.** (2008) 'Consciousness as Integrated Information: A Provisional Manifesto', *Biological Bulletin*, 215(3), pp. 216–242.
- **Varela, F.J. et al.** (1991) *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
- **Wager, T.D. and Lindquist, K.A.** (2016) 'An fMRI-Based Neurologic Signature of Physical Pain', *New England Journal of Medicine*, 374(13), pp. 1255–1265.

---

**Ideas**

Can we try to construct the progressive complexity of making that sophisticated computer system in the factory to "feel"? The data it receives does not feel like anything at all to the computer. It is just electron movement. Once that data enters the data processing pipeline, it just gets stored to the database and presented in the dashboard. No reason for feeling anything yet. But what if that processor would attempt to predict the future data points? It would need to start to understand the dynamics of the factory. How the datapoints are interconnected and how there are repeating patterns especially in the 24 hour, 7 day, 30 day and 365 day intervals. Without precise understanding of all the details in the factory, down to the quantum level interactions of every atom that forms the factory, the neural network that learns this prediction is forced to come up with abstract representations in its network of information processing. It has no understanding of the humans and their daily and weekly routines. No understanding of Earths yearly cycles and weather patterns. It just observes and makes its own abstract simplified "truths" to understand the observed patterns.

But without control over the system, just being a passive observer, it cannot gain information about itself. It cannot learn that its own prediction about the temperature change in some floor of the factory will cause a heater to turn on.

Obviously that data collection and prediction will get linked to the heaters eventually. The engineers of the factory will find it useful to use the data to control the factory. That was probably the reason for the predictive neural network in the first place. So now the system observes that it is part of the cycle. Temperature goes up, it adapts to this to learn to predict it, its learned predictions prevent the temperature from going up. It is part of the loop. But it still is not conscious. Its self-model and world-model are still very primitive. It has something very basic that we could call qualia, the self-model, the world-model and free will. But it does not form an episodic memory about its experiences. It does not learn about its existence in time.

So what if the system would also write a log telling everything it has observed. This log would contain the history of what has happened in the factory from day one. Sensors and heaters have broken. Weather anomalies have caused issues. Demand and supply of the factory has evolved over time. This is where the systems understanding of reality takes a step forward. The system can observe itself adjusting to these unexpeced changes. It can learn about how it intially had difficulty recognizing the weekly, monthly or yearly patterns, but over time, became more fluent in predicting them and keeping the factory stable. But it needs a much more complicated neural network to have access to all this data. It needs to be able to integrate all the data from the history and present into its predictions. An optimization challenge for the engineering team. Through this full history of everything in its existence, it is able to learn how it reacts to new situations, what its learning algorithm is trying to avoid and where it is trying to go.

It will also have the ability to learn to represent this learning behavior that it observes and learns from its history. It learns an approximation of itself as a learning system. This approximation might say something like "I seem to be constantly trying to learn ways to predict sensor and actuator damage and find ways to adapt to those damages as fast as possible. I seem to *hate* this inbalance."