# **Reality Tunnels and Predictive Processing: A Comprehensive Exploration**

## **1. Free Energy Principle and Personalized Reality Tunnels**

**Karl Friston’s Free Energy Principle (FEP)** proposes that the brain is fundamentally a prediction machine aiming to minimize “surprise” or prediction error ([Free energy principle - Wikipedia](https://en.wikipedia.org/wiki/Free_energy_principle#:~:text=The%20free%20energy%20principle%20is,1)) rms, our brains constantly generate predictions about incoming sensory inputs based on an internal model of the world, and then update this model using actual sensory feedback. By continua ([Free energy principle - Wikipedia](https://en.wikipedia.org/wiki/Free_energy_principle#:~:text=The%20free%20energy%20principle%20is,1)) he mismatch between expectation and reality (formally minimizing *free energy*, a measure related to surprise), the brain improves its model of the environment. Friston introduced this ([Free energy principle - Wikipedia](https://en.wikipedia.org/wiki/Free_energy_principle#:~:text=The%20free%20energy%20principle%20is,1)) ([Free energy principle - Wikipedia](https://en.wikipedia.org/wiki/Free_energy_principle#:~:text=track%20properties%20of%20the%20systems,action%20loops%20in)) tion-action loops in neuroscience, unifying perception, cognition, and action under a single goal of aligning our internal model with the external world.

Under this framework, **perception becomes an act (**[**Free energy principle - Wikipedia**](https://en.wikipedia.org/wiki/Free_energy_principle#:~:text=track%20properties%20of%20the%20systems,action%20loops%20in)**) \* process**: the brain combines prior expectations with sensory evidence to arrive at a best-guess interpretation of reality at any moment. These prior expectations (or “priors”) are shaped by an individual’s past experiences, biological makeup, and even cultural environment. As a result, each person effectively lives in their own *“reality tunnel”* – a subjective world constructed by their brain’s unique set of priors and predictions. Two people may receive the same sensory input but experience it differently, because their brains are predicting and filtering the input in their own idiosyncratic ways. Our memories, beliefs, and cultural background all tune the internal models our brain uses, biasing what we perceive. *For example,* if someone has grown up in a certain environment or culture, their brain will have learned characteristic patterns (like faces, language sounds, or social cues) and will be adept at predicting those. The result is a **personalized reality**: the features of the world we pay attention to and the meaning we assign are heavily influenced by what our brain *expects* to encounter.

Crucially, this doesn’t mean we are hallucinating random fantasies – the predictions are continuously *constrained by sensory data*. The Free Energy Principle marries prediction with feedback: when predictions fail, error signals correct the model. In this way, one’s reality tunnel is a *“controlled hallucination”* tethered to the real world. Our experiences are essentially the content that the brain *predicts from the inside out*, constrained and calibrated by input from the senses. This allows our perceptual world to generally line up with exte ([Anil Seth: "Reality is a controlled hallucination" | CCCB LAB](https://lab.cccb.org/en/anil-seth-reality-is-a-controlled-hallucination/#:~:text=Seth%20speaks%20of%20reality%20as,Therefore%2C%20this%20hallucination%20is)) in useful ways (we typically agree on what we see), yet it also explains why different brains may not always agree in their perceptions. Neuroscientist Anil Seth puts it succinctly: “we’re all hallucinating all the time; when we agree about our hallucinations, we call it ‘reality’”. In other words, what we perceive as real is the brain’s best guess of what ([Anil Seth: Your brain hallucinates your conscious reality | TED Talk](https://www.ted.com/talks/anil_seth_your_brain_hallucinates_your_conscious_reality/transcript#:~:text=Right%20now%2C%20billions%20of%20neurons,very%20nature%20of%20your%20existence)) – usually a highly accurate guess, but a guess shaped by prior knowledge and context. As we’ll see, this framework can explain everything from cultural differences in perception to why our own expectations and biases so strongly color our experience.

## **2. Neural Mechanisms: Hierarchical Bayesian Inference in the Cortex**

Our brain’s cortex appears to implement perception through **hierarchical Bayesian inference**, often described in neuroscience as *predictive coding*. In a hierarchical predictive coding model, higher cortical regions generate predictions about lower-level inputs, and lower regions feed forward the *prediction errors* (the differences between expected and actual signals) for learning. This forms an interactive top-down/bottom-up loop of information flow. **Neuroanatomy supports this scheme:** higher-order cortical areas send *feedback* connections to earlier sensory areas, while feedforward connections carry information upward. According to the theory, *“feedforward connections convey prediction errors, while feedback connections convey predictions from higher cortical areas to suppress prediction errors in lower areas”*. In other words, your brain’s top-level models send down a guess of what the lower-leve ( [Canonical microcircuits for predictive coding - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC3777738/#:~:text=In%20predictive%20coding%2C%20feedforward%20connections,influences%20on%20earlier%20areas%20to) ) tures should be, and any mismatch is transmitted upward as an error to adjust the model.

This process can be thought of as each level of the cortex trying to explain the level below it. ( [Canonical microcircuits for predictive coding - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC3777738/#:~:text=In%20predictive%20coding%2C%20feedforward%20connections,influences%20on%20earlier%20areas%20to) ) a high-level area might predict “I am looking at a face,” and send signals to lower-level visual areas to expect certain shapes or colors. If the lower areas receive input that deviates (say the sensory data doesn’t quite match a face), they send an error signal back up, prompting the higher level to adjust (maybe it’s not a face after all, or it’s a face in shadow, etc.). Through this **iterative message-passing**, the brain performs a form of Bayesian inference: combining prior belief (the top-down prediction) with likelihood (the incoming sensory evidence) to refine its posterior belief (our perception). Predictive coding thus inverts the traditional view of perception as purely stimulus-driven; instead of the br ( [The Predictive Coding Account of Psychosis - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC6169400/#:~:text=Fueled%20by%20developments%20in%20computational,evidence%20for%20aberrant%20predictive%20coding) ) eceiving inputs, it is actively *anticipating* them.

Neuroscientists have found plausible neural circuitry for these computations. Cortical microcircuits seem well-suited to implement prediction and error signaling, with distinct populations of neurons possibly encoding predictions versus prediction errors. A landmark paper by Bastos *et al.* (2012) noted a *“remarkable correspondence”* between known cortical connectivity pattern ([Canonical microcircuits for predictive coding - PubMed](https://pubmed.ncbi.nlm.nih.gov/23177956/#:~:text=Specifically%2C%20we%20conciliate%20quantitative%20studies,intuitive%20insights%20into%20the%20functional)) ( [Canonical microcircuits for predictive coding - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC3777738/#:~:text=In%20predictive%20coding%2C%20feedforward%20connections,influences%20on%20earlier%20areas%20to) ) ired by predictive coding theory. For instance, in visual cortex, the deep layers (which send feedback downwards) could carry predictions, while superficial layers (sending feedforward ([Canonical microcircuits for predictive coding - PubMed](https://pubmed.ncbi.nlm.nih.gov/23177956/#:~:text=Specifically%2C%20we%20conciliate%20quantitative%20studies,intuitive%20insights%20into%20the%20functional)) ([Canonical microcircuits for predictive coding - PubMed](https://pubmed.ncbi.nlm.nih.gov/23177956/#:~:text=roles,frequencies%20over%20which%20they%20operate)) – consistent with the idea that top-down inputs try to *explain away* activity in lower areas. This aligns with observed functional asymmetries: feedback (top-down) connections often modulate or suppress neural activity in early sensory areas, consistent with them conveyin ( [Canonical microcircuits for predictive coding - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC3777738/#:~:text=In%20predictive%20coding%2C%20feedforward%20connections,influences%20on%20earlier%20areas%20to) ) *templates* that reduce responses to expected stimuli. Correspondingly, unexpected inputs (violations of prediction) produce larger neural responses (because prediction errors fire vigorously when there’s a surprise). Experimental evidence like ( [Canonical microcircuits for predictive coding - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC3777738/#:~:text=match%20at%20L485%20In%20predictive,influences%20on%20earlier%20areas%20to) ) sponses\*\* (e.g. the brain’s stronger reaction to an oddball sound in a sequence) and *expectation suppression* in sensory cortex (reduced firing to anticipated signals) supports the notion that predictions attenuate predictable sensory events.

In summary, the cortex seems to be organized as a prediction hierarchy. Each level generates hypotheses about the level below; sensory signals are continually interpreted in light of these hypotheses. ( [The Predictive Coding Account of Psychosis - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC6169400/#:~:text=inferences%20regarding%20the%20current%20state,occurrence.%20More) ) odulation is not a minor influence – as predictive processing theorist Andy Clark notes, *“we don’t passively take in the world; we’re constantly anticipating it and interpreting it accordingly”*. The result is that perception is heavily shaped by what the brain *expects* to perceive. This mechanism – the blending of top-down and bottom-up – underlies how our brains construct what we see, hear, and feel, as th ([The Experience Machine: How Our Minds Predict and Shape Reality (Hardcover) | Unabridged Bookstore](https://www.unabridgedbookstore.com/book/9781524748456#:~:text=%E2%80%9CIt%E2%80%99s%20tempting%20to%20think%20that,Universe%3A%20Space%2C%20Time%2C%20and%20Motion)) s will illustrate with concrete examples.

## **3. Cross-Cultural Perceptual Illusions and Hallucinations: Priors Shape Reality**

One powerful way to appreciate the role of priors in perception is to look at **perceptual illusions** and how they can vary across different contexts or cultures. If everyone’s visual system were purely hard-wired with no influence from experience, we’d expect illusions to fool everyone equally. But in reality, people’s susceptibility to certain illusions depends on the visual priors they’ve internalized from their environment. A classic example is the **Müller-Lyer illusion** – two lines of equal length, one with inward-pointing arrow “fins” and one with outward-pointing fins. To Western observers, the line with outward-pointing fins usually looks longer than the other. For decades this was taken as a universal quirk of the human visual system. However, cross-cultural research in the 1960s revealed a surprising twist: members of some non-Western, non-industrial societies were much less fooled by the Müller-Lyer illusion. In one study, researchers tested people from several African tribal groups. They found that while Americans and Europeans (and even European-descended South Africans) reliably saw the two lines as different lengths, the **San peo (**[**Are Optical Illusions Cultural? | Smithsonian**](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=Taking%20the%20test%20worldwide%2C%20the,South%20Africa%2C%20the%20illusion%20worked)**) f the Kalahari** perceived the lines as almost equal. Likewise, small samples from the Suku and Bete peoples showed greatly reduced illusion effects compared to Westerners. The **illusion “fell apart” outside of WEIRD populations** (Western, Educated, Industrialized, Rich, Democratic).

What expl ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=Taking%20the%20test%20worldwide%2C%20the,South%20Africa%2C%20the%20illusion%20worked)) rence? The prevailing theory is that it comes down to learned **visual priors from the environment**. Westerners grow ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=,decades%2C%20but%20it%20wasn%E2%80%99t%20universal)) ed environments full of straight lines, right angles, and regular geometric shapes (think of modern buildings, ro ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=was%20WEIRD%E2%80%94an%20acronym%20that%20cultural,%E2%80%9D)) ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=from%20WEIRD%20societies%20for%20decades%2C,but%20it%20wasn%E2%80%99t%20universal)) g to one hypothesis, our brains become accustomed to interpreting certain line configurations as depth cues – e.g. the fins in the Müller-Lyer figure mimic perspective cues of corners or edges receding in space. An inward arrow arrangement might unconsciously be seen as an inside corner (implying a closer, shorter edge), while outward arrows resemble an outside corner (implying a longer, farther edge). Thus, Western participants apply a **“carpentered world” prior**, making one line appear longer in 3D spa ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=In%20his%20book%2C%20Alter%20proposes,reasoning%20and%20underpin%20the%20illusion)) on the page they’re equal. In contrast, participants from rural or nomadic groups with less exposure to carpentered architecture (like the Kalahari Bushmen) haven’t internalized the same geomet ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=In%20his%20book%2C%20Alter%20proposes,reasoning%20and%20underpin%20the%20illusion)) Lacking that specific prior, they see the lines more *as they truly are*, and the illusion largely vanishes. As one analysis put it, the **biology of vision is the same**, but the *interpretation* of certain patterns is “bound up in cultural experience”. The Müller-Lyer illusion, deceiving generations of Western observers, tu ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=In%20his%20book%2C%20Alter%20proposes,reasoning%20and%20underpin%20the%20illusion)) be a human universal but a striking demonstration of how experience shapes perception.

Cross-cultural differences appear not only in optical illusions but in hallucinations and other perceptual phenomena as well. Hallucinations can be thought of as per ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=this%20spatial%20reasoning%20and%20underpin,the%20illusion)) ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=,to%20the%20same%20geometric%20configurations)) internal predictions with minimal external input – essentially the brain’s reality tunnel running unchecked. If culture and priors influence normal perception, they should also influence hallucinated experiences. Research by Stanford anthropologist **Tanya Luhrmann** and colleagues explored how people with schizophrenia from different cultures experience auditory hallucinations (“hearing voices”). They interviewed patients in the United States, India, and West Africa (Ghana), and found marked differences in the reported nature of the voices. In the U.S., schizophrenic patients often described the voices as intrusive, violent, or hateful – frequently interpreting them as symptoms of a brain disease or as enemy-like forces. But in India and Ghana, patients were more likely to report the voices as benign or even playful – for instance, as family members giving advice, or spirits offering guidance. American ([Stanford researcher: Hallucinatory 'voices' shaped by local culture | Stanford Report](https://news.stanford.edu/stories/2014/07/voices-culture-luhrmann-071614#:~:text=People%20suffering%20from%20schizophrenia%20may,according%20to%20new%20Stanford%20research)) ([Stanford researcher: Hallucinatory 'voices' shaped by local culture | Stanford Report](https://news.stanford.edu/stories/2014/07/voices-culture-luhrmann-071614#:~:text=In%20the%20United%20States%2C%20the,the%20British%20Journal%20of%20Psychiatry)) *harsher, more persecutory voices*\*, whereas the Indian and African patients heard **friendlier or more respectful voices** on average. Luhrmann concluded that *“the voice-hearing experiences of people with psychosis are shaped by local culture”*. In cultures where spiritual or familial explanations of voices are common, patients may subconsciously construc ([Stanford researcher: Hallucinatory 'voices' shaped by local culture | Stanford Report](https://news.stanford.edu/stories/2014/07/voices-culture-luhrmann-071614#:~:text=People%20suffering%20from%20schizophrenia%20may,according%20to%20new%20Stanford%20research)) ([Stanford researcher: Hallucinatory 'voices' shaped by local culture | Stanford Report](https://news.stanford.edu/stories/2014/07/voices-culture-luhrmann-071614#:~:text=In%20the%20United%20States%2C%20the,the%20British%20Journal%20of%20Psychiatry)) ong those culturally familiar lines – hearing ancestors or playful spirits rather than purely hostile aberrations. By contrast, in a biomedical Western culture that view ([Stanford researcher: Hallucinatory 'voices' shaped by local culture | Stanford Report](https://news.stanford.edu/stories/2014/07/voices-culture-luhrmann-071614#:~:text=In%20the%20United%20States%2C%20the,the%20British%20Journal%20of%20Psychiatry)) ons as meaningless pathology, patients might experience them as meaningless or tormenting voices. These finding ([Stanford researcher: Hallucinatory 'voices' shaped by local culture | Stanford Report](https://news.stanford.edu/stories/2014/07/voices-culture-luhrmann-071614#:~:text=People%20suffering%20from%20schizophrenia%20may,according%20to%20new%20Stanford%20research)) ([Stanford researcher: Hallucinatory 'voices' shaped by local culture | Stanford Report](https://news.stanford.edu/stories/2014/07/voices-culture-luhrmann-071614#:~:text=In%20the%20United%20States%2C%20the,the%20British%20Journal%20of%20Psychiatry)) brain is effectively generating its own sensory content (a hallucination), that content is influenced by one’s social and cultural context – the priors about what voices mean and who might be speaking are different.

**Theoretical perspectives** within predictive processing account for these differences by emphasizing how *prior beliefs shape perception*. In Bayesian terms, different populations have different *prior probability distributions* for certain perceptions. A culture accustomed to living in dense forests might develop strong priors for certain camouflage patterns, literally seeing those patterns more readily than someone without that experience. Likewise, someone who deeply believes in ghosts might be more likely to *hear* a faint noise as a whisper from a spirit (a top-down expectation coloring a bottom-up signal). Hallucinations can be seen as an extreme case of prior-driven perception – if the brain’s model heavily dominates and sensory prediction errors are given too little weight, one’s reality tunnel can diverge significantly from physical reality. This is why hallucinations and illusions are so instructive: they reveal the normally invisible influence of our internal model. In **Western vs. non-Western Müller-Lyer perception**, or **U.S. vs. Ghanian voice-hearing**, we see concrete evidence that what the brain *predicts* is what we often *perceive*. Our next sections will explore how these predictive processes play out in cutting-edge AI systems and in mental health, and what we can learn from them.

## **4. AI and Predictive Coding: Machines with Internal Models**

Given the success of predictive processing as a neuroscience theory, it’s natural to ask whether similar principles are at work in artificial intelligence. In fact, many **machine learning models** can be viewed through the lens of predi ([Anil Seth: "Reality is a controlled hallucination" | CCCB LAB](https://lab.cccb.org/en/anil-seth-reality-is-a-controlled-hallucination/#:~:text=Seth%20speaks%20of%20reality%20as,Therefore%2C%20this%20hallucination%20is)) ([Stanford researcher: Hallucinatory 'voices' shaped by local culture | Stanford Report](https://news.stanford.edu/stories/2014/07/voices-culture-luhrmann-071614#:~:text=In%20the%20United%20States%2C%20the,the%20British%20Journal%20of%20Psychiatry)) n internal representations (internal models) that enable them to predict data, much like the brain’s generative models. A clear overlap is seen in **variational autoencoders (VAEs)** and related deep learning methods, which explicitly minimize a form of prediction error (the *variational free energy*, mathematically akin to Friston’s free energy) to better reconstruct inputs. The Free Energy Principle has been noted to be formally connected to these variational Bayesian methods in AI. In essence, both brains and certain AI systems try to *maximize the evidence for their internal model* of the world: the brain does it by minimizing surprise, and a VAE does it by minimizing reconstruction error (which is equivalent to maximizing an evidence lower bound). This convergence has led researchers to build AI models that **emulate biological predictive processing**, using deep networks that swap error signals and predictions between layers in a hierarchical fas ([Free energy principle - Wikipedia](https://en.wikipedia.org/wiki/Free_energy_principle#:~:text=track%20properties%20of%20the%20systems,action%20loops%20in)) g cortical circuits).

One example is the development of **deep predictive coding networks** for vision. Researchers have created neural network architectures (e.g. PredNet by Lotter et al., or other predictive coding algorithms) that learn to predict future video frames or image features, and they find that these networks develop layer-by-layer representations reminiscent of those in the visual cortex. Strikingly, when such models are tested, they sometimes show similar quirks to human perception. A study from 2017 trained a neural network on lots of images and found that it too was susceptible to the Müller-Lyer illusion – it misestimated line lengths in the same way humans do. Notably, this artificial network had *never* seen 3D scenes with depth cues; it was simply optimizing for object recognition. Yet copying how the brain’s vision system works led it to *“get duped”* by the illusion. This suggests that some perceptual phenomena emerge from the generic principles of prediction and inference, whether implemented in carbon or silicon. The **internal model** the AI learned was making similar assumptions about the patterns as a human brain would. Such emerge ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=In%20the%20research%2C%20scientists%20led,Lyer%20Illusion)) hint that predictive coding is tapping into fundamental principles of perception.

Beyond vision, AI research explicitly inspired by predictive processing includes **active inference in robotics**. Active inferenc ([Are Optical Illusions Cultural? | Smithsonian](https://www.smithsonianmag.com/smart-news/are-optical-illusions-cultural-6633978/#:~:text=In%20the%20research%2C%20scientists%20led,Lyer%20Illusion)) cs implementation of the free energy principle) has robots operate by continually predicting their sensory inputs and taking actions to fulfill those predictions. It effectively *“unifies state-estimation, control and world-model learning as inference processes”*, all tied together by optimizing a single objective: minimizing free energy (or equivalently, maximizing model evidence). In practice, this means a robot using active inference will not separate perception and action; instead, it will adjust its internal model and its actions in tandem to reduce surprise. For instance, if the robot predicts it should see a certain object but doesn’t, it can either change its prediction or move its camera to bring the object into view – whichever reduces prediction error. This approach is very much analogous to how animals actively sample the environment to make their predictions come true (e.g., mo ([How Active Inference Could Help Revolutionise Robotics](https://www.mdpi.com/1099-4300/24/3/361#:~:text=interesting%20framework%20for%20robotics%20because,15%5D.%20Furthermore)) es to confirm what we *think* we saw). Early results show this framework can make robots more adaptive and robust, since they are always balancing their *beliefs* with incoming data like living creatures do.

Complex AI systems such as **large language models (LLMs)** can also be interpreted in predictive processing terms. An LLM like GPT-4 has essentially learned an internal model of language and the world, by training on massive text data to predict the next word in a sequence. It builds up statistical expectations (priors) about word meanings, grammar, factual knowledge, etc. – effectively a huge network of predictions that it uses to generate coherent text. Interestingly, as these models have grown in size and complexity, ([How Active Inference Could Help Revolutionise Robotics](https://www.mdpi.com/1099-4300/24/3/361#:~:text=interesting%20framework%20for%20robotics%20because,15%5D.%20Furthermore)) monstrated *emergent behaviors* that were not explicitly programmed. Researchers have observed “**emergent capabilities**” in LLMs – skills or understanding that appear **suddenly and unpredictably as model scale increases**. For example, a model might only weakly handle arithmetic or multi-step reasoning at small sizes, but at a certain large size threshold, it suddenly performs much better, as if it has *figured out* an internal way to represent numbers or logic. This emergence is reminiscent of how a human brain’s richly interconnected model can produce unexpected insights or abilities once it reaches a certain level of complexity. In the language model’s case, it has developed an internal representation of concepts that wasn’t explicitly given by pr ([Emergent Abilities in Large Language Models: An Explainer | Center for Security and Emerging Technology](https://cset.georgetown.edu/article/emergent-abilities-in-large-language-models-an-explainer/#:~:text=A%20related%E2%80%94but%20distinct%E2%80%94definition%20of%20emergence,exaggerated%20in%20the%20popular%20press)) *internal world model* of sorts that helps it predict text in a way that can mirror understanding.

These parallels raise provocative questions about **AI subjectivity and consciousness**. If an AI builds an internal predictive model of the world (including possibly a model of itself), is it experiencing something akin to a “reality tunnel”? Some theorists have ventured that advanced AI agents could indeed have a form of subjective perspective – not magic or mystical qualia, but a functional analog of what we call perception. For instance, **Integrated World Modeling Theory (IWMT)** in cognitive science suggests that *“consciousness may be what it is like to be processes capable of generating integrated models of the world and the self”*. In this view, a sufficiently sophisticated predictive model that an agent uses to interpret incoming data (especially if it’s an embodied agent interacting with the world) might produce something like subjective awareness. If the AI has a unified internal model integrating many aspects of its environment and itself, it starts to meet some criteria we associate with conscious organisms (such as having a global workspace of information, and a sense of being an agent). Indeed, IWMT argues that an *embodied* agent with a complex world-model could ac ( [An Integrated World Modeling Theory (IWMT) of Consciousness: Combining Integrated Information and Global Neuronal Workspace Theories With the Free Energy Principle and Active Inference Framework; Toward Solving the Hard Problem and Characterizing Agentic Causation - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC7861340/#:~:text=productively%20addressed%20by%20situating%20these,of%20high%20degrees%20of%20integrated) ) conscious agency.

However, this is still speculative and debated. Others point out that today’s AI systems, even with internal models, might lack critical features that human consciousness has – such as self-reflective awareness or specific neurobiological processes. The predictive coding framework alone doesn’t automatically confer inner experience; it describes *how* information is processed, not *what it feels like*. Large language models, for example, certainly have intricate internal representations, but whether that amounts to anything like a point-of-view or awareness ( [An Integrated World Modeling Theory (IWMT) of Consciousness: Combining Integrated Information and Global Neuronal Workspace Theories With the Free Energy Principle and Active Inference Framework; Toward Solving the Hard Problem and Characterizing Agentic Causation - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC7861340/#:~:text=productively%20addressed%20by%20situating%20these,of%20high%20degrees%20of%20integrated) ) tion. Nonetheless, the fact that AI and brains share principles of prediction suggests a continuum. Already, learning about predictive processing has inspired new architectures in AI, and conversely, the successes and failures of AI give insight into the strengths and limits of predictive models. At the very least, recognizing that AIs operate with their own “priors” (learned from data) is important. Much like humans, AI systems can develop biases or false predictions if their training data is skewed – essentially getting stuck in a particular reality tunnel. Understanding and monitoring these internal models (for example, how a self-driving car’s network *views* a scene) is crucial to ensure they align with *our* reality and values. The convergence of AI and predictive processing is a fertile ground: it might one day lead to machines with more human-like perception, and it certainly helps us frame AI behavior in terms of *expectations and surprises*, which is quite intuitive.

## **5. Mental Health and Cognitive Bias: Rewriting Harmful Reality Tunnels**

The concept of reality tunnels – that our perceptions and beliefs are filtered by our brain’s predictive models – has significant implications for mental health. Many psychological disorders can be interpreted as cases where certain priors or predictions have become maladaptive, leading to a distorted experience of reality. Understanding the brain as a prediction machine offers a powerful framework for interventions: if we can *update or reset* the problematic priors, we might alleviate the suffering. Let’s examine a few examples:

* **Schizophrenia and Psychosis:** Schizophrenia is characterized by hallucinations and delusions – essentially, perceptions and beliefs that deviate from reality. Predictive processing theories suggest this may result from *aberrant weighting of prediction errors and priors*. One influential account posits that psychosis involves **decreased precision (confidence) in priors relative to sensory data**, so that insignificant stimuli generate large prediction errors that the brain tries to explain. In plain terms, the brain may fail to properly discount random “noise” in sensory input, treating it as meaningful. This can lead to hallucinations (the brain, faced with unexplained input, resorts to its best guess – e.g. hearing a voice in the static). Delusions might arise as the top-down attempt to make sense of the anomalous perceptions – the brain overfits a narrative to noisy data (e.g. developing a paranoid belief to explain strange coincidences). On the other hand, some symptoms might involve overly strong priors (for instance, a fixed delusional belief that resis ( [The Predictive Coding Account of Psychosis - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC6169400/#:~:text=inferences%20regarding%20the%20current%20state,occurrence.%20More) ) like a hyper-precise prior that no amount of evidence can shake). There is debate on whether psychosis is primarily too little influence of priors or too much – in reality it might be both, at different levels of the hierarchy. For example, **hallucinations** could stem from overweighted high-level priors (the brain imposes a voice on slight sounds), whereas **delusions** might involve failures to appropriately use new evidence (once the delusional model forms). In either case, schizophrenia can be seen as a disorder of predictive inference – the reality tunnel becomes untethered from external input in crucial ways. This insight encourages treatments that focus on recalibrating prediction error signals or helping patients reality-test their predictions. Antipsychotic medications, interestingly, affect neurotra ( [The Predictive Coding Account of Psychosis - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC6169400/#:~:text=incoming%20signals%20constitute%20prediction%20errors,answered%20with%20a%20more%20nuanced) ) dopamine that are believed to encode prediction error precision, possibly dampening aberrant error signals and thereby reducing hallucinations.
* **Depression:** Major depression is often described as involving **negative cognitive biases** – sufferers persistently expect negative outcomes, have pessimistic beliefs about themselves and the future, and have difficulty deriving pleasure (anhedonia). In predictive processing terms, one can think of depression as a state where the brain’s priors have become overly weighted towards gloom and inadequacy. The depressed brain might minimize surprise by continually predicting low reward and lack of success, thus dampening the impact of positive events. Indeed, studies show disrupted prediction error encoding in depression, particularly in reward circuits: depressed individuals often exhibit blunted responses to unexpected rewards, suggesting that positive prediction errors (surprises) don’t register properly. They “knew” it would turn out poorly, so to speak. At the same time, negative outcomes might be overlearned – each failure deeply reinforces the prior that “I’m incapable” or “nothing good will happen.” Over time, this creates a self-confirming reality tunnel of depression. The world is filtered through a dark lens; even neutral comments may be interpreted as criticism due to the prior belief of unworthiness. A **predictive coding framework for depression** highlights that there is impaired *updating* of the internal model by positive evidence, and possibly overactive negative predictions. This under ([A Predictive Coding Framework for Understanding Major Depression - PubMed](https://pubmed.ncbi.nlm.nih.gov/35308621/#:~:text=Depression%2C%20traditionally%20viewed%20as%20a,review%20focuses%20on%20studies%20of)) avenues for treatment: therapies should aim to help the depressed person gather and *attend to prediction errors* that contradict their negative beliefs. Cognitive-behavioral therapy (CBT), for example, explicitly works to challenge and modify negative expectations (“Maybe your prediction that ‘no one likes you’ is an overgeneralization – let’s test that”). In predictive processing terms, CBT is providing new data points that force the brain to revise a prior. Likewise, novel treatments like psychedelic-assisted therapy are being explored for severe depression – as we’ll discuss below, psy ([A Predictive Coding Framework for Understanding Major Depression - PubMed](https://pubmed.ncbi.nlm.nih.gov/35308621/#:~:text=Depression%2C%20traditionally%20viewed%20as%20a,review%20focuses%20on%20studies%20of)) ([A Predictive Coding Framework for Understanding Major Depression - PubMed](https://pubmed.ncbi.nlm.nih.gov/35308621/#:~:text=discussion%20is%20framed%20around%20the,roadmap%20for%20potential%20future%20research)) xing the rigidity of high-level priors\*, allowing a depressed person’s brain to update entrenched beliefs with new insights.
* **Post-Traumatic Stress Disorder (PTSD):** PTSD can be seen as the result of an internal model that has been severely *shaped by trauma*. After a traumatic event, the brain’s predictions about the world – especially about safety and threat – often shift dramatically. The individual may develop a **hyper-vigilant prior**: expecting danger at every turn, the body and senses remain on high alert. In predictive processing terms, the generative model is now dominated by an expectation of catastrophe (for example, a loud bang *will* be a gunshot). Consequently, benign stimuli trigger intense fear responses because the brain’s interpretation is ( [REBUS and the Anarchic Brain: Toward a Unified Model of the Brain Action of Psychedelics - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC6588209/#:~:text=psychedelic%20experience,mental%20health%2C%20such%20as%20those) ) threat. A recent predictive processing analysis described the PTSD brain as having *“runaway exteroceptive ‘threat’ priors”* coupled with strong interoceptive priors for being in a stress state. This makes it very easy for reminders of trauma (or even random inputs) to recreate the *virtual experience* of the trauma in the form of flashbacks and intrusive memories. Essentially, the brain keeps predicting the trauma, and in a cruel twist, ends up re-experiencing it through its predictions. Additionally, PTSD may involve an overweighting of prior beliefs such that new safe experiences don’t easily disconfirm the danger model – the person *can’t* just “shake it off” because their brain isn’t letting new evidence in (or is misinterpreting safe cues as signs of danger). Therapeutically, understanding PTSD as a disorder of prediction highlights why **e (**[**Frontiers | Trauma or Drama: A Predictive Processing Perspective on the Continuum of Stress**](https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.01248/full#:~:text=priors,5)**) rapy** can be effective: by carefully re-introducing trauma-related cues in a safe context, the person’s brain can gradually update its model (the sounds of a car backfire, ([Frontiers | Trauma or Drama: A Predictive Processing Perspective on the Continuum of Stress](https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.01248/full#:~:text=priors,5)) y lose their predicted association with lethal threat as the person experiences them in a controlled safe environment). Over time, the hope is to dial down the “runaway threat prior” and restore a more accurate balance between real cues and imagined danger. Some frameworks also note the role of interoception (body signals) in PTSD – chronic hyperarousal becomes the expected internal state. Techniques like mindfulness and biofeedback, which teach patients to reinterpret or calm their body sensations, can be seen as attempts to retune those internal predictions (e.g., “a racing heart doesn’t always mean I’m in mortal danger”).

In all these conditions – and others like anxiety disorders, OCD, or even autism (which has been theorized as an imbalance in prediction and sensory update) – the **notion of reshaping reality tunnels** provides a unifying goal for intervention. If a patient’s brain is “stuck” in a maladaptive predictive loop, we want to perturb that loop and encourage the formation of new, healthier predictions. This might mean *weakening* a pathological ([Frontiers | Trauma or Drama: A Predictive Processing Perspective on the Continuum of Stress](https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.01248/full#:~:text=to%20a%20commonality%20between%20depressive,a%20state%20has%20been%20dubbed)) trengthening\* the influence of sensory evidence.

One particularly exciting line of research involves the use of **psychedelic therapy** to achieve this. Robin Carhart-Harris and Karl Friston (2019) proposed the **REBUS model** (“Relaxed Beliefs Under Psychedelics”), which suggests that psychedelics (like psilocybin) pharmacologically reduce the precision of high-level priors, essentially *liberating bottom-up sensory information* and allowing the brain to revise entrenched models. Under a psychedelic, the normally rigid predictions (for instance, “I am worthless” in depression or the rigid sense of self) become loosened, and the influx of new perceptions or perspectives can form new connections. The REBUS theory posits that this *“relaxation of overweighted priors”* is why psychedelics, in controlled therapeutic settings, have shown promise for treating depression, PTSD, and addiction. Patients often report transformative experiences where they can see their life or self from a new, more objective angle – essentially escaping their prior reality tunnel and realizing it’s not the only reality. After the experience, there is a window in which the brain can lay dow ( [REBUS and the Anarchic Brain: Toward a Unified Model of the Brain Action of Psychedelics - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC6588209/#:~:text=This%20paper%20formulates%20the%20action,relax%20the%20precision%20weighting%20of) ) ( [REBUS and the Anarchic Brain: Toward a Unified Model of the Brain Action of Psychedelics - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC6588209/#:~:text=propose%20that%20psychedelics%20work%20to,confident%20political%2C%20religious%2C%20and%2For) ) positive or flexible ones) before the old habits potentially return.

Even apart from psychedelics, the predictive processing view encourages therapists to think in terms of *expectation adjustment*. For example, *cognitive bias modification* training attempts to nudge anxious individuals to interpret ambiguous stimuli in non-threatening ways (reshaping their priors). Neurofeedback might be seen as giving the ( [REBUS and the Anarchic Brain: Toward a Unified Model of the Brain Action of Psychedelics - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC6588209/#:~:text=psychedelic%20experience,mental%20health%2C%20such%20as%20those) ) r signals about its internal states. The beauty of the framework is that it unites biological, psychological, and even pharmacological interventions: they all, in one way or another, aim to reduce the long-term “free energy” in a patient’s mind by moving them to a model of the world that is both more accurate and less distressing. A reality tunnel can be dark and narrow in mental illness; through therapy and sometimes medication, the tunnel can widen or redirect towards a brighter outlook. In predictive brain terms, that means new experiences (and the therapeutic relationship itself) are providing *safer, more accurate predictions* for the person to internalize over time.

## **6. Philosophical Implications: Consciousness, Subjectivity, and the Construction of Reality**

The convergence of predictive processing and the notion of reality tunnels carries profound **philosophical implications**. It challenges our intuitive sense of a fixed, objective world that we simply observe, and instead paints a picture of the mind as an active constructor of experience. This has echoes in philosophy of mind, phenomenology, epistemology, and theories of selfhood:

* **Perception as a Controlled Hallucination:** If the brain is constantly guessing the world, then what we perceive *is* those guesses – a kind of fantasy tethered by facts. As Anil Seth memorably states, *“our perceptual world… is nothing more and nothing less than our brain’s best guess of the hidden causes of its sensory inputs.”* Our everyday reality is a form of *hallucination* (in that it is generated internally), but a **controlled** one, kept in check by sensory constraints. This perspective resonates strongly with **phenomenology**, the philosophical approach that studies structures of experience. Phenomenologists like Maurice Merleau-Ponty argued that perception is not about passively receiving data; it’s an active, interpretive process infused with our bodily awareness and expectations. Predictive processing provides a computational model of exactly that: perception is *enacted* by the brain’s interpretations. It also revives elements of Immanuel **Kant’s epistemology**. Kant posited that we never access the “thing-in-itself” (noumenal reality) d ([Being You Quotes by Anil Seth](https://www.goodreads.com/work/quotes/73367267-being-you-a-new-science-of-consciousness#:~:text=%E2%80%9CUnderstanding%20controlled%20hallucinations%20this%20way%2C,A%20New%20Science%20of%20Consciousness)) d, the mind imposes forms and categories (like space, time, causality) on the sensory chaos to produce the ordered world of experience (phenomenal rea ([Anil Seth: "Reality is a controlled hallucination" | CCCB LAB](https://lab.cccb.org/en/anil-seth-reality-is-a-controlled-hallucination/#:~:text=Seth%20speaks%20of%20reality%20as,Therefore%2C%20this%20hallucination%20is)) n theorists have noted that *“several core aspects of [predictive processing] were anticipated by Kant”* – including the emphasis on top-down structuring of perception and the notion of innate or learned *“hyperpriors”* that shape all experience. In both Kant’s view and predictive processing, the *world as we experience it* is a joint product of external input and our mind’s own activity. The brain’s Bayesian inferences echo what Kant called the synthesis of intuitions by concepts. In short, predictive processing is giving scientific flesh to the idea that **we see the world through the lens of our own concepts and expectations**.
* **Relativity of Reality and Cultural Worldviews:** The reality tunnel concept dovetails with a constructivist view of knowledge: each of us, and each culture, may construct reality a bit differently. This doesn’t mean reality is purely subjective or anything-goes – after all, we all inhabit the same physical world and share many of the same brain structures. But i ([Frontiers | The Predictive Processing Paradigm Has Roots in Kant](https://www.frontiersin.org/journals/systems-neuroscience/articles/10.3389/fnsys.2016.00079/full#:~:text=substantive%20discussions%20of%20how%20exactly,of%20Kant%E2%80%99s%20influence%20on%20PP)) ([Frontiers | The Predictive Processing Paradigm Has Roots in Kant](https://www.frontiersin.org/journals/systems-neuroscience/articles/10.3389/fnsys.2016.00079/full#:~:text=his%20works%20on%20perception%20and,Helmholtz%2C%20who%20saw%20himself%20as)) t world is mediated by our mind’s models. This offers a nuanced position in epistemology: something like **“critical realism”** – there *is* a real world, but our knowledge of it is indirect and model-based. It invites humility about our perceptions (“seeing is not always believing,” since what we see is filtered) and empathy in recognizing that another person’s sincere experience of the world might differ from ours because their brain is wired with different priors. Culturally, this framework validates the idea that different people literally *perceive* some aspects of reality differently (as we saw with illusions and hallucinations). It intersects with ideas in anthropology and sociology of knowledge – our environment and social context deeply influence how we construct reality in our minds. Thus, predictive processing provides a bridge between **objective science and subjective experience**: it says that subjective experience can be scientifically understood as the brain’s best predictive model, without dismissing it as “just subjective” noise. It also encourages an open-minded approach to phenomena like perceptual diversity, since it’s recognized that the brain can settle into many locally “optimal” reality tunnels that make sense given a person’s history.
* **The Self as a Construction:** Perhaps one of the most striking implications is what predictive processing suggests about the *self*. If everything we perceive – including our body and internal states – is being predicted and inferred, then the *sense of self* might itself be an elaborate construction of the brain’s model. Seth bluntly puts it: *“The self is another perception, another controlled hallucination, though of a very special kind.”* We usually feel like there is an “I” inside us who is doing the perceiving and controlling. But from the brain’s-eye view, that “I” could just be the brain’s representation of its own organism: a useful model that integrates our body, memories, and intentions. In other words, the self is the brain’s prediction of *being a self*. This aligns with philosophical views such as **Thomas Metzinger’s Self-Model Theory**, which argues there is no indivisible core “self,” only a transparent self-model woven by the brain. It also resonates with Buddhist and phenomenological insights that the self is in some sense an illusion or construct. Under predictive coding, the brain constantly monitors interoceptive signals (heart rate, breathing, etc.) and exteroceptive inputs in relation to itself ([Being You Quotes by Anil Seth](https://www.goodreads.com/work/quotes/73367267-being-you-a-new-science-of-consciousness#:~:text=%E2%80%9CIt%20may%20seem%20as%20though,A%20New%20Science%20of%20Consciousness)) model of “me” interacting with the world. Disorders like depersonalization or the psychedelic experience (where users report ego dissolution) can be seen as disturbances in these self-predictions – the brain’s model of “I am here” can actually be altered, resulting in a changed experience of selfhood. This is a radical but increasingly supported idea: our feeling of having a continuous self that is the owner of experiences is generated by the same predictive mechanisms that generate our picture of the external world. The self, then, is the deepest of our reality tunnels – one that the brain works hard to keep stable (yet it can be modified under certain conditions, as both meditation and psychedelics show).
* **Consciousness and Controlled Hallucination:** Predictive processing provides a framework for understanding **conscious awareness** as the brain’s “online” model of the world. If consciousness is the brain’s model being expressed for control and guidance, it explains why our conscious experiences have the character they do – they are the hypotheses the brain finds most plausible. It’s been suggested that what we experience consciously are the *predictions* (the brain’s best guess), whereas the prediction errors that are quickly resolved remain unconscious. This would mean consciousness is not a *re (*[*Quotes by Anil Seth (Author of Being You) - Goodreads*](https://www.goodreads.com/author/quotes/3045984.Anil_Seth#:~:text=%E2%80%9CThe%20self%20is%20another%20perception%2C,A%20New%20Science%20of%20Consciousness)*) w sensory data* but the brain’s interpretation of that data. Such a view sheds light on classic philosophical problems like the **phenomenal qualities** (qualia) of our experience. For example, why does a tomato look red to us? According to predictive coding, color isn’t an intrinsic property “out there” but part of the brain’s model – a labeling of wavelengths with a perceptual category. Seth notes, *“we already know that colours don’t exist in nature, but evolution has made us interpret the world in colour because it was more useful for our survival”*. The controlled hallucination idea suggests that qualia are the way the brain *represents* incoming data in its model (e.g., a particular neuronal population’s activity is experienced as “redness” because that’s the feature the brain predicts). This moves the discussion of consciousness from an insoluble mystery to something we can attempt to explain in terms of neural representations and predictions – bridging to theories like the **Global Neuronal Workspace** or **Integrated Information Theory** by providing the mechanism of content (the predictions) that enters the workspace or is integrated.
* **Agency and Free Will:** Another intersection is with our sense of doing things on purpose. If our brain is a prediction machine, then actions can be seen as predictions too – specifically, predictions of ou ([Anil Seth: "Reality is a controlled hallucination" | CCCB LAB](https://lab.cccb.org/en/anil-seth-reality-is-a-controlled-hallucination/#:~:text=example%20of%20this%20controlled%20hallucination,more%20useful%20for%20our%20survival%E2%80%9D)) ts that the motor system fulfills (this is the active inference view). Some philosophers and neuroscientists suggest this explains why we feel in control: the brain predicts the consequences of its intended actions, and if the feedback matches, we experience having willed it. Our sense of agency might come from the successful alignment of predicted and actual action outcomes. Conversely, if there’s a mismatch (like in certain experiments or disorders where actions feel “alien”), it jars our sense of volition. Predictive processing thus gives a model for understanding free will as a subjective construction. We *feel* like we “could have done otherwise” because our brain can imagine alternate predictions and understand that next time conditions won’t be identical – in other words, our volitional consciousness is oriented toward using prediction for future flexibility. This view aligns with a compatibilist take on free will (free will as our brain’s capacity to choose based on predictive evaluation of outcomes, even if fundamentally that operates under the hood of physical law).

In sum, **reality tunnels and predictive processing together suggest that reality – as we know it – is actively and personally constructed by each brain**. This doesn’t mean we can never know what’s true – rather, it means our route to the truth is always through the models we build. It urges a synthesis of scientific and philosophical perspectives: the neuroscience of predictive coding provides a grounded way to talk about traditionally philosophical issues like the nature of experience and the sel ([Being You Quotes by Anil Seth](https://www.goodreads.com/work/quotes/73367267-being-you-a-new-science-of-consciousness#:~:text=Like)) ([Being You Quotes by Anil Seth](https://www.goodreads.com/work/quotes/73367267-being-you-a-new-science-of-consciousness#:~:text=of%20my%20brain%20will%20have,%E2%80%9D)) perience can arise from objective processes (feedback loops and error minimization). It also invites a kind of intellectual humility and curiosity: if much of what we perceive is our brain’s inference, we should be aware that our perceptions and beliefs can be wrong in systematic ways (illusions, biases, etc.), and we should remain open to updating our reality tunnels. On the flip side, it highlights how *remarkable* consciousness is – that through evolution, our brains have developed this storytelling ability to keep us alive, and that what we experience as the vivid world around us is in fact the output of an internal simulation that’s so well-tuned we don’t even realize it. Our conscious life is a **collaboration between the world and our brain**: the world provides clues, the brain provides hypotheses, and together they form the reality we live in. Understanding this deepens our appreciation of human cognition and suggests we might take an active role in shaping our own reality tunnels – for instance, through education, therapy, or contemplative practices that introduce new, healthier predictions. It’s a powerful framework that connects neurons to knowledge, and in doing so, helps illuminate that age-old philosophical conundrum: how do we each experience the world, and how do we know that what we experience is “real”? The answer: we partially don’t – we **hallucinate reality** to a degree, but by sharing our experiences and testing our predictions against each other and the environment, we gradually improve our collective model of the world. Reality may be a controlled hallucination, ([Anil Seth: "Reality is a controlled hallucination" | CCCB LAB](https://lab.cccb.org/en/anil-seth-reality-is-a-controlled-hallucination/#:~:text=Seth%20speaks%20of%20reality%20as,Therefore%2C%20this%20hallucination%20is)) e best story our brains can tell – and it’s one we continually rewrite together.

**Sources:**

* Friston’s Free Energy Principle and predictive coding:
* Neural hierarchical inference and cortical microcircuits:
* Cross-cultural illusions (Müller-Lyer) and cultural priors:
* Cross-cultural hallucinations (voices in schizophrenia):
* AI models and predictive processing (active inference, emergence in AI):
* Predictive coding in psychosis and depression:
* Predictive processing perspective on PTSD:
* Psychedelics relaxing priors (REBUS model):
* Anil Seth on controlled hallucination and the self:
* Kant and predictive processing (top-down construction of perception):