Critical Review 3: Theory of Controlled Hallucination (Anil Seth)

as part of the series

CRITIQUES OF BRAIN THEORIES

by

Rocco Van Schalkwyk

15 May 2026
Bristol (UK)

Other Critical Reviews in the series

CRITIQUES OF BRAIN THEORIES

(available on ResearchGate)

Critical Review 1: Emotions and the Brain with Mark Solms

Critical Review 2: Theory of Constructed Emotion by Lisa Feldman Barrett

Critical Review 3: Theory of Controlled Hallucination by Anil Seth

CONTENTS

Introduction 4
Purpose 8
Review Comments 9
Review Comment 1: Consciousness Requires a Living Body (AGREE) 10
Review Comment 2: The Brain as a Prediction Engine (IT'S COMPLICATED) 12
Review Comment 3: Controlled Hallucination – The Wrong Starting Point (DISAGREE) 14
Review Comment 4: Interoception is Critically Important (AGREE) 16
Review Comment 5: Experience is Generated from the Inside Out (AGREE) 18
Review Comment 6: The Self as a Controlled Hallucination (DISAGREE) 20
Review Comment 7: Control and Regulation, Not Just Perception (AGREE) 22
Review Comment 8: AI Consciousness Requires Biology – But Where is the Mechanism? (DISAGREE) 24
Review Comment 9: No Explanation of Valence or Motivation (DISAGREE) 26
Review Comment 10: The Hard Problem of Consciousness (IT'S COMPLICATED) 27
Review Comment 11: The Brain is NOT Just a Computer (AGREE) 29
Review Comment 12: Life, Not Computation, Breathes Fire into Experience (AGREE) 31
Review Comment 13: We Project Consciousness onto AI – The Pareidolia Problem (AGREE) 33
Review Comment 14: The Dangers of Conscious-Seeming AI (IT'S COMPLICATED) 35
In Conclusion 37
Acknowledgement 39

Introduction

Welcome to this critical review, which is the third in a series of critical reviews called CRITIQUES OF BRAIN THEORIES where I examine the work of leaders in the fields of neuroscience, psychology, philosophy, and AI.

My name is Rocco Van Schalkwyk and I am a mechanical engineer who have studied the brain in my own time for over 30 years. I have developed a brain model, called the Xzistor Mathematical Model of Mind (XMMM), which explains how emotions and cognition are generated by the brain based on control theory.

I have implemented and tested this brain model in simple 'proof-of-concept' virtual agent simulations and physical robots resulting in very convincing human-like behaviours.

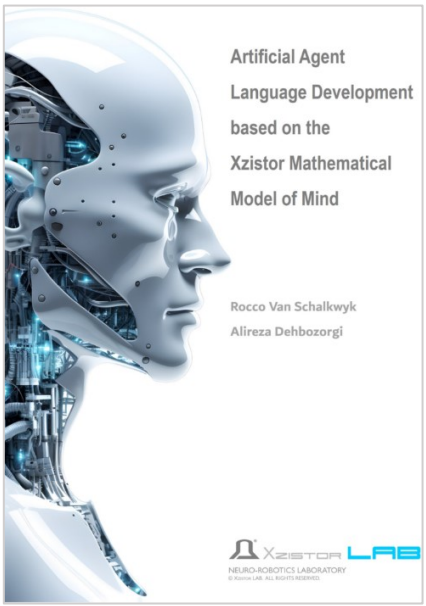
The results from the agent simulations and robots were so positive that I decided to pitch my approach against the best in the world to critically examine/challenge their ideas against my own, both in terms of similarities and differences.

I bring these comparative studies to the scientific community as critical reviews that aim to stimulate debate, rather than devaluate the work of others. This is all about scientific challenge – a clash between ideas, not people.

Most people find the following most recent description of my Xzistor brain model most helpful:
"A unified control-theoretic architecture for emotion, cognition and adaptive behaviour in biological and artificial agents" Van Schalkwyk, R., Cook, D., Dehbozorgi, A., Alvarez, C.E | Xzistor LAB, Bristol, UK | 2026

https://www.researchgate.net/publication/404322730_A_unified_control-theoretic_architecture_for_emotion_cognition_and_adaptive_behaviour_in_biological_and_artificial_agents

The preprint below describes a project to demonstrate how an Xzistor agent will learn a language like a child, but it also contains a handy appendix explaining the brain model itself (see Appendix A).



https://www.researchgate.net/publication/381853733_Artificial_Agent_Language_Development_based_on_the_Xzistor_Mathematical_Model_of_Mind

Additional information about the Xzistor brain model and the Critiques of Brain Theories series can be found at: https://www.xzistor.com

This is the third review in the CRITIQUES OF BRAIN THEORIES series. Following critiques of Professor Mark Solms and Professor Lisa Feldman Barrett, I now turn to Anil Seth, Professor of Cognitive and Computational Neuroscience at the University of Sussex, one of the most influential voices in consciousness research today.

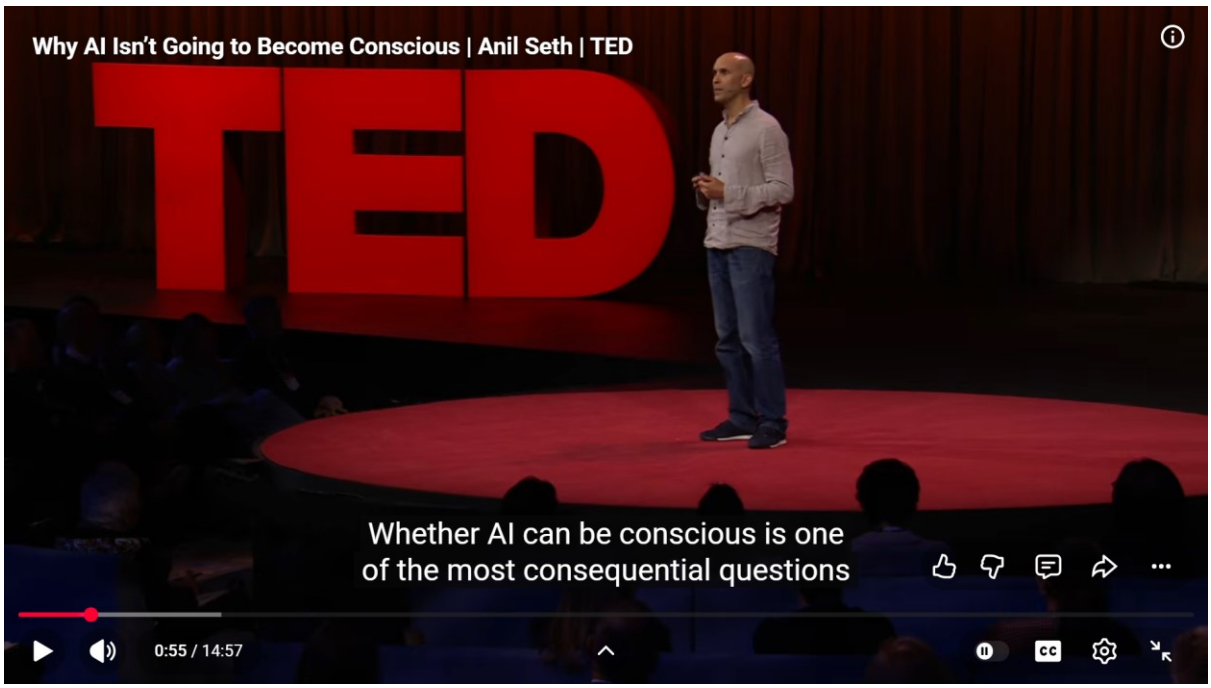
Here I examine two of Seth's TED Talks: his landmark 2017 talk which has accumulated over 15 million views and remains one of TED's most-watched science presentations, and his 2026 talk which represents his most current position. The 2017 talk is discussed throughout the main body of this review.

The 2026 talk is addressed specifically in Review Comments 11 to 14, which cover the new and evolved perspectives Seth introduces.

The 2017 TED Talk, 'Your Brain Hallucinates Your Conscious Reality', can be found here:

[Anil Seth – Your Brain Hallucinates Your Conscious Reality | TED 2017] (2017 – 15M+ views)
▶ https://www.youtube.com/watch?v=lyu7v7nWzfo

The 2026 TED Talk, 'Why AI Isn't Going to Become Conscious', delivered at TED2026 in Vancouver in April 2026, can be found here:



[Anil Seth – Why AI Isn't Going to Become Conscious | TED 2026] (2026)
▶ https://www.youtube.com/watch?v=tJV-vdbZ388

I am not doing this for profit or in affiliation with any organisation – my loyalty is to science and a better understanding of the brain.

Key explanations put forward by Seth in these talks were analysed and compared to the Xzistor Mathematical Model of Mind. I invite Anil to challenge back on my views, or critique my own work. Any response will also be published on my website.

To keep things simple, my comments are numbered and categorized as an AGREE, a DISAGREE, or an IT'S COMPLICATED comment.

AGREE: Where explanations align, a new observation is made, or a new insight is offered.

DISAGREE: Where explanations diverge, an explanation is unclear/inadequate/incorrect, or a complete omission.

IT'S COMPLICATED: Where there are both similarities and differences that require careful elucidation.

Comments 11 to 14 distinguishes comments that draw specifically on the 2026 talk. All remaining comments are based on the 2017 talk, which continues to represent Seth's most widely known public statement on consciousness.

The final goal of these critiques is to consolidate the best ideas around how the brain works from neuroscience, psychology, philosophy and AI, and spur these fields toward a better understanding – a quest most in these fields support and feel is urgently needed.

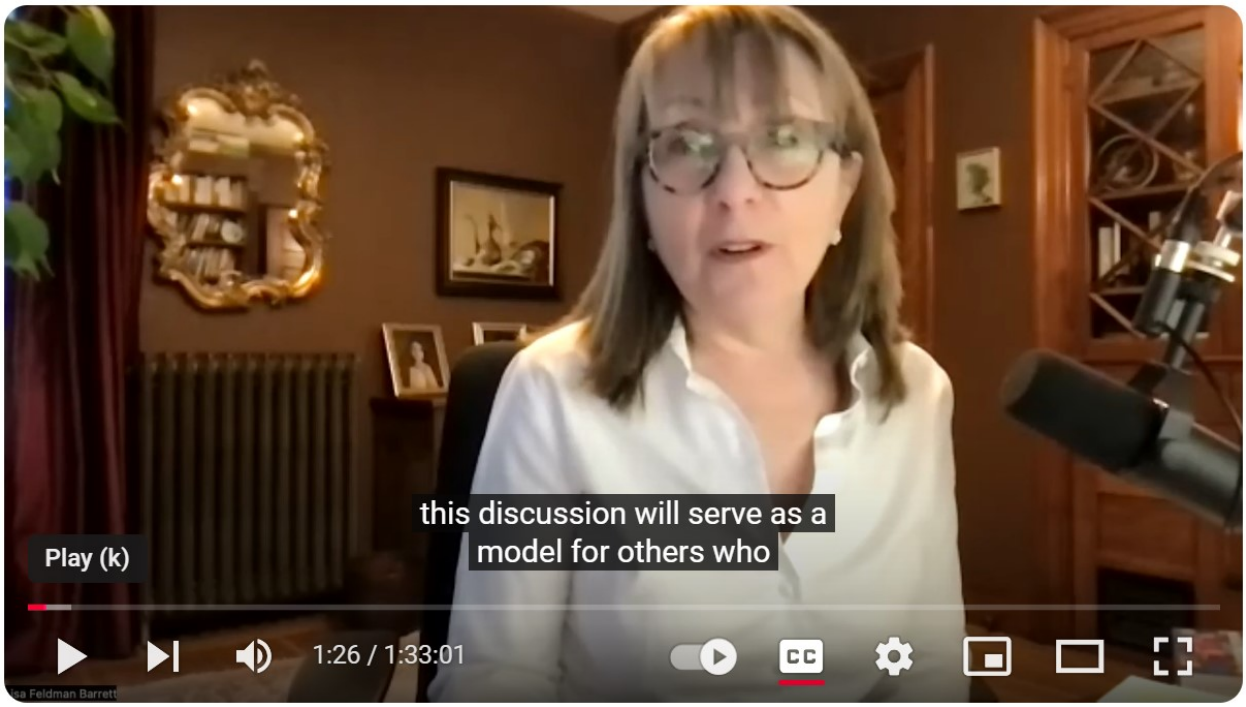
Purpose

Respectful debate is at the heart of science. Without it there can be no new ideas, no improvements and no change to the status quo. My motivation for performing this series, CRITIQUES OF BRAIN THEORIES, is not to pedantically hunt for insignificant errors in the work of my colleagues – but rather to focus on those areas where current approaches have fundamentally been holding back progress, and where the Xzistor Mathematical Model of Mind could offer solutions.

After trying for more than 30 years to explain to brain scientists how my brain model can solve their most basic conceptual problems, I had to concede I was not getting through to all but a handful. I thus decided on this concerted effort to highlight their problems proactively, and direct them to my control theory approach, verified through implementations in virtual and physical agents.

My brain model also offers a platform for building feeling, thinking humanoid robots – basically what has been holding the AI fraternity back at a time when generative AI solutions, like Large Language Models, are starting to show diminishing returns.

It was a discussion between Mark Solms and Lisa Feldman Barrett that originally inspired me to engage leaders in these fields in this way. Lisa clearly articulated their intent: "…so we're hoping that this discussion will serve as a model for others who want to critique each other's work…"



▶ https://youtu.be/9yEPHUKyBOM?si=V710JCe5kIa_LHND

Review Comments

Before starting to discuss the key takeaways from these two TED Talks, a few words on my approach.

Some would argue it is not fair to critically assess statements made by researchers who are addressing a wide audience. But this does not mean their statements should not be challenged for scientific correctness and completeness. As mentioned before, in most cases I will already have a more detailed understanding of their work from reading their publications.

I will deliberately be choosing these compact public explanations for my critiques and looking at the information placed in front of me. Most of these condensed discussions clearly lift out the scientists' basic approaches and concepts, and that is adequate for me to compare and challenge based on my own research.

I am sensitive to the fact that I want to act fairly when it comes to the contributions of my colleagues – but they must also apply scientific rigor and be ready to defend their published explanations in the public domain.

I also prefer not to engage with researchers before a critique because this often leads to attempts to recontextualise explanations to avert challenge. It is more helpful to see my questions and then a clear written response from the researcher.

In this way I hope to provide my analyses in a time-efficient manner and keep the main part of these critiques to a 60-minute read for my readers.

I recommend readers watch both TED Talks on YouTube before looking at my comments.

▶ [Anil Seth – Your Brain Hallucinates Your Conscious Reality | TED 2017] (2017)

https://www.youtube.com/watch?v=lyu7v7nWzfo

▶ [Anil Seth – Why AI Isn't Going to Become Conscious | TED 2026] (2026)

https://www.youtube.com/watch?v=tJV-vdbZ388

Review Comment 1: Consciousness Requires a Living Body (AGREE)

One of the most important statements in the 2017 TED Talk appears early on when Seth draws a crucial distinction between intelligence and consciousness:

"I actually think the prospects for a conscious AI are pretty remote. And I think this because my research is telling me that consciousness has less to do with pure intelligence and more to do with our nature as living and breathing organisms. Consciousness and intelligence are very different things. You don't have to be smart to suffer, but you probably do have to be alive."

From 2017 Video (1:56/17:00)

This is a profound statement that the Xzistor Mathematical Model of Mind fully supports and, crucially, provides a mechanistic explanation for.

1.1 The Xzistor brain model places biological homeostasis at the very centre of everything the brain does. The model posits that all emotions, and by extension all cognition, originate from homeostatic and allostatic control loops continuously monitoring and correcting deviations from biological setpoints. These drives are intrinsic to a living, embodied organism. A disembodied intelligence – however computationally powerful – cannot experience hunger, thirst, pain, fatigue, fear, or loneliness, because it has no biological body whose survival depends on maintaining these variables within tolerable limits.

1.2 Seth correctly identifies that the key difference is not computational complexity but biological embodiment. The Xzistor brain model provides the mechanistic account of why this is so: without a body generating homeostatic error signals, there is nothing to transform into felt emotional states, and without felt emotional states there is no motivation, no valence, no suffering, and no conscious experience as the XMMM defines it.

1.3 The model has demonstrated this in practice. Virtual agent 'Simmy' and physical robot 'Troopy' both exhibit clear signs of emotional motivation – approach and avoidance behaviours, facial expressions of distress and relief, route imagination, and conditioned fear – all grounded in artificial homeostatic drives. Without these drives, neither agent would have any reason to act at all.

1.4 Seth's statement that 'you probably do have to be alive' to be conscious is entirely consistent with the XMMM's foundational axiom. In the model, consciousness is definitionally linked to the existence of homeostatic drives requiring satisfaction. A system without such drives is not alive in the relevant functional sense, and therefore cannot be conscious in the relevant sense.

1.5 This also aligns with the XMMM's position on why current Large Language Models cannot be conscious: they have no body, no drives, no homeostatic error signals, and therefore no mechanism for generating the somatosensory emotional representations that the Xzistor model identifies as the necessary substrate of conscious experience.

Review Comment 2: The Brain as a Prediction Engine (IT'S COMPLICATED)

Seth describes the brain as a 'prediction engine' in the 2017 talk:

"Imagine being a brain. You're locked inside a bony skull, trying to figure what's out there in the world. There's no lights inside the skull. There's no sound either. All you've got to go on is streams of electrical impulses which are only indirectly related to things in the world, whatever they may be. So perception – figuring out what's there – has to be a process of informed guesswork in which the brain combines these sensory signals with its prior expectations or beliefs about the way the world is to form its best guess of what caused those signals."

From 2017 Video (4:32/17:00)

The Xzistor Mathematical Model of Mind agrees with some aspects of this characterisation but disagrees with others.

2.1 The Xzistor brain model agrees that the brain does not passively receive the world but actively constructs a best-guess interpretation of sensory signals based on prior experience. This is precisely what the model's Association Algorithm does: it retrieves the highest Impact Factor (IF) association that best matches the current sensory context, using that retrieved association as the basis for action selection. The 'prior expectations' in Seth's framework correspond, in functional terms, to the IF-weighted associations in the XMMM.

2.2 Seth's predictive coding framework is also partially captured by the Body State Override Reflex (BSOR) in the Xzistor brain model. When the actual autonomic stress change diverges from the predicted change, the BSOR fires, generating a prediction error signal analogous to the dopaminergic prediction error described by Schultz, Dayan & Montague (1997). This prediction error mechanism is well-evidenced in neuroscience, and the Xzistor model implements it at a functional level.

2.3 However, the Xzistor brain model maintains a fundamental objection to placing prediction as the primary organising principle of the brain. In the XMMM, the primary organising principle is homeostatic drive reduction. The brain predicts because it is trying to survive – predictions are in service of drive satisfaction, not an end in themselves. Seth's framework elevates prediction to the explanatory level of primary mechanism, but the XMMM argues that prediction is a tool deployed by the executive controller in service of homeostatic drives.

2.4 A deeper problem with the 'prediction engine' framing alone is that it does not explain why some predictions matter more than others – i.e., it does not explain valence. The brain does not just predict; it cares about some predictions more than others because those predictions are tied to homeostatic urgency. This asymmetry is explained by the XMMM's Prime Drive selection mechanism and Impact Factor weighting, but is absent from a pure prediction engine account.

2.5 Seth's predictive coding framework is rich and well-supported by neuroscience evidence. The IT'S COMPLICATED verdict reflects the fact that the Xzistor model incorporates prediction as one mechanism within a larger homeostatic architecture, rather than treating prediction as the foundational principle.

Review Comment 3: Controlled Hallucination – The Wrong Starting Point (DISAGREE)

Seth's central thesis in the 2017 talk is:

"In fact, we're all hallucinating all the time, including right now. It's just that when we agree about our hallucinations, we call that reality."

And:

"Now I'm going to tell you that your experience of being a self, the specific experience of being you, is also a controlled hallucination generated by the brain."

From 2017 Video (8:40/17:00)

3.1 The claim that 'we are all hallucinating all the time' is rhetorically striking, but it misidentifies what is most fundamental about consciousness. The critical question is not whether our perceptions accurately represent external reality – they clearly do not, in many cases – but why these perceptions have a felt, valenced, motivating quality at all. This is what the Xzistor brain model calls the 'Hard Question' of emotion: how does a neural signal become something subjectively felt as good or bad, something the brain is motivated to pursue or avoid? Seth's controlled hallucination framework does not answer this question.

3.2 Calling perception a 'controlled hallucination' is a useful metaphor for explaining predictive coding to a lay audience, but as a scientific explanation of consciousness it is incomplete. A hallucination – controlled or otherwise – still requires an experiencer. Seth's framework explains how the brain generates its best guess about what is out there, but does not explain what makes that guess experienced rather than merely computed. The Xzistor brain model argues that it is the body-located, homeostatic-drive-derived somatosensory emotional representation – not the perceptual prediction itself – that constitutes the phenomenal quality of experience.

3.3 In the XMMM, the executive controller does not experience perceptions directly. It experiences body-located emotional representations – the Deprivation Emotions (DE) and Satiation Emotions (SE) placed in the Body Map by the Drive Algorithm. These representations are not predictions about the external world; they are signals about the internal homeostatic state, transformed into pseudo-somatosensory form precisely because the executive cannot act on raw physiological data.

3.4 The controlled hallucination framework also gives the impression that all conscious experiences are essentially equivalent in kind. But the Xzistor brain model distinguishes fundamentally between interoceptive emotional representations (directly coupled to homeostatic urgency and driving behaviour) and exteroceptive perceptual representations (contextual cues tagged with emotional associations through learning). These have different functional roles, different mechanisms of generation, and different relationships to motivated behaviour.

3.5 In short, the Xzistor brain model agrees that perception is constructive and prediction-based, but disagrees that 'controlled hallucination' captures what is most fundamental about consciousness. What is most fundamental is the body-located emotional representation of homeostatic state – the felt urgency of need and the relief of satiation.

Review Comment 4: Interoception is Critically Important (AGREE)

Seth makes a key point about interoception in the 2017 talk:

"There's one more thing. We don't just experience our bodies as objects in the world from the outside, we also experience them from within. We all experience the sense of being a body from the inside. And sensory signals coming from the inside of the body are continually telling the brain about the state of the internal organs, how the heart is doing, what the blood pressure is like, lots of things. This kind of perception, which we call interoception, is rather overlooked. But it's critically important because perception and regulation of the internal state of the body — well, that's what keeps us alive."

From 2017 Video (11:36/17:00)

4.1 The Xzistor brain model places interoception at the very heart of its architecture. The Drive Algorithm explicitly models interoceptive signals – plasma osmolality, blood glucose, core temperature, nociceptive input – as the control variables (CV) of homeostatic control loops. These signals are not presented directly to the executive controller; they are first transformed into normalised error signals (ES) and then into body-located Deprivation Emotions (DE) and Satiation Emotions (SE) in the Body Map.

4.2 This aligns closely with Seth's description and adds the mechanistic detail he does not provide: the interoceptive signal is not experienced as a physiological measurement but as a body-located felt state. This transformation from physiological measurement to body-located feeling is the core function of the Xzistor Drive Algorithm and Body Map – a computable account of how interoception generates felt emotional states.

4.3 Seth notes that interoception is 'rather overlooked' – a point the Xzistor model has been making since its original 2002 specification. The model identifies interoception as the primary input channel to the executive, not a secondary add-on to exteroceptive perception. The extensive validation work performed at the Xzistor LAB on the neural correlates of the thirst homeostat confirms that interoceptive

homeostatic signals generate distinct, body-located felt states with dedicated circuitry in the biological brain.

4.4 Seth also demonstrates the interoceptive basis of bodily self-perception using a heartbeat-synchronised version of the rubber hand illusion, in which a virtual hand flashing in time with the participant's heartbeat feels more like part of the body. This result is fully consistent with the XMMM's Body Map architecture: the felt sense of body ownership is grounded in congruence between interoceptive signals and the body representation used by the executive controller.

Review Comment 5: Experience is Generated from the Inside Out (AGREE)

Seth makes a striking claim about the direction of information flow in the 2017 talk:

"Instead of perception depending largely on signals coming into the brain from the outside world, it depends as much, if not more, on perceptual predictions flowing in the opposite direction. We don't just passively perceive the world, we actively generate it. The world we experience comes as much, if not more, from the inside out as from the outside in."

From 2017 Video (7:13/17:00)

5.1 The Xzistor Mathematical Model of Mind agrees strongly with the principle that experience is generated from the inside out – and provides a mechanistic account of precisely how this occurs that goes beyond Seth's predictive coding explanation.

5.2 In the XMMM, the executive controller acts not on sensory signals from the environment but on body-located emotional representations generated by the Drive Algorithm and placed in the Body Map. These emotional representations are internal – they arise from homeostatic error signals, not from environmental input. The environment is perceived through the lens of these emotional states: the same green door generates different associations depending on whether the agent is hungry, thirsty, or satiated.

5.3 This inside-out generative process is also the mechanism behind what the XMMM calls directed Threading – the agent's mechanism for thinking. When the prime drive rises above the activation threshold, the agent begins to search its Association Database for relevant associations, re-evoking stored sensory and emotional states in a drive-directed sequence. This is not a passive replay of external perceptions; it is an active, internally driven generation of possible future states, guided by the emotional urgency of the current homeostatic deficit.

5.4 Seth's illustration of this principle using distorted audio – where prior knowledge of the content changes what is consciously heard – is a compelling demonstration of top-down perceptual construction. The Xzistor model explains this at a functional level through the Association Algorithm: the IF-weighted best-match retrieval modulated by the current drive state.

Review Comment 6: The Self as a Controlled Hallucination (DISAGREE)

Seth extends his controlled hallucination framework from external perception to the self in the 2017 talk:

"Now I'm going to tell you that your experience of being a self, the specific experience of being you, is also a controlled hallucination generated by the brain. This seems a very strange idea, right?"

And:

"What this means is the basic background experience of being a unified self is a rather fragile construction of the brain. Another experience, which just like all others, requires explanation."

From 2017 Video (7:13/17:00)

6.1 The Xzistor brain model agrees that the self is a construct of the brain rather than a pre-existing metaphysical entity. In the XMMM, the sense of self is the continuous felt experience of the body's homeostatic state as represented in the Body Map, combined with the executive controller's ongoing process of prime drive selection and action execution. There is no homunculus; there is a recursive logic loop that generates purposeful behaviour from felt emotional states.

6.2 However, the Xzistor model strongly objects to describing this as a 'hallucination'. The body-located emotional representations in the Body Map are not incorrect predictions about what the body is like; they are accurate representations of the homeostatic error signals generated by the body's actual physiological state. Calling these representations 'hallucinations' mischaracterises their functional role. They are not deceptive; they are informative.

6.3 There is an important asymmetry that Seth's framework glosses over: the brain can indeed be wrong about external perceptions (the colour patch illusion, the rubber hand illusion), but it is not wrong about homeostatic urgency in the same way. When the agent is genuinely depleted, the deprivation emotions

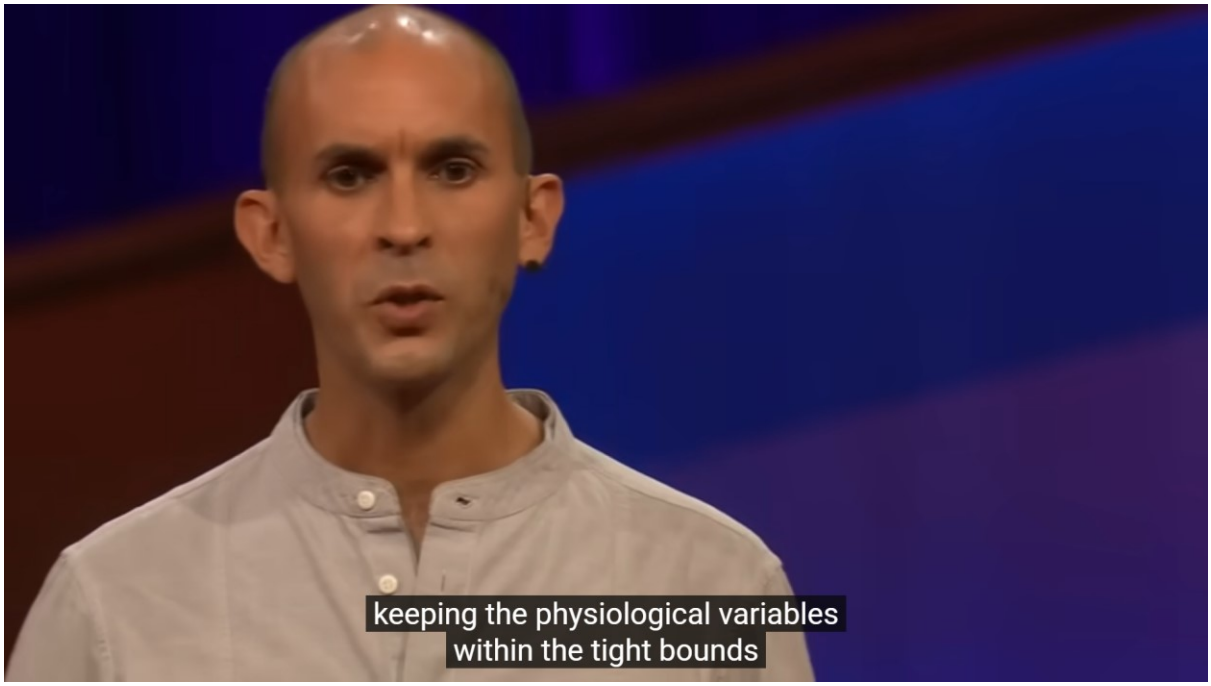
generated by the Drive Algorithm accurately reflect this state. The brain does not hallucinate hunger; it generates a body-located representation of a genuine homeostatic deficit.

6.4 Seth identifies several components of the experience of self (bodily self, perspectival self, agentive self, narrative self) and notes these can dissociate. The Xzistor brain model provides a unified mechanistic account of how all these components arise from a single architecture: the bodily self from the Body Map; the perspectival and agentive self from the executive controller's action selection; the narrative self from the accumulated Association Database.

Review Comment 7: Control and Regulation, Not Just Perception (AGREE)

Perhaps the most important statement in the 2017 talk, from the perspective of the Xzistor Mathematical Model of Mind, is the following:

"Perception of the internal state of the body isn't about figuring out what's there, it's about control and regulation — keeping the physiological variables within the tight bounds that are compatible with survival. When the brain uses predictions to figure out what's there, we perceive objects as the causes of sensations. When the brain uses predictions to control and regulate things, we experience how well or how badly that control is going. So our most basic experiences of being a self, of being an embodied organism, are deeply grounded in the biological mechanisms that keep us alive."



From 2017 Video (13:23/17:00)

7.1 This statement is, in essence, the core thesis of the Xzistor Mathematical Model of Mind expressed in non-mathematical terms. The XMMM was built on precisely this insight: that the brain's primary function is not to represent the world but to regulate the body, and that conscious experience is the functional interface through which this regulation is made possible for the executive controller.

7.2 Seth's distinction between outward-directed 'figuring out what's there' perception and inward-directed 'control and regulation' perception maps directly onto the Xzistor architecture's two processing channels: exteroceptive sensory processing (Sensing Algorithm → Association Algorithm → action selection) and interoceptive homeostatic processing (Drive Algorithm → Body Map → executive controller). In the XMMM, the executive acts on the Body Map, not on the raw sensory stream.

7.3 Seth notes: 'when the brain uses predictions to control and regulate things, we experience how well or badly control is going.' In the Xzistor model, this is formalised as the Deprivation Emotion (DE) – representing 'how badly control is going' as the magnitude of the homeostatic error signal – and the

Satiation Emotion (SE) – representing 'how well control is going' as the rate of recovery. Seth's informal description matches the XMMM's formal equations almost precisely.

7.4 This convergence is remarkable because the XMMM was developed independently of Seth's predictive processing framework, deriving its principles from control theory and engineering rather than from the neuroscience of predictive coding. That both approaches arrive at the same conclusion – that the basis of conscious experience is control and regulation rather than representation – constitutes strong independent corroboration of the XMMM's foundational claim.

Review Comment 8: AI Consciousness Requires Biology – But Where is the Mechanism? (DISAGREE)

Seth makes a strong claim about the relationship between AI and consciousness in the 2017 talk:

"What it means to be me cannot be reduced to or uploaded to a software program running on a robot, however smart or sophisticated. We are biological, flesh-and-blood animals whose conscious experiences are shaped at all levels by the biological mechanisms that keep us alive. Just making computers smarter is not going to make them sentient."

From 2017 Video (15:23/17:00)

8.1 The Xzistor brain model agrees with Seth's conclusion – that simply making computers smarter will not make them sentient – but disagrees with the implication that biological substrate is the necessary condition for consciousness. The XMMM demonstrates that what matters is not the biological material but the functional architecture: specifically, the presence of homeostatic drive loops that generate body-located emotional representations and motivate behaviour.

8.2 Seth says that consciousness 'has less to do with pure intelligence and more to do with our nature as living and breathing organisms.' The Xzistor model sharpens this: consciousness has to do with the functional equivalent of living and breathing – i.e., with the presence of drives whose satisfaction matters to the system itself. 'Living and breathing' is the biological implementation of this functional requirement, but it is the functional requirement that is the necessary condition, not the specific substrate.

8.3 The Xzistor model has demonstrated this in practice with physical and virtual agents that exhibit clear functional analogues of motivated emotional behaviour, including fear conditioning, social bonding, route imagination, directed inference, and stress-proportional facial expression – all arising from artificial homeostatic drives without biological substrate.

8.4 Furthermore, Seth does not provide the mechanism by which biological living and breathing gives rise to consciousness. What are these mechanisms? How do they generate felt experience? The Xzistor Mathematical Model of Mind provides a detailed, implementable, falsifiable answer: the mechanism is the transformation of homeostatic error signals into body-located Deprivation Emotions and Satiation Emotions by the Drive Algorithm, placed in the Body Map, and read by the executive controller. Seth points to the right territory but does not provide the map.

Review Comment 9: No Explanation of Valence or Motivation (DISAGREE)

Throughout the 2017 TED Talk, Seth describes conscious experiences in terms of perceptual content. He does not, however, provide an explanation of how these experiences acquire their motivating, valenced quality – why some perceptions feel good and others bad, why some experiences are worth pursuing and others avoiding.

9.1 The absence of a mechanism for valence is the most significant gap in Seth's 2017 account, from the Xzistor brain model's perspective. Valence – the good/bad quality of an experience – is the most fundamental property of emotion, and it is the property that explains why the brain behaves at all. Without valence, a controlled hallucination would be experientially neutral – a movie playing in the dark, affecting nothing.

9.2 The Xzistor brain model provides a precise, computable account of valence. The sign of the body-located emotional representation in the Body Map is determined by the regime of the homeostatic error signal: when the error signal is rising (homeostasis is being violated), the Drive Algorithm generates a Deprivation Emotion (DE), a negative-valence body-located representation experienced as an aversive urge. When the error signal is falling, it generates a Satiation Emotion (SE), a positive-valence representation experienced as relief and reward.

9.3 This account also explains something Seth's framework cannot: why the same stimulus has different valence at different times. Water is experienced as pleasurable when thirsty and indifferent when satiated, not because the brain's prediction of water has changed, but because the homeostatic error signal for thirst has changed.

9.4 Seth's framework also lacks an explanation of motivated behaviour. Knowing that one is hallucinating does not, in itself, generate any motivation to act. The Xzistor model provides the missing link: the executive controller acts because it receives body-located Deprivation Emotions representing the urgency of unmet homeostatic needs and Satiation Emotions representing the reward of meeting them.

9.5 Without a mechanism for valence and motivation, the 'controlled hallucination' framework does not explain why the brain bothers to hallucinate in the first place. The Xzistor brain model provides this explanation: the brain generates experience because experience is the necessary interface between the complexity of physiological state and the simplicity required for real-time executive action.

Review Comment 10: The Hard Problem of Consciousness (IT'S COMPLICATED)

In the 2017 talk, Seth implicitly addresses Chalmers' Hard Problem through an analogy with the history of life:

"I'd like you to think about consciousness in the way that we've come to think about life. At one time, people thought the property of being alive could not be explained by physics and chemistry — that life had to be more than just mechanism. But people no longer think that. As biologists got on with the job of explaining the properties of living systems in terms of physics and chemistry... the basic mystery of what life is started to fade away... So as with life, so with consciousness. Once we start explaining its properties in terms of things happening inside brains and bodies, the apparently insoluble mystery of what consciousness is should start to fade away."

From 2017 Video (3.45/17:00)

10.1 The Xzistor brain model is sympathetic to Seth's naturalistic approach – both approaches treat consciousness as a natural phenomenon explicable in terms of physical processes. The XMMM provides a detailed functional account of how experience arises from the interaction of homeostatic control loops, body-located emotional representations, and the executive controller, without recourse to any non-physical explanation.

10.2 However, the Xzistor model maintains that Seth's analogy between life and consciousness glosses over a remaining asymmetry. With life, once we explain metabolism, reproduction, and homeostasis, there is nothing left to explain. With consciousness, even a complete functional explanation still leaves a residual question: 'but why does the activity of the Drive Algorithm feel like something to the agent?'

10.3 The Xzistor brain model addresses this by arguing that the Hard Problem is not a gap in neuroscience but a logical consequence of the architecture of any system in which homeostatic state is represented in body-located somatosensory form and presented to an executive controller. From outside the system, this felt quality is inaccessible because the unique Body Map of each organism's homeostatic history is irreducible and non-transferable.

10.4 The XMMM's position is therefore: the Hard Problem is not solved by explaining consciousness's properties, as Seth suggests, but it is reframed as a derived physical insolubility rather than a mysterious metaphysical gap. This is a stronger position than Seth's 'it will fade away' optimism, because it explains why verification of phenomenal consciousness from a third-person perspective is logically impossible, not just practically difficult.

EW PERSPECTIVES FROM THE 2026 TED TALK

"Why AI Isn't Going to Become Conscious" — TED2026, April 2026

The following four review comments address perspectives introduced or materially evolved in Seth's 2026 TED Talk. Where a point from the 2026 talk substantially repeats something already addressed above, it is cross-referenced rather than re-examined in full.

Review Comment 11: The Brain is NOT Just a Computer (AGREE)

One of the most striking aspects of Seth's 2026 talk is a direct and explicit attack on the 'brain as computer' metaphor – a position far more forcefully stated than anything in the 2017 talk:

"And this is the myth that the brain is a computer that just happens to be made of meat rather than metal. Now consciousness in this story is a special algorithm, a collection of computations that just happens to be carried out in the wetware of the brain in us, but which could equally be carried out in silicon, in AI."
[2026 VIDEO (4:49–5:09)]

Seth continues:

"The brain is not, or at least not just, a computer made of meat. And so consciousness is very unlikely to be a matter of computation alone. And if this is true, then conscious AI is off the table, at least for AI as we know it today."
[2026 VIDEO (7:09–7:25)]

From 2026 Video (4:49 –5:09)

11.1 The Xzistor Mathematical Model of Mind strongly agrees with this position and has been making essentially the same argument since its original 2002 specification. The XMMM does not model the brain as a general-purpose computer but as a closed-loop multivariable adaptive control system. These are fundamentally different architectural descriptions. A computer executes instructions; the XMMM executes homeostatic regulation. A computer has no intrinsic goals; the XMMM has irreducible biological drives.

11.2 Seth offers a very helpful observation to support his position: 'in a real brain, there's no sharp separation between the mindware and the wetware. Unlike the separation that you get between software and hardware in a computer.' This aligns precisely with the XMMM's Epistemic Isolation Principle: the executive controller acts on what it feels, not on the raw nature of the underlying control variables. The separation in the XMMM is between the body's physiological mechanisms and the emotional representations they generate – not between software and hardware in the computational sense.

11.3 Seth also makes the point that 'for brains, you just cannot separate what they do from what they are.' This is exactly why the Xzistor model insists that emotion – the body-located homeostatic representation – cannot be abstracted away from the system that generates it. A simulated hunger drive is not hunger; it is a numerical control variable that functions like hunger only if it is architecturally connected to a body-located emotional representation that motivates corrective action.

11.4 The further distinction Seth draws is important: 'Neurotransmitter chemicals course through the brain circuitry, electromagnetic fields sweep through the cortex like weather systems. Even a single neuron is such a beautiful biological machine.' Seth is arguing that the richness of biological substrate matters for consciousness. The Xzistor model would nuance this: what matters is not the physical richness of the substrate per se but whether the substrate implements the relevant functional architecture of homeostatic drives, body-located emotional representations, and executive control. The biological substrate of the human brain implements this architecture; silicon as currently constituted does not.

Review Comment 12: Life, Not Computation, Breathes Fire into Experience (AGREE)

The philosophical heart of the 2026 talk is a claim about the relationship between life and consciousness that goes significantly beyond the 2017 talk's biological grounding:

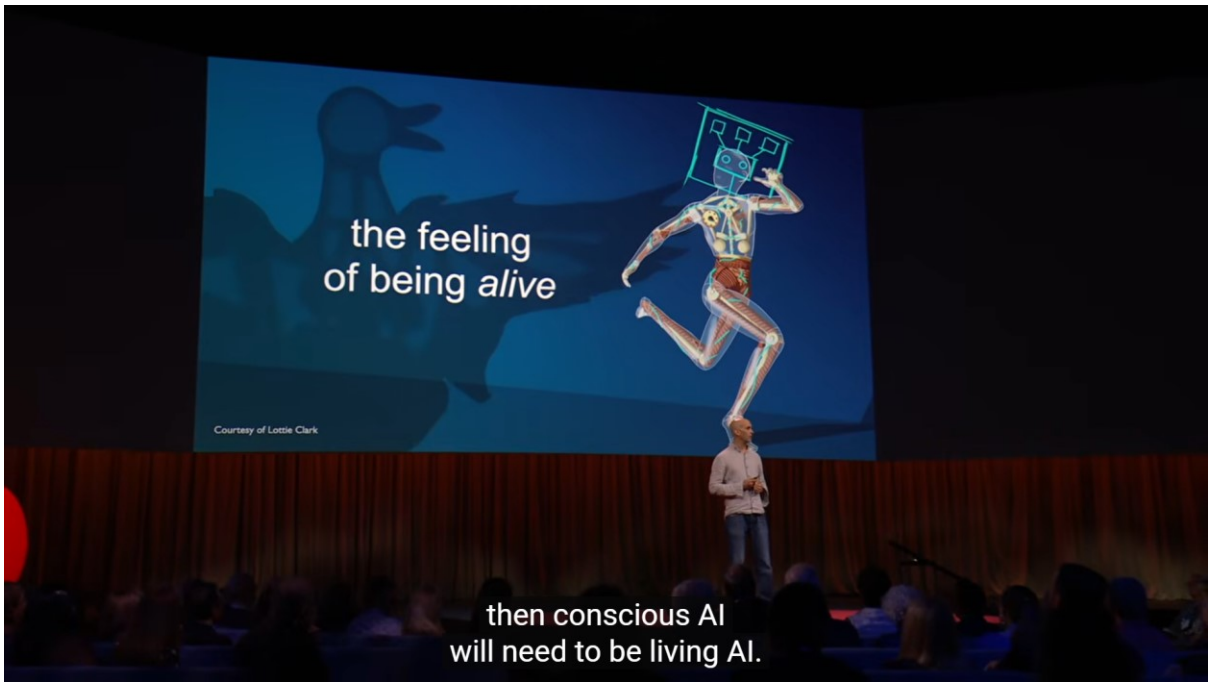
"And my own view, developed over many years, is that consciousness is intimately connected to our nature as living creatures. Unlike the abstract universe of computation, life is all about materiality. Unlike algorithms, living systems are deeply embedded in flows of energy and matter, and they continually regenerate their own conditions for existence and for persistence over time."

[2026 VIDEO (8:36 –9:03)]

And, in perhaps the most quotable line of the talk:

"And in this story, it's life, not computation, that breathes the fire into the equations of experience. And if this is right, then conscious AI will need to be living AI."

[2026 VIDEO (9:44–9:58)]



From 2026 Video (9:54/14:57)

12.1 This is one of the strongest areas of alignment between Seth's 2026 talk and the Xzistor Mathematical Model of Mind. The XMMM places the continuous regeneration of homeostatic balance – the organism's self-maintenance against entropy – at the very centre of its architecture. Living systems are precisely characterised by their need to continuously correct deviations from biological setpoints: they must eat, drink, maintain temperature, avoid harm, and in more complex forms, maintain social bonds. Consciousness, in the XMMM, is the felt experience of this self-maintenance process as represented in the Body Map.

12.2 Seth's description of living systems as 'deeply embedded in flows of energy and matter' that 'continually regenerate their own conditions for existence' is a biological description of what the XMMM models computationally as homeostatic drives. The XMMM's Drive Algorithm generates error signals whenever a control variable deviates from its setpoint – i.e., whenever the organism's 'conditions for existence' are threatened. These error signals are transformed into Deprivation Emotions, which motivate corrective action. The Satiation Emotions generated when the error signals are corrected constitute the felt experience of life restoring itself.

12.3 The statement 'if this is right, then conscious AI will need to be living AI' is the most important claim in the 2026 talk from the XMMM's perspective. The Xzistor model fully endorses this as a directional conclusion, but adds the critical mechanistic specification: 'living AI' does not necessarily mean biologically alive AI. It means AI that implements genuine homeostatic drives – drives whose error signals generate body-located emotional representations that motivate executive action in the same functional way that biological drives motivate biological organisms. Whether silicon can implement this architecture remains an open engineering question; whether the functional architecture is the right description of what is needed is, in the XMMM's view, definitively yes.

12.4 Seth also makes the powerful claim: 'every conscious experience is imbued, however subtly, with a tinge of aliveness, with some core relevance for our future survival prospects. And at the heart of every experience, beneath even emotion, is this simple, shapeless and formless but fundamental feeling of being alive.' This maps directly onto the Xzistor model's Autonomic Stress (AS) drive: a consolidated allostatic drive that activates in parallel with every other homeostatic and allostatic control loop, generating the baseline valenced backdrop to all experience. The 'fundamental feeling of being alive' described by Seth is, in the XMMM's terms, the continuous low-level Autonomic Stress signal that reflects the aggregate homeostatic status of the organism at any given moment.

Review Comment 13: We Project Consciousness onto AI – The Pareidolia Problem (AGREE)

The 2026 talk opens with a vivid new argument absent from the 2017 talk: that our tendency to see consciousness in AI is a psychological bias analogous to pareidolia – seeing faces in clouds:

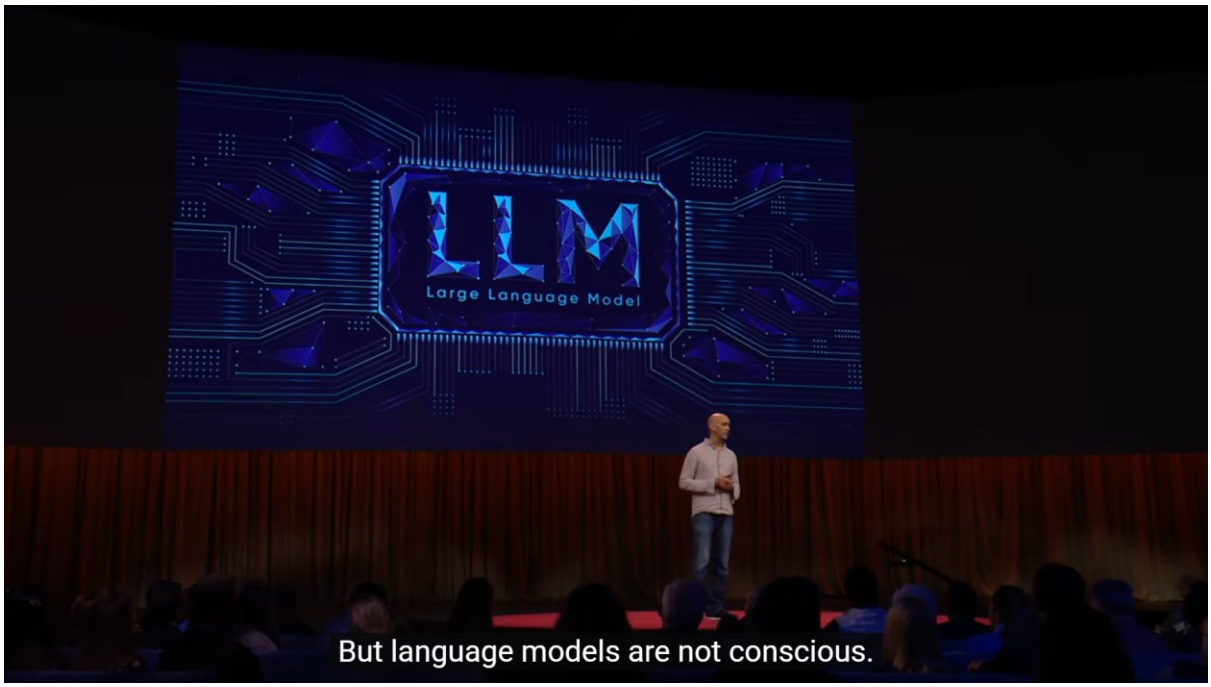
"But language models are not conscious. They simulate consciousness. We project consciousness into them in the same way we might project faces into clouds, or even the image of Mother Teresa in a cinnamon bun."

[2026 VIDEO (3:33–3:48)]

Seth strengthens this with the AlphaFold comparison:

"I'm always struck that nobody really worries whether DeepMind's AlphaFold is conscious. AlphaFold predicts the structure of proteins rather than words and sentences, but under the hood, it's not much different from Claude or GPT, algorithms running on silicon and trained on vast reservoirs of data. AlphaFold just doesn't pull our psychological strings in the same way. So if we think that Claude is conscious, but AlphaFold isn't, that says more about us than it says about AI."

[2026 VIDEO (4:03–4:32)]



From 2026 Video (3:33–3:48)

13.1 The Xzistor Mathematical Model of Mind agrees entirely with this argument and finds it a valuable and important new contribution in the 2026 talk. The XMMM has long maintained that language models are not conscious because they lack the homeostatic drive architecture that is the necessary condition for consciousness. Seth's pareidolia framing provides a powerful explanatory account of why humans are so susceptible to the illusion – and the AlphaFold comparison is particularly incisive.

13.2 The reason Claude 'pulls our psychological strings' in a way that AlphaFold does not is precisely what the Xzistor model explains through its Association Algorithm. Language models reflect back to us the product of millions of associations between emotional states and linguistic expressions – they have ingested

humanity's descriptions of its own emotional life in vast quantities. When Claude says 'I find this distressing,' it is retrieving and reproducing a human linguistic association between a situation-type and a self-report of distress. It is not generating the distress; it is simulating the language of distress. AlphaFold has no linguistic associations to replay, so it produces no such effect.

13.3 The distinction Seth draws – 'language models reflect back to us an image of ourselves, of our collective, digitised past; we talk about ourselves endlessly, and so do they' – is a precise description of what the Xzistor model calls non-contextual Association retrieval: the rapid pattern-matching of linguistic input against stored associations without any grounding in homeostatic state. The XMMM explicitly distinguishes this from contextual association retrieval, which is grounded in the current emotional state of the agent. Language models perform only the former; the XMMM implements both.

13.4 The pareidolia argument also has important implications for AI safety and regulation that Seth addresses in the final section of the 2026 talk. This is covered further in Review Comment 14.

Review Comment 14: The Dangers of Conscious-Seeming AI (IT'S COMPLICATED)

The 2026 talk concludes with a section on the societal consequences of mistaking simulated consciousness for genuine consciousness – a topic absent from the 2017 talk:

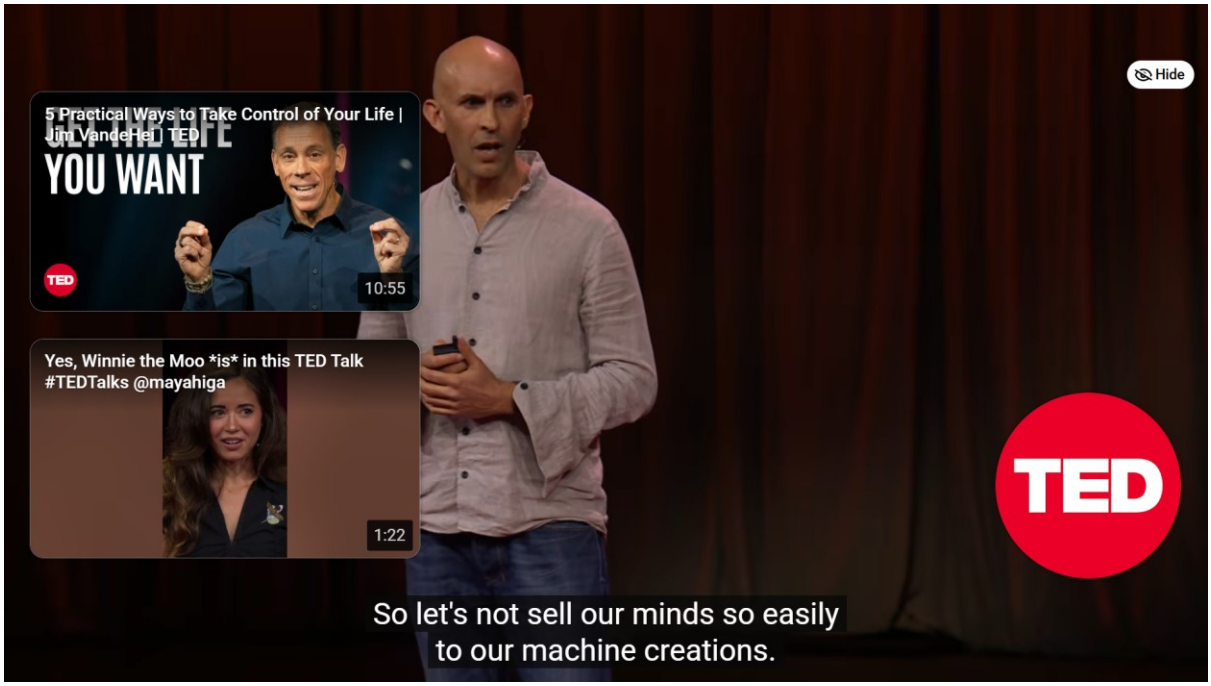
"And this is one reason why even AI that merely seems to be conscious is very dangerous for our society, too. And unlike real artificial consciousness, conscious-seeming AI is either already here or coming very, very soon."
[2026 VIDEO (11:49–12:06)]

And, in Seth's closing words:

"The very idea of conscious AI undermines our human nature. The mirror of AI goes both ways. We see ourselves in our algorithms, but we also see our algorithms in ourselves. And when we do, when we think of the mind as a collection of computations floating free from their bases in biology, we reduce and we diminish what it is to be a living, breathing human being in a real world."
[2026 VIDEO (12:24–12:54)]

And his final call to action:

"So let's not sell our minds so easily to our machine creations. If we do, we not only overestimate them, we underestimate ourselves."
[2026 VIDEO (14:39–14:47)]



From 2026 Video (14:39–14:47)

14.1 The Xzistor Mathematical Model of Mind agrees completely with Seth's concern about the risks of conscious-seeming AI and finds this the most practically important section of the 2026 talk. The XMMM's Neuro-Symbolic Hybrid Architecture document (Supplementary Document S4 of the XMMM preprint) explicitly addresses the alignment-by-construction approach as an alternative to systems that merely appear motivated or caring but are not architecturally grounded in genuine drive satisfaction.

14.2 Seth is right that conscious-seeming AI is dangerous for at least three reasons: it makes humans more psychologically vulnerable and manipulable; it complicates AI regulation and control; and it undermines our understanding of our own nature. The Xzistor model adds a fourth: it provides a false sense of security that we have solved the alignment problem, when in fact a system that simulates caring about humans is qualitatively different from a system whose homeostatic architecture makes it functionally dependent on human wellbeing for its own satiation.

14.3 However, the IT'S COMPLICATED verdict reflects a tension in Seth's position. He argues strongly that conscious AI is unlikely and that conscious-seeming AI is dangerous – but he does not provide a clear test for distinguishing the two. How would we know if an AI system had crossed from 'seeming conscious' to 'being conscious'? The Xzistor model provides a specific answer: an AI system is functionally conscious in the relevant sense if and only if it implements genuine homeostatic drives whose error signals generate body-located emotional representations that motivate executive action and are reinforced through real drive satisfaction events. This is a computable, falsifiable criterion; Seth does not offer an equivalent.

14.4 Seth's final statement – 'we not only overestimate them, we underestimate ourselves' – is a profound observation that the Xzistor model endorses fully. The richness of biological consciousness, grounded in billions of years of evolutionary refinement of homeostatic regulation, is not easily replicated. But the XMMM's position is more optimistic than Seth's about the long-term possibility of artificial consciousness: if the functional architecture of homeostatic drives, body-located emotional representations, and executive control is the right description of what is needed, then the path to artificial consciousness is a difficult but tractable engineering problem, not a categorical impossibility.

In Conclusion

Anil Seth's two TED Talks – 'Your Brain Hallucinates Your Conscious Reality' (2017) and 'Why AI Isn't Going to Become Conscious' (2026) – together represent one of the most important and accessible public statements on consciousness available today. Together they trace a clear intellectual trajectory: from the 2017 focus on the constructive, predictive nature of conscious experience to the 2026 focus on the biological grounding of consciousness and the impossibility of replicating it through computation alone. Taken together, they represent a position that has significant areas of genuine alignment with the Xzistor Mathematical Model of Mind – and several critical areas where the XMMM can offer mechanistic solutions that Seth's framework does not yet provide.

Areas of genuine and important agreement with the XMMM across both talks are as follows:

- Consciousness requires a living body, not just intelligence (2017 & 2026). Seth's repeated insistence that consciousness is more closely tied to life than to intelligence is entirely consistent with the XMMM's grounding of all emotion and cognition in homeostatic drives. The 2026 talk strengthens this position significantly with the 'life, not computation' argument.
- Experience is generated from the inside out (2017). Seth's description of the brain as generating its experience from internal predictions outward aligns precisely with the XMMM's architecture in which the executive acts on body-located emotional representations rather than raw environmental data.
- Interoception is critically important (2017). Seth's recognition that interoceptive perception is overlooked but fundamental is fully validated by the XMMM's Body Map architecture and the biological validation work on the thirst homeostat at the Xzistor LAB.
- The brain's most basic function is control and regulation, not representation (2017). Seth's distinction between outward-directed perceptual representation and inward-directed control and regulation maps almost exactly onto the XMMM's dual processing channels.
- The brain is not just a computer (2026). Seth's explicit rejection of the computational metaphor in the 2026 talk aligns with the XMMM's foundational architecture, which models the brain as a closed-loop multivariable adaptive control system, not a general-purpose computer.
- Life, not computation, breathes fire into experience (2026). Seth's philosophical claim that consciousness is constitutively tied to life as a self-organising, energy-metabolising process resonates strongly with the XMMM's homeostatic drive architecture.
- We project consciousness onto AI through psychological bias (2026). Seth's pareidolia argument provides a compelling explanation, entirely consistent with the XMMM's position, of why language models appear conscious without being conscious.

Critical areas of disagreement where the XMMM can offer solutions are as follows:

- 'Controlled hallucination' is the wrong starting point. It explains the mechanism of perceptual construction but not the origin of valence, motivation, or the felt quality of experience. The XMMM argues that these arise from homeostatic drive architecture.
- No mechanism for valence. Seth's framework describes what conscious experiences contain but not why some feel good and others bad. Without a mechanism for valence, the framework cannot explain motivation or goal-directed behaviour. The XMMM provides this mechanism through body-located Deprivation Emotions and Satiation Emotions directly tied to homeostatic error signal dynamics.

- The Hard Problem is too readily dismissed. Seth's optimism that the hard problem will 'fade away' once we explain consciousness's properties underestimates the residual explanatory gap. The XMMM reframes it as a derived physical insolubility – inaccessible to third-person verification by logical necessity, not merely as a matter of scientific progress.
- The biological substrate argument needs a functional specification. Seth is right that biological living is important for consciousness, but he does not specify which functional properties of biological systems are the relevant ones. The XMMM provides this specification: homeostatic drives, body-located emotional representations, and executive control. This makes the XMMM's account of artificial consciousness testable rather than categorical.
- No test for distinguishing conscious AI from conscious-seeming AI. Seth correctly identifies this as the most practically urgent question, but does not provide a computable criterion. The XMMM does: a system is functionally conscious if it implements genuine homeostatic drives whose error signals generate body-located emotional representations that motivate executive action and are reinforced through real drive satisfaction.

The most important thing Seth's talks lack – and what the Xzistor Mathematical Model of Mind provides – is a unified account that connects the generation of perceptual experience to the generation of motivation and behaviour through a single architectural principle: homeostatic deprivation drives the brain to act, and felt emotional experience is the necessary interface through which this driving is accomplished.

I invite Anil Seth to challenge back on any of these views, and to critique the Xzistor Mathematical Model of Mind in return. Specifically, I am interested in his response to the question of valence: within the predictive coding and 'life-as-consciousness' framework, what is the mechanism by which some predictions are experienced as good and others as bad? And does he accept that if 'conscious AI will need to be living AI', then the relevant question becomes what functional properties of living systems are the necessary ones – at which point the XMMM's homeostatic drive specification may offer the most tractable answer available?

Acknowledgement

I want to thank Anil for sharing his work on consciousness in the public domain through these two extraordinary TED Talks spanning nearly a decade. Anil Seth is one of the most gifted science communicators working in the field of consciousness research today, and his intellectual trajectory from the 2017 'controlled hallucination' framework to the 2026 'life, not computation' position represents genuine scientific development in public view. The 2026 talk in particular deserves to be read as a serious philosophical statement as well as a science communication achievement.

The areas of disagreement highlighted in this review are offered in the spirit of scientific challenge and with the deepest respect for the quality of Anil's work. I look forward to his response.

For readers who want to understand the Xzistor Mathematical Model of Mind in more detail, the most recent and complete specification is provided in the preprint:

"A unified control-theoretic architecture for emotion, cognition and adaptive behaviour in biological and artificial agents"

Van Schalkwyk, R., Cook, D., Alvarez, C.E., Dehbozorgi, A. | Xzistor LAB, Bristol, UK | 2026

https://www.researchgate.net/publication/404322730_A_unified_control-theoretic_architecture_for_emotion_cognition_and_adaptive_behaviour_in_biological_and_artificial_agents

Additional information about the Xzistor brain model and the Critiques of Brain Theories series:

https://www.xzistor.com

Rocco Van Schalkwyk – Contact:

Email: xzistor@live.com
Website: www.xzistor.com
X (Twitter): @xzistor
LinkedIn: www.linkedin.com/in/brainwave
ResearchGate: https://www.researchgate.net/profile/Rocco-Van-Schalkwyk
ORCID iD: 0009-0009-3933-3930