

Mark Solms, Consciousness as Precision Optimization

Some Physiological and Philosophical Considerations

ActInf GuestStream [#016](#)

Authors:

Mark Solms ¹, David S. Douglass ², Stephen Sillett ³, Daniel Ari Friedman ⁴

ORCID, Affiliations

1. 0000-0003-4828-3882, Department of Psychology, University of Cape Town, Cape Town, South Africa
2. 0000-0001-7894-8019, Active Inference Institute
3. 0000-0003-1046-1946, Active Inference Institute, Community Psychologist, Facilitator, Participatory Theatre Director PhD Candidate, Salomons Institute for Applied Psychology, Canterbury Christ Church University. Founder of InFusion Lab, Member of Global Psychosocial Network. Co-Founder of Aiding Dramatic Change in Development
4. 0000-0001-6232-9096, Active Inference Institute & Department of Entomology and Nematology, University of California, Davis, USA.

Live presentations and group discussions of the 2018 paper "How and Why Consciousness Arises: Some Considerations from Physics and Physiology" by Mark Solms and Karl Friston, https://discovery.ucl.ac.uk/id/eprint/10057681/1/Friston_Paper.pdf

Produced by Active Inference Institute in 2022

Consciousness as Precision Optimization

Some Physiological and Philosophical Considerations

Mark Solms

Neuroscience Institute,
University of Cape Town

Mark Solms

CONTENTS

Contents

CONTENTS	2
Session #016.1, Jan. 22, 2022	5
Introductions and Welcome	5
Summary of Points to be Covered	6
Slide: Points to present	6
Claim 1: Affect is the foundational form of consciousness and it is intrinsically conscious.	8
Slide: 'It is surely of the essence of an emotion... known to consciousness'	8
Question: Other possible 'foundations' for consciousness	9
Question: Feelings and interoception	9
Question: 'Affect' versus 'feelings'	10
Claim 2: This claim is based on empirical evidence, not on semantic arguments.	11
Slide. Reticular Activating System and Periaqueductal Grey	11
Case 1: Patients Without Cortex	12
Slide. Labeled diagram of side view of human brain	13
Image. Brain scans of hydranencephalic child	14
Slide. Hydranencephalic girl with infant brother (Merker 2007)	14
Image. Color photo of little girl laughing	15
Affective responses by hydranencephalic children	16
Case 2: Loss of pre-frontal cortex or of basal ganglia	16
Image. Drawing of human brain superimposed on shadow skull, with prefrontal cortex (PFC) in red	17
Image. Patient W: Two views of human brain with no PFC	18
Slide: Dialog between Solms and patient lacking PFC	19
Case 3: Damasio's patient with bilaterally destroyed insular cortex	20
Image. Painting of human brain sectioned to show insula (in red)	20
Case 4: Deep brain stimulation of the substantia nigra	23
Substantia nigra marked in red	23
Slide. Interview with deep-brain-stimulation patient (Blomstedt et al, 2008)	24
Human brain during four affective states	24
Colorized human brain; two images each during four affective states (Damasio et al 2005)	25
	25
Slide. Colorized human brain labeled by activity of serotonin, dopamine, and noradrenaline	26

Question: "Philosophical zombies?"	27
Question: The problem with uptake	28
Question: Cerebral palsy and other neurological conditions	29
Question: Hypervigilance and resistance	30
Question: Coma and evidence for consciousness	30
Question: Our inherited priors and new evidence	32
Question: Affects - ascending? descending?	32
Claim 3: Affect is not synonymous with interoceptive inference	34
Claim 4. This casts new light on the 'Hard Problem.'	37
Perception and Learning Without Awareness	40
Question: Pre-arousal + awareness	41
Question: Consciousness as metaphor	42
Perception as fantasy	43
Session wrapup	45
Question: Other thinkers	45
The importance of Active Inference	45
Session #016.2, June 11, 2022	46
Introduction; Recap of Session 1	46
Claim 1: Affect is the foundational form of consciousness and it is intrinsically conscious.	48
Claim 2: This claim is based on empirical evidence, not on semantic arguments.	49
Case 1: Patients Without Cortex.	49
Case 2: Loss of Pre-Frontal Cortex or of Basal Ganglia	51
Case 4: Deep Brain Stimulation of the Substantia Nigra	55
Claim 3: Affect is not synonymous with interoceptive inference.	58
Claim 4. This casts new light on the 'hard problem.'	59
Image. Functional Mechanisms of Visual Information Processing	60
Claim 5: Affect is an extended form of homeostasis.	63
Question: The Emphasis on Action	67
Question: Homeostasis is only part of the story	68
Question: Hormesis and homeostasis	69
Claim 6: Complex organisms require multiple (categorical) homeostats	71
Claim 7: These homeostats must be prioritized	73
The periaqueductal gray	73
Question: The "Tomkins/Panksepp Tradition?"	74
The control centers are in the brain	75
Claim 8: The mechanism of perceptual (and cognitive) consciousness is precision modulation of allostatic prediction errors	77

Case 5: Mr. S's tumor, and his confabulatory amnesia	81
Session Wrapup	88
Claim 9 (preview): The predictive hierarchy is arranged concentrically, for progressive tolerance of uncertainty	88
Claim 10 (preview): Consciousness can be artificially engineered	89
Looking forward	90
Supplemental Lists	92
Subjects	92
References	92
Mentioned Names & People	93
Other Resources	93

Session #016.1, Jan. 22, 2022

First participatory group discussion.

Session contributors: Prof. Mark Solms, Daniel Ari Friedman, Stephen Sillett, David S. Douglass

Videos (English with public comments): <https://www.youtube.com/watch?v=NHfGuwoCXI8>

Scrolling Closed Captioning available on YouTube:

- English (word for word)
- Chinese (simplified)
- Chinese (traditional)
- Dutch
- French
- German
- Italian
- Japanese
- Korean
- Portuguese
- Russian
- Spanish

Introductions and Welcome

00:07 *Daniel Ari Friedman:*

Hello, Everyone! Welcome to ActInfLab GuestStream #016.1 with Mark Solms. It's January 22nd 2022. It's going to be an awesome discussion! We really appreciate everyone taking the time out to participate in the live stream, and in the discussion. We're going to have some intermittent presentation intervals by Professor Solms, and also a period for discussion alternating with that. So we'll each just introduce ourselves and then pass it directly to the presentation section. So, I'm Daniel. I'm a researcher in California. And I'll pass to Dave.

00:51 *David S. Douglass:*

I'm Dave. I'm a software developer and student of psychology. I live two hundred kilometers north of Manila. Stephen?

01:02 *Stephen Sillett:*

Hello! I'm Stephen. I'm based in Toronto, Canada, and I'm doing a practice-based PhD around social topographies, and how people engage in community development. And I'll pass it over to Mark now. Welcome!

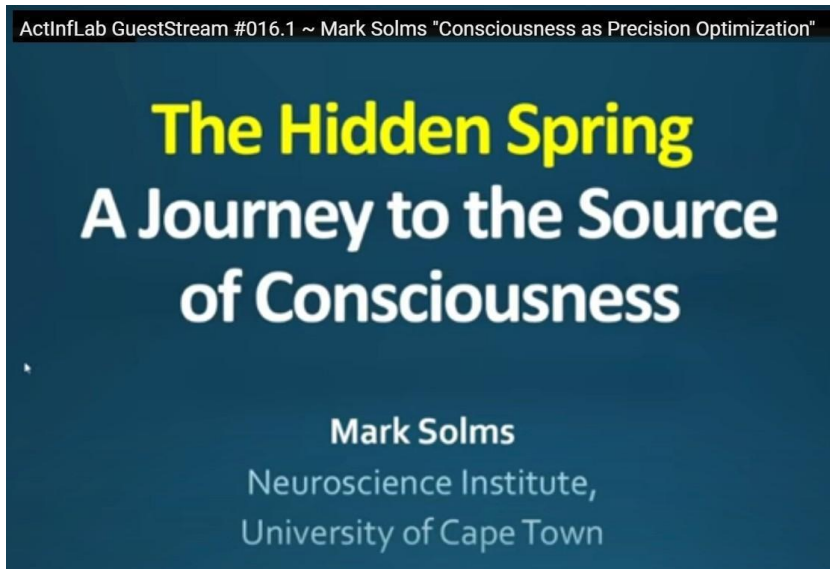
01:21 *Prof. Mark L. Solms:*

Thanks very much! In a moment I'll share my screen and you'll see my slides. I'm going to start with the summary slide of all the points that I propose to cover. And then I'll take them one by one, each of those points. And at the end of each of them, I'll stop and see if anyone wants to comment or ask a question. And then we slowly plow through it like that. We've got two hours so it should be okay. But if worse comes to worst and we don't get to the end of the list of topics, we will arrange a follow-up meeting to conclude them.

So I'm very glad to be here - thanks for inviting me! And I'm really looking forward to this discussion. Here come my slides.

Summary of Points to be Covered

02:24 On the screen is the title of my book.



And here is the promised summary of what I'm going to cover.

Slide: Points to present

- Affect is the foundational form of consciousness and it is intrinsically conscious.
- This claim is based on empirical evidence, not on semantic arguments.
- Affect is not synonymous with interoceptive inference.
- This casts new light on the 'hard problem'.
- Affect is an extended form of homeostasis.
- Complex organisms require multiple (categorical) homeostats.
- These must be prioritized.
- The mechanism of perceptual (and cognitive) consciousness is precision modulation of allostatic prediction errors.
- The predictive hierarchy is arranged concentrically, for progressive tolerance of uncertainty.
- Consciousness can be artificially engineered.

As I said, I'll cover each of these points in turn; and then we'll have questions and comments in between. And if we don't get to the end that's fine. Rather, let's rather give each one of the topics the time that it needs.

02:53 I start with the assertion (1) that "affect is the foundational form of consciousness, and that affect is intrinsically conscious."

I then move to make the point (2) that "that claim that *affect is intrinsically conscious* - in other words, that you can't have a feeling that you don't feel - that this is not based on semantic arguments, but rather on empirical evidence." And I will summarize the evidence.

I'll then make the point (3) that "*affect is not synonymous with interoceptive inference*, which is widely believed these days;" and I'll tell you why that matters when I get there.

And then I'll move on to making the point (4) that "all of this casts *new light on the 'hard problem'*" - particularly if affect is not just an interoceptive form of perception. Because there are lots of problems about perception that have got us into this hard problem in the first place.

04:03 I'll argue (5) that "affect, rather than being interoceptive inference, that it's an *extended form of homeostasis*." I'll tell you what I mean by that when we get there.

And I'll make the point (6) that "complex organisms like ourselves require multiple Homeostats, and that these are to be treated as categorical variables" - which matters.

And that (7) "these variables must be prioritized:" one can't meet all one's needs simultaneously.

This one is a large point (8) - the empirical evidence about affect - "The mechanism of perceptual and cognitive consciousness is *precision modulation of allostatic prediction errors*." That's a big claim and so we must spend some time on it.

04:55 I'll then talk about (9) the nature of the predictive hierarchy. I seem to see it a little differently from most people. So I want to set out my stall on that score, and explain why I think "*the predictive hierarchy is arranged concentrically*, and that as one moves toward the periphery of the hierarchy, there's progressively increasing tolerance of uncertainty."

And then lastly (10) I'll say a word or two about why I think that "consciousness can be *artificially engineered*."

So even if we don't get to the end of all of these topics, one way or another, even if we're in the second meeting, those are the topics we will cover. And we'll see how far we get tonight. And at least you know what the overall... When I say tonight, I'm in Berlin - I know it's morning for some of you. But at least you get the overall picture from that slide.

Claim 1: Affect is the foundational form of consciousness and it is intrinsically conscious.

05:57 So let's start with "Affect is the foundational form of consciousness and it's intrinsically conscious."

Slide: 'It is surely of the essence of an emotion... known to consciousness'

"It is surely of the essence of an emotion that we should be aware of it, i.e. that it should become known to consciousness. Thus the possibility of the attribute of unconsciousness would be completely excluded as far as emotions, feelings and affects are concerned."

Sigmund Freud (1915)

The discoverer of the unconscious - I quote him because Freud, more than anybody, wanted us to realize just how much of mental life is unconscious. Even Freud said that "It is surely of the essence of an emotion that we should be aware of it, that is, that it should become known to consciousness. Thus the possibility of the attribute of unconsciousness would be completely excluded as far as emotions, feelings and affects are concerned." It's quite a startling statement from (as I said) the discoverer of the unconscious, to say that "the possibility of the attribute of unconsciousness would be completely excluded as far as emotions, *feelings* and affects are concerned!"

06:56 Now I know this is a controversial issue, and that there are all sorts of ways in which people want to argue that there *are* such things as unconscious emotions and what not. And this has partly to do with the fact that different people mean different things by these words "emotions," "affects," and so on. And there are also all kinds of methodological operationalizations of these things. I won't go into all of that. I'm happy to, if any of you want to discuss it. But I've emphasized the word "feelings" to make clear that... If anyone wants to argue that they think that "the word 'emotion' can apply to unconscious processes," or "'affects' can apply to unconscious processes," that's fine by me.

07:52 But the word "feelings" - I don't see how you can ever argue that feelings can be unconscious! It just doesn't make any sense that a feeling could exist without being felt! If it were not felt it would not be a feeling. And so I'm going to focus everything that I'm going to say on this process - this thing - this mental state called "feelings." And I'm arguing that feelings are the foundational form of consciousness, and that *this* form of consciousness is intrinsically conscious. That's a sort of a dogmatic statement! I'm not sure if anyone wants to discuss that. Perhaps you want to hold your horses. But that's the end of my first point. So anyone who wants to interrupt here, go ahead. Otherwise I'll move on.

08:56 *Friedman:*

Thanks for the digestible, claim-based format!

Question: Other possible 'foundations' for consciousness

09:02 *Friedman:*

My question would be,

What else has been posited (in your view) as a foundational form of consciousness? So what are the counterfactuals for that first claim? What other horses are in the race?

09:13 *Solms:*

Let me say that the whole search for a **neural correlate of consciousness** (which is in a way where the trouble began, with Francis Crick) - it was focused on visual perception. And there're good reasons to focus on visual perception -- mainly because, if you take a look at your (you'll excuse the pun) -- take a look at your consciousness right now, it's probably dominated by visual perception. Visual **qualia** predominate in our consciousness. And a very large part of our cortex is given over to visual processing. And everybody's always believed that cortex is where it's at, when it comes to consciousness. And we also understand a very great deal about visual cortical mechanisms and visual mechanisms altogether.

10:21 So that was a perfectly reasonable place to start. So that's the other main horse that's in the race. But vision was thought to be a good model example, not because it is the totality of consciousness, but because it was assumed that, "if we can crack the problem of how *visual* consciousness arises, then we can generalize our conclusions to other forms."

And I'm going to argue that there's something profoundly different about visual consciousness and affective consciousness. And, moreover, that there's something profoundly different about cortical consciousness.

11:02 The brainstem mechanisms, which is where affective consciousness arises -- and I know that I'm saying that boldly now -- and I'm about to argue that point in the second of the bullet points on the screen.

Affective consciousness is not a cortical form of consciousness; it's a brainstem form -- and that the cortical form of consciousness is contingent upon the brainstem form. And this is one of the main reasons - apart from evolutionary reasons (which sort of interdigitate with and flow from that anatomical localization) - why I think that affect is the foundational form. So it's really vision that I'm using as my kind of foil, to answer the question simply.

11:56 *Friedman:*

Stephen.

Question: Feelings and interoception

12:00 *Sillett:*

One question with feelings is --

Would you say that feelings are integrated with proprioception, in the sense that they are a felt sense of where the body is as well? So it brings it into kind of an action-oriented domain? Or is it enough to have sort of a sense of how things are going, sort of an overview of just an emotional state? And then whether those two connect in some way?

12:32 *Solms:*

As I'm going to argue when I get to my third point, I don't think that affect is synonymous with interoceptive inference altogether - interoceptive perception altogether. And (I don't know) -- some people would class proprioception with interoception; some people would not; some people would say it's a sort of a hybrid between interoceptive and exteroceptive perception. But to the extent that by the word "proprioception" you're referring to a form of *perception*, I would disagree with that. Because, I don't think that affect is a form of perception.

If by "proprioception" you mean what neurologists normally mean by that word, it's part of the broader category of **somatic sensation**, and therefore it does not qualify, I think, as *affective*. Because, as I say, I'm going to argue that affect is *not* a form of perception.

13:47 I draw a distinction, as do... I think it's perfectly fine to draw a distinction between exteroceptive bodily sensations and interoceptive bodily sensations. But I think that the interoceptive type is *nothing other than perception*, and not intrinsically different. So even if you think of proprioception as *straddling* those two domains, I still think that it's a modality of perception, and suffers from all the same problems as vision for that reason.

I see we have a comment in the chat line.

Question: 'Affect' versus 'feelings'

14:32 *Friedman:*

Yeah, I'll read this. It was a question in the live chat from Ingrid. They wrote, "What differentiates affect from feeling, from your perspective? Is affect conscious, or pre-conscious? Thank you."

14:47 *Solms:*

You know, this is a perennial problem. We don't all mean the same thing by these terms. So that's why I've focused on the word "feeling." Because there can be little doubt what "feeling" denotes: it denotes something you *feel*. In other words, something that you are *necessarily* aware of. If you were not aware of it, it wouldn't be called a feeling, because you wouldn't feel it. "Affects" and "emotions" denote many different things to different people. So when you ask me, "is affect pre-conscious or conscious" and so on, it depends what you mean [chuckles] by "affect."

15:33 To *my* mind (and I can only just make clear how *I* use these words) -- I use the word "feeling" to denote "the subjective experience of affect." "Affect" is a kind of abstract term for "a function that we *feel*". But that doesn't mean that you all have to use the word in the same way. And it's precisely in order to not get lost in this kind of semantic tangle that I'm focusing on the word "feeling." And I'll make clear in the very point that I'm moving to next: what I mean by feeling in terms of its physiological and anatomical and other objective instantiations.

Claim 2: This claim is based on empirical evidence, not on semantic arguments.

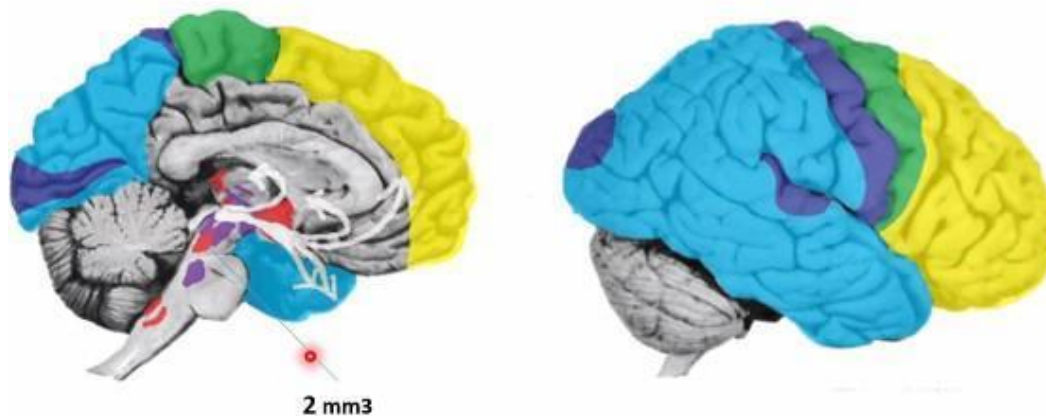
16:37 Solms:

So, the second point is, "This claim is based on empirical evidence, not on semantic arguments."

As you saw, we've just been talking about *the meaning of the words*. And I've staked my claim for the word "feeling" to make clear that that's what I mean when I say that "affect is intrinsically conscious" - I mean that it is necessarily *felt*. I'm also saying it's the *foundational* form of consciousness. That doesn't flow from the word, except perhaps from the word in the way that [Thomas] Nagel and [David] Chalmers define "consciousness." And many people seem to have sort of agreed with this way of defining it, that "there is something it is like to be conscious" - that this **something-it-is-likeness** seems to be more or less the same thing as feeling - it seems to mean the same thing - but "feeling" in a rather broader sense of the word than "*affective* feeling."

17:37 So I want to be clear now why (other than those linguistic or semantic points) - why I'm making this claim about feeling. It's not by any means based in the semantics that "you have to feel a feeling!" I want to be clear why I think that this is the foundational form of consciousness.

Slide. Reticular Activating System and Periaqueductal Grey



RETICULAR ACTIVATING SYSTEM and PERIAQUEDUCTAL GREY

We've all known since 1948, when there was this big surprise that Magoun and Moruzzi found, that *all* consciousness is contingent upon what we subsequently called "**reticular activating arousal**," in other words upon the "something" that's contributed by these core brainstem structures, which after Magoun and Moruzzi's work came to be known as the **reticular activating system**.

18:33 I want to be clear: I'm not only talking about the reticular activating system, but also about the **periaqueductal gray**, which is not technically part of the reticular activating system. But it is part of the brain structures which seem to be fundamental to the generating of consciousness.

What Magoun and Moruzzi showed, among other things, was that, if you make a small lesion above the reticular activating system, thereby severing it, disconnecting it, from the forebrain and from the cortex in particular, then consciousness is obliterated entirely. Cortical consciousness is *100% dependent* upon brainstem arousal.

19:27 That was a shock! And a shock that I don't think we have fully come to terms with, even all these years later. It should have led us to realize that we've been looking in the wrong place for the neural correlate of consciousness if we're looking to cortex. Because it's plain, even from just that study performed in 1948, that cortical consciousness is derived from elsewhere. Cortical consciousness *cannot occur* in the absence of reticular activating arousal.

20:07 But the way that Magoun and Moruzzi fudged it was to say - and everyone more or less in their wake has said the same - that: "It turns out there're *two* kinds of consciousness. There's the *brainstem* kind, which we call '**wakefulness**' or the '**level of consciousness**;' and then there's the *cortical* kind, which we call '**awareness**' or the '**phenomenal contents of consciousness**.'"

So this distinction - between a *quantitative level* contributed by the brainstem. and *qualitative contents* contributed by the cortex - is the way in which we saved the cortical theory of consciousness, following Magoun and Moruzzi's seminal observations.

21:05 In case you're wondering what *this* refers to [indicating label "2 mm³" or "2 cubic millimeters" on the Figure] - it refers to a more recent study by [David B.] Fischer and colleagues, which showed that the smallest lesion necessary to produce **coma** in human beings is a two cubic millimeter lesion - in other words, a lesion the size of a match head, in the **parabrachial complex**.

So there's no doubt, on the basis of Magoun and Moruzzi's studies all those years ago, and on the basis of these much more recent observations, that consciousness cannot exist without whatever-it-is that this part of the brain contributes, this primitive brainstem core.

But this idea - that it's merely a level, or a quantitative dimension of consciousness - is what I think is unsustainable.

22:06 Because I'm sure many of you have heard me talking about this before, I'm going to very quickly whiz through the evidence for why [it's wrong] - this notion that "the brainstem contributes only a quantitative level, a sort of a wakeful prerequisite for consciousness, rather than the quality and the content."

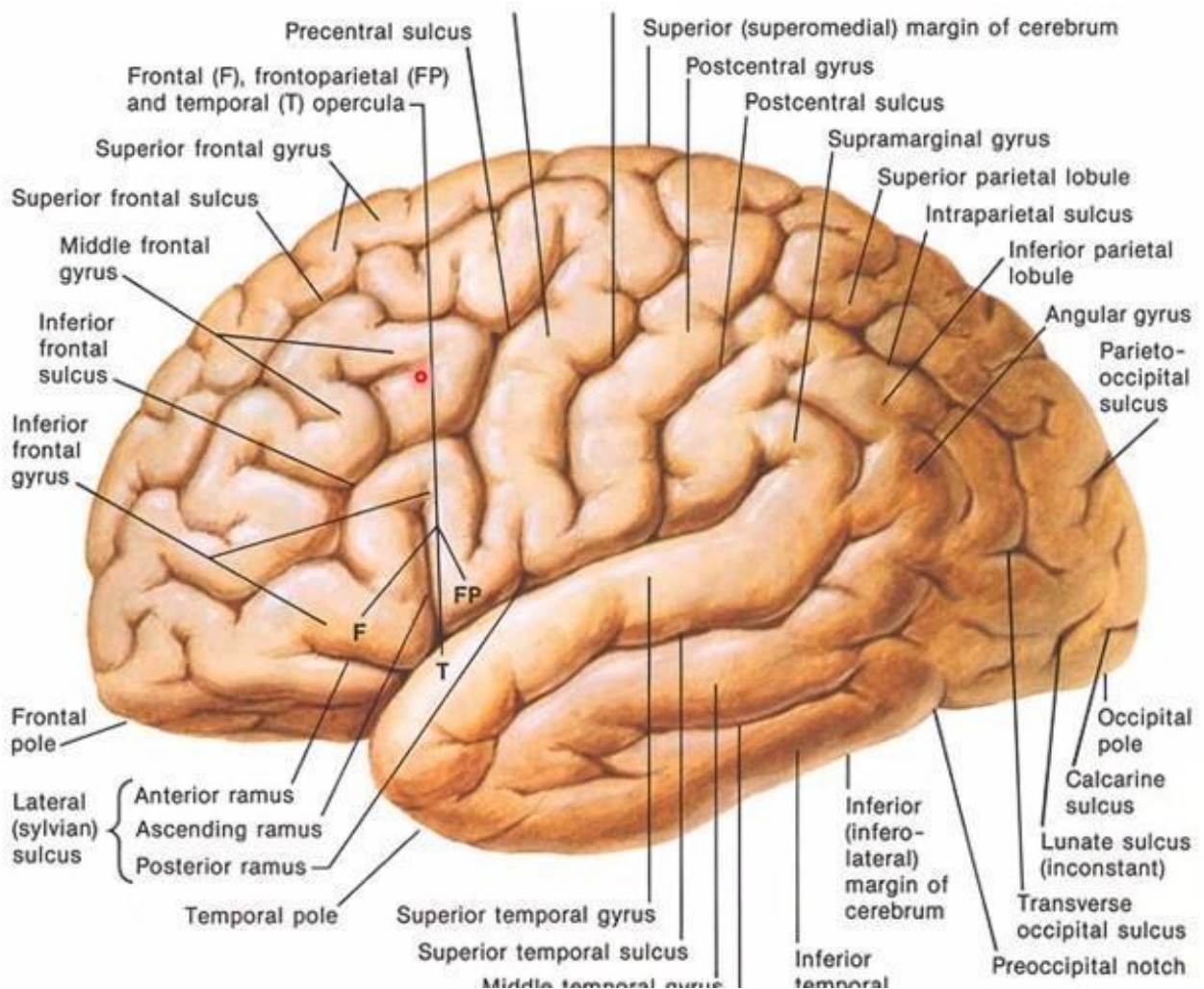
But I want to - before I do that - to just give you a clear picture of what is meant by this distinction. The distinction is analogous to the difference between a television set and its power supply. The idea that the cortex is responsible for the contents and qualities of consciousness, for phenomenal consciousness, the real stuff - the real *qualia* of consciousness - the thing that all the arguments are about -- it's that this is like the television set, and it has to be plugged in, of course. A television set can't do its televisual thing unless it's powered up! But to claim therefore that "the real source of television," and all that that word represents, *is* the power supply, the electrical source in the wall - is ludicrous.

23:31 And that's pretty much how we've all seen - until very recently - how we've all seen what the brainstem contributes to consciousness. So nobody's doubting that it's necessary - that it's prerequisite - that without it you can't have cortical consciousness. But it's been reduced to a mere power supply - something which is not actually *consciousness itself* - in the sense that we're all arguing about.

Case 1: Patients Without Cortex

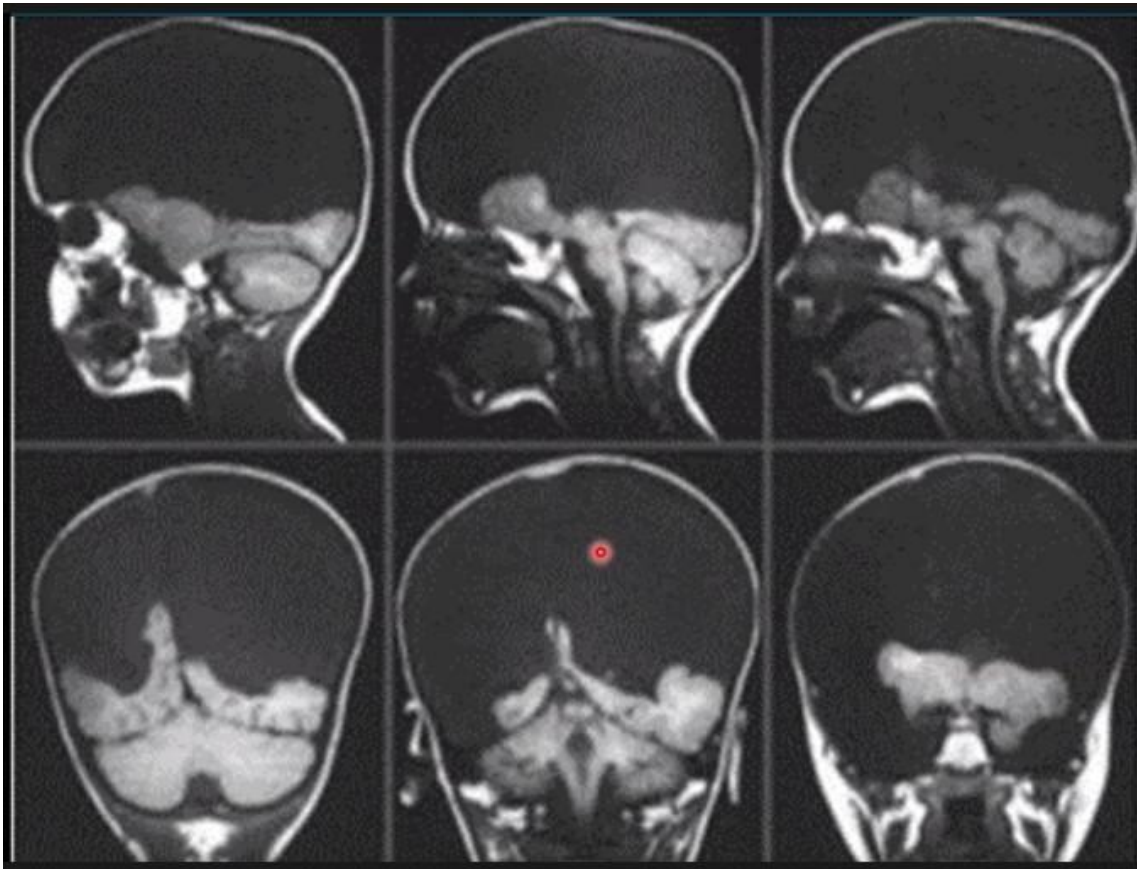
24:02 So, why do I think that that's not correct?

Slide. Labeled diagram of side view of human brain



Well, I'll start with the most dramatic evidence - which is patients who have no cortex. If the theory that the cortex is necessary for the contents and qualities of consciousness [is correct], then patients without cortex should be in a coma. Or at the very least, if we claim that the brainstem produces nothing (other than this sort of background *level*), then patients with no cortex should have some kind of **blank wakefulness**. I suppose the closest natural kind of thing that tallies with "blank wakefulness" is the **vegetative state**, which is also called **non-responsive wakefulness**. In other words, there's wakefulness "without anybody being there" - without there being anything-it-is-like for the organism that has blank wakefulness due to the absence of cortex.

Image. Brain scans of hydranencephalic child



25:09 But when you look at cases who have no cortex -- here's the scan of a girl with hydranencephaly - born with no cortex -- and you examine the girl herself...

Slide. Hydranencephalic girl with infant brother (Merker 2007)



You see that she's not only conscious (in other words, not in a coma) - she's conscious in the sense that she wakes up in the morning, goes to sleep at night - but more interestingly, she is responsive.

Remember my definition, or *the* definition of the vegetative state, the state of blank wakefulness; it is "*non-responsive* wakefulness." And clearly this girl is not non-responsive. Here she's responding to the placement of her baby brother on her lap. And she responds emotionally in a way that any kid -- any little girl, if you place a baby on her lap she'll -- most will express pleasure. Most kids like their little baby siblings being placed on their laps!

Image. Color photo of little girl laughing



26:13 And it's not just her; it's most of these kids.

If they're raised in something other than a state of complete deprivation (like Romanian orphans, who fall into what might look like a vegetative state - although their cortex is entirely intact!) -- If you actually raise these kids in reasonably humane conditions, they are conscious *and* responsive - *affectively* responsive. That's the main point that I'm making: these children with absolutely no cortex - their consciousness has a quality and the content - that quality we call "affect." *It feels like something* to be these kids. And the content is not monotonous affect; it's *all kinds* of affect.

Affective responses by hydranencephalic children

"They express **pleasure** by **smiling** and **laughter** and **aversion** by **fussing**, arching their backs and **crying**, their faces being **animated** by these **emotional** states. Familiar adults enlist their responsiveness to build up **play** sequences, predictably progressing from **smiling**, through **giggling**, to **laughter** and **great excitement** on the part of the children. They respond most vigorously to the voices and actions of their parents and other people they are familiar with, and they show **preferences** for certain situations over others. For example, they appear to **enjoy** specific toys, tunes, or videos, and they even come to expect the regular presence of such things in the course of daily routines. Though behaviour varies from child to child, some of them clearly show **initiative** (within the limitations of their motor disabilities), for example by kicking noise-making trinkets hanging in a special frame constructed for the purpose, or by activating **favourite toys** using switches. Such behaviours are accompanied by **situationally appropriate** signs of **pleasure** or **excitement** on the part of the child."

Bjorn Merker (2007)

27:05 Here's a summary slide extracting what Bjorn Merker - who studied many, many, many of these kids - what he said about them.

They show all of these different emotions. And most importantly of all, they show them in situationally appropriate ways. In other words, they respond with laughter if they're tickled. They respond with arched backs and protests if something that they want is denied them. They respond with a startle if there's a sudden unexpected noise, and so on. So the evidence is that these kids are not only conscious in the sense of being awake - which is interesting in itself.

In fact, I can't imagine what "a state of blank wakefulness" might be like! [Laughs.] But there you have it. The concept derived from Magoun and Moruzzi's work is that "there is such a theoretical fiction as 'blank wakefulness.'"

28:09 But these kids are not blank! These kids are emotionally responsive, in exactly the way that you would expect if they had consciousness, at least in the form of affect. So I'm arguing that, with no cortex at all, these kids display the conscious contents and qualities, in the form of affects.

Now of course, the big methodological problem is that, "How do you *know* what it's like to be these kids? How do you know if there's *anything* it's like to be these kids? They can't *tell* us anything about what it's like to be them!"

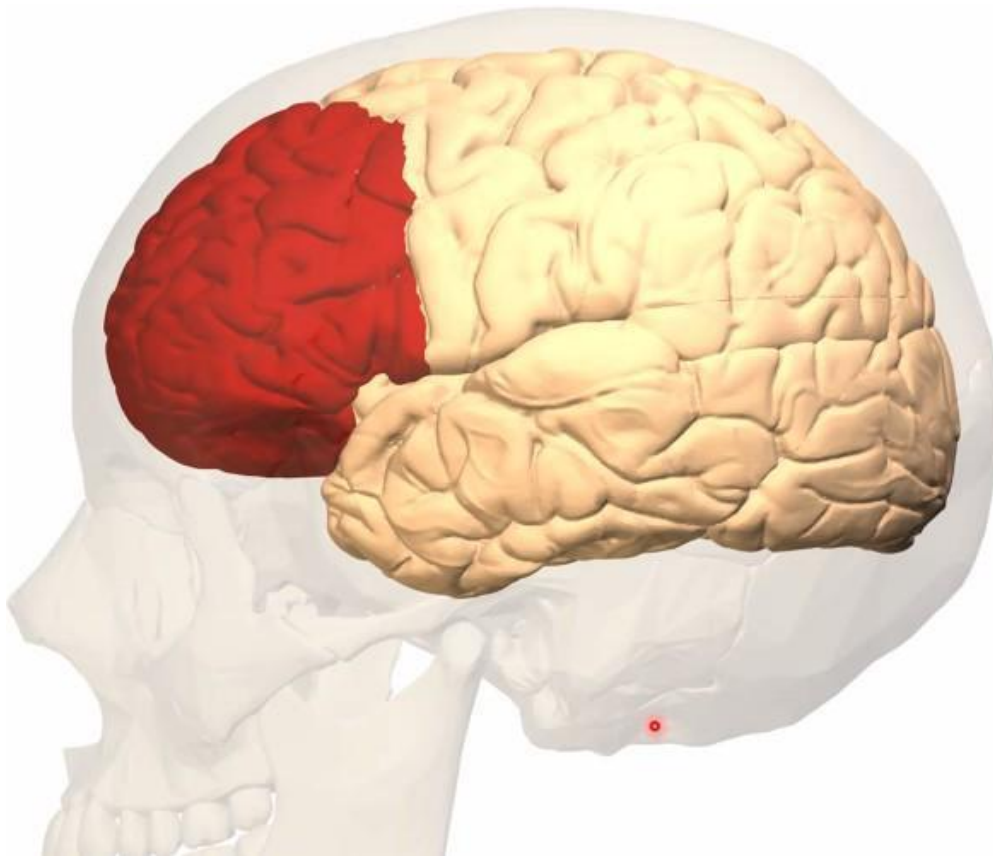
So all you can do in this situation is look at *other lines* of evidence. If we can find multiple converging lines of evidence, converging on the conclusion that the brainstem is where the affects are actually generated, then we can do no better. We can do no more. We can do nothing other than that in science.

Case 2: Loss of pre-frontal cortex or of basal ganglia

29:14 So I want to show you now what happens, if we look at cases who *can* speak, who've lost *great swathes* of cortex.

(They can't have lost *all* of their cortex, because otherwise they wouldn't be able to speak to us.) So I'm wanting to show you examples of cases who have lost great swathes of cortex. And the cortex in question is the cortex that has been most closely identified with **sentient** being, with the subjective presence of a conscious mind.

Image. Drawing of human brain superimposed on shadow skull, with prefrontal cortex (PFC) in red



The first example is of *prefrontal* cortex (PFC). I think that it's fair to say that the mainstream view is that "where all of this information processing comes together, and is received by a sentient subject of the mind -- the mainstream view is that would be in the prefrontal cortex." Now there are few patients there. It's rare to have a patient with absolutely no prefrontal cortex.

Image. Patient W: Two views of human brain with no PFC



But here's one example. And I'm just giving you examples in this presentation. I'm speaking about individual cases - those individual hydranencephalic kids that you just saw -- they are representative of the class of patient that I'm talking about.

Slide: Dialog between Solms and patient lacking PFC

Me: Are you consciously aware of your thoughts?
Patient W: Yes, of course I am.
Me: In order to confirm that, I am going to ask you to solve a problem that will require you to consciously picture a situation in your mind.
Patient W: OK.
Me: Imagine that you have two dogs and one chicken.
Patient W: OK.
Me: Do you see them in your mind's eye?
Patient W: Yes.
Me: Now tell me how many legs do you see, in total?
Patient W: Eight.
Me: Eight? •
Patient W: Yes; the dogs ate the chicken.

30:38 Here is a patient of my own who has no prefrontal cortex, but happily has a sliver of language cortex. And so he's able to tell us what it's like to be him.

I asked him if he was conscious of his thoughts, and he said, "Yes, of course I am." And I explained to him that, according to classical theory, he shouldn't be.

So in order to confirm that he's conscious of his thoughts, I asked him if he would be willing to solve a problem that would require him to consciously picture a situation in his mind. And he agreed to do that. So I said, I want him to picture in his mind, I want him to imagine, that he has two dogs and one chicken. Which he then says, OK, he'll do. And then I ask him, "Do you see them in your mind's eye?" and he says, Yes he does. I then ask him to count the legs. You know, this is not an everyday task - it's something that's unexpected. So there he has these two dogs in his mind's eye and a chicken, and he has to count the legs. And of course the answer is ten. I ask him "how many legs do you see?" - and he says, to my great disappointment, he says "Eight." I say, "Eight?" and he says, "Yes, the dogs ate the chicken!"

31:55 Which, I think, is enough evidence for me that this chap is there! You know - that there's "somebody at home," who's looking at these dogs and the chicken, and imagining (perhaps concretely, you know) that the dogs actually ate the chicken.

Or perhaps he's pulling a fast one - he's making it a joke! And in fact the making of puerile jokes is a very common feature of patients with the frontal lobe syndrome; and this is linked traditionally with the commonplace observation that these patients are *emotionally disinhibited*.

32:40 And why I draw your attention to that, is that, if it were true that "the cortex, and the prefrontal cortex in particular, is where affects are *generated*," then how come, if you have an *absence* of prefrontal cortex, not only is affect preserved, but affect is *excessive*? - There's *more* affect than normal - there's there's *too much* affect! This is entirely incompatible with the idea - espoused by many respectable neuroscientists - Joseph

LeDoux for example, who's a great authority on affect. Joseph LeDoux says that, "of course the amygdala and other subcortical machinery for anxiety doesn't actually generate any *feeling* at all - that the *feeling* of anxiety is 'read out' (to use his term) - 'labeled' - by the prefrontal cortex, and that's when it actually turns into a conscious experience."

33:44 So I say again: If that were true, then how come not only *my* patient - this one making this joke to us - but *all* frontal lobe patients - the *less* frontal cortex, the *more* affect they have? So the less "capacity" they have to "read out and label their feelings" - somehow it seems as if they're *beset by more* feeling, rather than less!

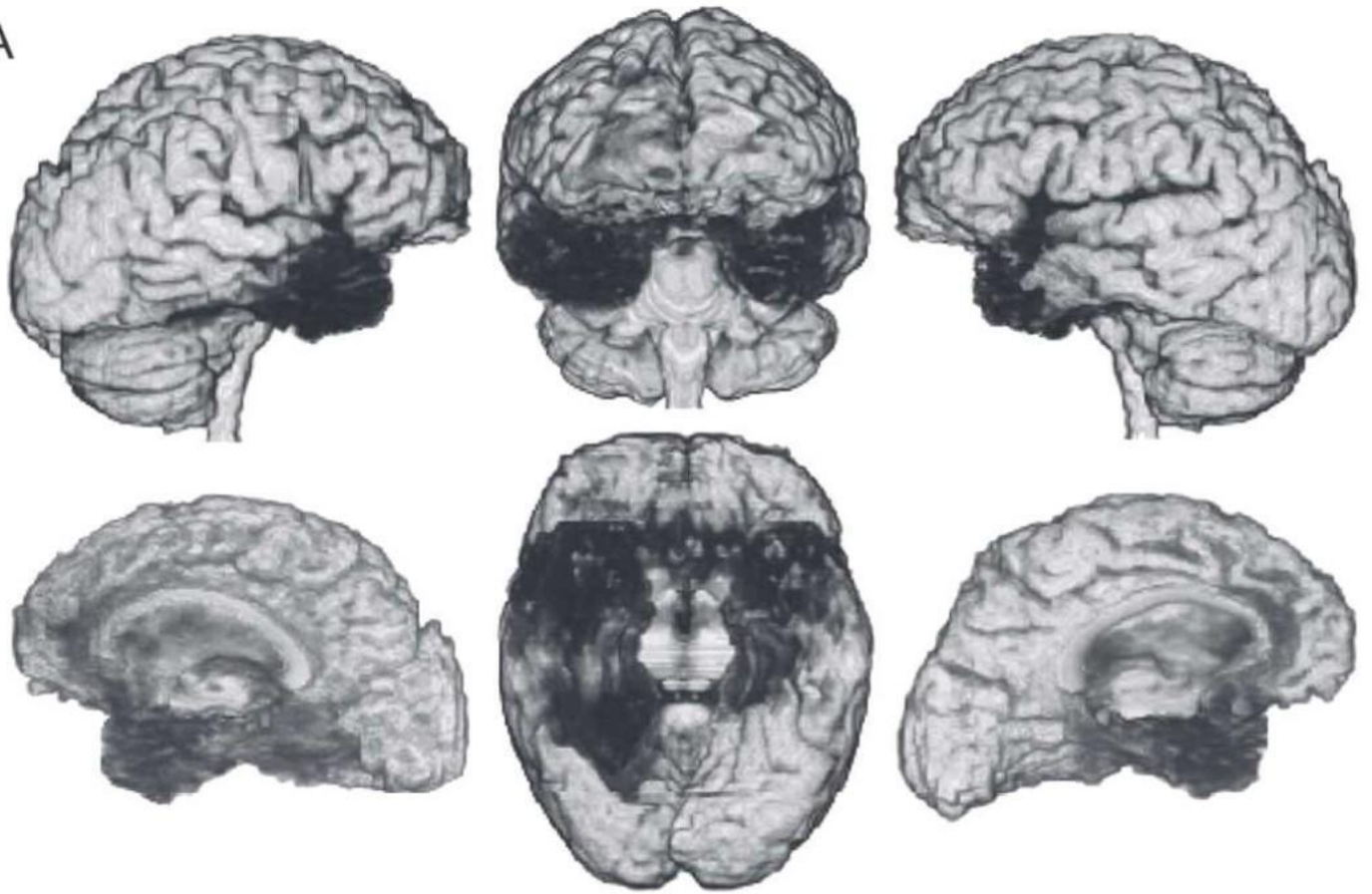
Case 3: Damasio's patient with bilaterally destroyed insular cortex

Image. Painting of human brain sectioned to show insula (in red)



34:12 Since I'm talking about feelings: The other major cortical candidate as being the seat of the sentient self, is the **insula**. Bud Craig claims that "the insula, or perhaps even just the anterior third of it, is where the sentient subject comes into being." And he links this particularly with affect. He says that "this is the part of the cortex which registers the state of **the own body**. '*How do you feel now?*' is registered by this part of the cortex." And it's on this basis that he claims that "this is where the sentient subject comes into being."

A



Three-dimensional reconstruction of Patient B's brain. Figure 1 from Damasio, Antonio; Hanna Damasio; Daniel Tranel (2013)

"Persistence of feelings and sentience after bilateral damage of the insula." DOI: 10.1093/cercor/bhs077

Slide. Interview by Antonio Damasio of patient without insula

Damasio: Do you have a sense of self?

Patient B: Yes, I do.

Damasio: What if I told you that you weren't here right now?

Patient B: I'd say you've gone blind and deaf.

Damasio: Do you think that other people can control your thoughts?

Patient B: No.

Damasio: And why do you think that's not possible?

Patient B: You control your own mind, hopefully.

Damasio: What if I were to tell you that your mind was the mind of somebody else?

Patient B: When was the transplant, I mean, the brain transplant?

Damasio: What if I were to tell you that I know you better than you know yourself?

Patient B: I would think you're wrong.

Damasio: What if I were to tell you that you are aware that I'm aware?

Patient B: I would say you're right.

Damasio: You are aware that I am aware?

Patient B: I am aware that you are aware that I am aware.

35:05 So again, in just one case, Victor [sic] Damasio's, who had absolutely no insula, bilaterally obliterated and he interviewed him in much the same way as I interviewed my patient with no *frontal* cortex. And Damasio asks him if he has a sense of self, then he says Yes, he does. I won't go through all of this, all of which is Damasio trying to ascertain whether or not this person is conscious, and in particular whether or not he possesses selfhood. Does he even understand what selfhood is?

And eventually Damasio says, "What if I were to tell you that I know you better than you know yourself?"

Patient says, "I would think you're wrong." Damasio says, "What if I were to tell you that you're aware that

I'm aware?" Patient says, "I'd say you're right." Damasio says, "You're aware that I'm aware?" Patient says, "I

am aware that you're aware that I'm aware."

Please note all these references to "I!" There is a sentient subject speaking to us here, with absolutely no insular cortex.

36:15 So I think the claim on the basis of the hydranencephalic children, the claim that the cortex is the seat of consciousness and of *feeling* in particular, has to be dismissed.

But they can't speak. So we look at the other major candidates who *can* speak - patients with no prefrontal cortex. They are conscious; there's a sentient subject there; and they feel like something. They *tell* us, and they *show* us. And likewise patients with no insular cortex.

Of course, the claim can be made, "Well, in these cases their consciousness arises from the *remnants* of the cortex - the remainder of the cortex - the cortex that is *not* removed." Bear in mind: the ones who have *no* cortex, who clearly show behavioral displays of affect - every prediction that you might make from the hypothesis that "it feels like something to be these kids" is confirmed by their behavior, by their responses to affective stimuli.

37:22 And in that case we say, "Well, we don't know, because they can't *speak!*" So then we use these other cases with no prefrontal cortex, no insular cortex, who *can* speak - and then we say, "Well, it must be the *other* bits of the cortex that are generating the consciousness!" It seems like a losing wicket to me!

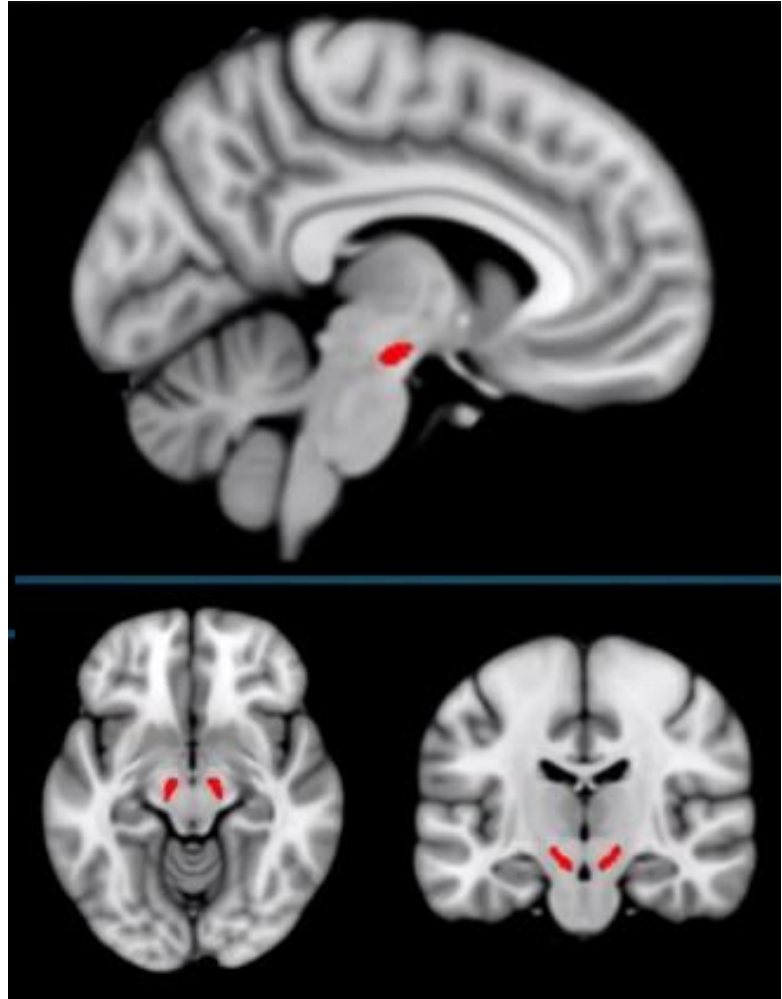
Case 4: Deep brain stimulation of the substantia nigra

37:41 But let's look at other evidence than lesion evidence. Remember we're looking for *converging* lines of evidence, converging on the conclusion that the feelings are generated in the brainstem - *in* the brainstem.

Substantia nigra marked in red

And so here we look at the case of a patient where an entirely different method is used - that is, deep brain stimulation of a part of the reticular activating system, in this case the **substantia nigra**. The surgeon was aiming for the subthalamic nucleus, but the electrode went into the substantia nigra. And when he stimulated it, the patient fell into a profound depression within five seconds.

38:33 If you read it yourself, you'll see this patient within five seconds fell into a profound depression. As she has no psychiatric history, never been depressed before, never had a diagnosis of any kind, let alone a diagnosis of major depression -- and here within five seconds you can generate - produce - cause a suicidal depressive state -- which within 90 seconds of switching off the stimulator -- after the stimulation was stopped, the depression stops.



Slide. Interview with deep-brain-stimulation patient (Blomstedt et al, 2008)

"The patient's face expressed profound sadness within five seconds [...] Although still alert, the patient leaned to the right, started to cry, and verbally communicated feelings of sadness, guilt, uselessness, and hopelessness, such as 'I'm falling down in my head, I no longer wish to live, to see anything, hear anything, feel anything ...' When asked why she was crying and if she felt pain, she responded: 'No, I'm fed up with life, I've had enough ... I don't want to live any more, I'm disgusted with life ... Everything is useless, always feeling worthless, I'm scared in this world.'
[ctd. ...]

[... ctd.]

When asked why she was sad, she replied: 'I'm tired. I want to hide in a corner ... I'm crying over myself, of course ... I'm hopeless, why am I bothering you' [...] The depression disappeared less than 90 seconds after stimulation was stopped. For the next five minutes the patient was in a slightly hypomanic state, and she laughed and joked with the examiner, playfully pulling his tie. She recalled the entire episode. Stimulation at other sites did not elicit this psychiatric response."

Blomstedt et al (2008)

39:10 Very bravely, she agreed to have the stimulator switched on and off, *either* in the subthalamic nucleus *or* in the substantia nigra of the reticular activating system, where she was blind as to where the stimulation was being applied. And every time it was applied in the reticular activating system, she fell into this severe depressive state again.

So this is what you would expect -- this is the prediction from the hypothesis that affects are actually generated in the brainstem. The prediction would be, if you stimulate these reticular activating nuclei you will generate affects; and that's exactly what you do.

39:59 This is also why I mentioned the periaqueductal gray earlier. Because, if you stimulate reticular activating nuclei, you generate intense affects. But the same applies to periaqueductal gray. In fact, the greatest intensity and range of affects anywhere in the brain can be elicited from deep brain stimulation of the periaqueductal gray. Brain stimulation of the cortical mantle produces very little if any affect.

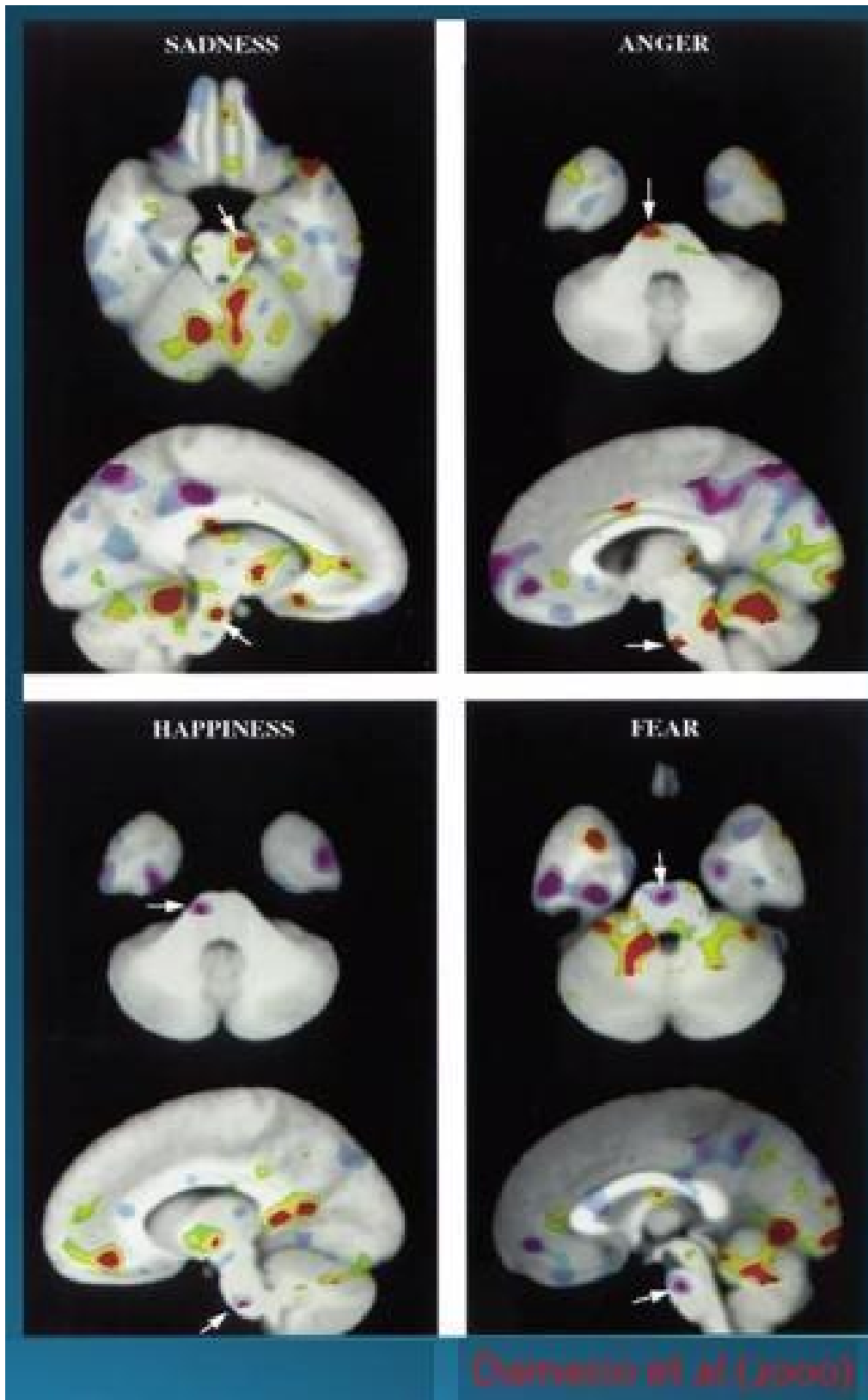
Human brain during four affective states

40:29 Another method: Positron emission tomography.

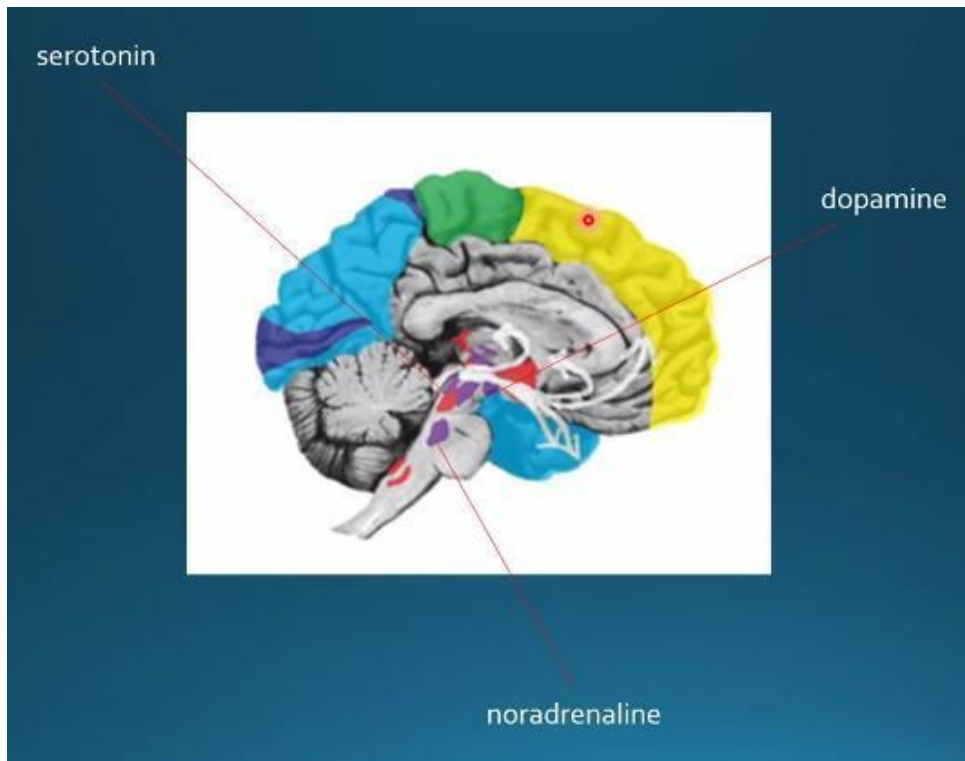
So, another study of Damasio's. Patients in states of anger, sadness, happiness, and fear. You see where the major brain activity is in all of these affective states. The major activity is subcortical. It's in the brain stem. It's not cortical.

So again, this is exactly what you would predict, if the affect is actually being generated *by the brain-stem*. Not being "read out" by the cortex, but generated by the brainstem.

Colorized human brain; two images each during four affective states (Damasio et al 2005)



Slide. Colorized human brain labeled by activity of serotonin, dopamine, and noradrenaline



41:06 And then there's also the matter of the neurotransmitters that psychopharmacologists tinker with. The mainstream psychiatric medications modulate neurotransmitters that are sourced in the reticular activating system. We don't normally think of it that way; but *it is simply the case* that serotonin, which we increase with antidepressants, is sourced in the reticular activation system. Dopamine, which we block with antipsychotics, is sourced in the reticular activating system. Noradrenaline, which is blocked in many anti-anxiety drugs, is sourced in the reticular activating system.

41:51 If the reticular activating system was just the "power source" that I referred to in my analogy about the television set, then it should only interest anesthetists! You shouldn't be able to change the channel - alter the programming of the television set - by pulling out the plug, or by attenuating the power supply! You might expect the screen to go blank, or just flicker -- but not to have whole new programs, with *new contents* and *new qualities*!

And so this is my argument: That the idea that *feeling states* - affective feeling states - of the kind that those children display, of the kind that these patients describe, of the kind that psychiatrists are directly trying to manipulate - are generated by upper brainstem structures. They're not generated by cortex. You don't need to *have* cortex in order to have these states!

42:53 And yet -- we all know and agree - *we all agree* - that *cortical consciousness is contingent upon reticular activating arousal*. If we now are forced to accept that reticular activating arousal is not contentless - is not a mere quantity - it's not just a *level* - but rather has content and quality that we call *affect* -- then we have to accept that *affect* is a prerequisite for all other forms of consciousness.

And it's on *this* basis, not on semantic one, that I'm claiming that affect is the foundational form of consciousness.

43:31 It also makes good *evolutionary sense*, that this ancient structure, rather than the cortical mantle (which evolves only with mammals) -- that the ability to register your own subjective state - how you're doing, affectively, in terms of hunger and thirst and sleepiness and too-hotness and whatnot -- that this should not require such a late evolutionary development -- but that's another point. I'm just wanting to slot in there that I also mean it in the evolutionary sense, when I say that affect is the foundational form of consciousness. I don't only mean it in the functional sense.

44:14 Now, I'll pause there, because that's a big claim, and a very controversial claim. In fact, I must tell you I'm still slightly *dismayed* that it remains controversial! I've been really surprised by the degree of pushback that I get when I argue... And I'm not the only one who argues it - I'm not the *first* one to argue it! But Damasio argues it; [Jaak] Panksepp was the first to argue it; [Bjorn] Merker argues it; and here I come, a good few years on, and argue the same thing, citing all of this evidence; and -- wow! People just don't want to accept it!

So let's hear what you have to say.

Question: "Philosophical zombies?"

45:08 *Friedman:*

I'll read a question from the chat first; and then we'll go to Stephen. Jonathan, in the chat, wrote, "I can imagine a modern large language model (the likes of GPT-3) creating such conversations, without apparently having a mind's eye. While I agree it is good evidence, I think we are seeing machine learning systems having similar conversations."

So: You presented case studies that are the tip of the iceberg of broader populations and analyses. How do we think about our evaluation of conversational dialogue, in questions about necessity and sufficiency, here?

45:48 *Solms:*

Yeah. You see, this is a kind of a relative of the "philosophical zombie" argument. The problem with this sort of thing, is that - in fact, the same applies to you and me! *I can't know for sure* that *Daniel* is conscious, that Stephen is conscious, that David is conscious, or that the person who asked that question is conscious! You know, the point you've just made really does apply in exactly the same way! There's no greater reason to believe that it applies to the patients that I've just presented to you, than that it applies to you and me! Because you too could be a philosophical zombie! In other words: You too could *look* as if you have internal states - *say* that you have internal states - but actually there's nobody there! *I can never disprove that!* That is a *metaphysical* problem. That's not a scientific problem.

46:52 The scientific method requires that your hypothesis has to be consistent with all the known facts. And then you have to make **falsifiable predictions** from that hypothesis. And if your predictions are not falsified then they are *upheld*, unless and until some evidence is found which *positively* - (this is what *positivism* is all

about!) - which *positively falsifies* it. So I'm saying these kids look as if they're conscious, you know? They *behave* as if they're conscious! Now, it's true that they might not be and Stephen might not be and Daniel might not be and David might not be. But, you know, my hypothesis is that they are!

47:38 So what do I do? I say, "Well, if Daniel feels pain like I do, then when I stick a pin in his hand he's going to go 'ouch!' and withdraw his hand, and avoid me in future." And that's a confirmation of my prediction. The same applies for these kids. But (you know) that's just one source of evidence. As I said, you don't accept just one. You then say, "Okay, well, what about this line of evidence? - What about that?" And I've given you *five* lines of evidence here, in which, in every instance, the prediction from the hypothesis that these people are conscious; and that the consciousness is generated in the brainstem, and that it has an affective quality -- every single line of evidence, the prediction is confirmed! What you would expect would happen *if* that were true, *is* what happens! So I'm afraid you can do no more than that in science.

48:41 And it's a curious fact that we raise the bar of evidence when it comes to the problem of consciousness! I think we must be very careful not to require any different sort of evidence in the case of consciousness than we require for any other matter in science, and in this case - in any other biological function. If you start with the assumption, that "consciousness is something *special* that the ordinary rules of science don't apply to," then you're going to have a self-fulfilling prophecy.

So, *yes!* - It's possible. It's possible that these kids and those patients are not there - that they're lying to me, that... You know - It's a little bit of a curious thing, that somebody without a sense of self wants to pretend that they've got one! I think you're in a kind of slippery slope there to irrationality! But I can't prevent you from holding that view. It's just not consistent with any of the evidence, I'm afraid.

Question: The problem with uptake

49:59 *Sillett:*

I was wondering how much you think the challenge with *uptake* is? Because people are sort of stuck in the idea that knowledge and information is a *signal* coming in. And *meaning*... "It's all about meaning, and the ability to communicate *meaning*" -- as opposed to engaging in meaningful *action* and skillful, enacted processes of the voice -- twitching the voice box, twitching and enabling words to be in the space around us, in some sort of way? And whether that issue is kind of a deeper, ingrained challenge to overcome?

50:43 *Solms:*

Well, I completely agree with that! I think that that is a fundamental source of the problem. Because, our common sense tells us "our consciousness is exteroceptive." As I said earlier, our conscious experience is dominated by visual qualia. And where do they come from? From (you know) things - light waves impinging on my retina from outside! And I'm seeing things *out there!* And so consciousness seems to be so obviously derived from something *out there*.

51:21 And this was at the heart of the experiments that Magoun and Moruzzi did. Magoun and Moruzzi weren't expecting to find what they found! They were expecting that a cat deprived of any exteroceptive stimuli will fall asleep, if not fall into a coma. And it was therefore a great surprise to find that that's not what happens, that the consciousness is generated endogenously - it comes from deep within the brain - doesn't flow in with sensory stimuli. And when they lesioned that core brain-stem structure, the lights went out! And they couldn't make sense of it!

And they came up with this theory -- which as I've shown you (at least as I've shown to anyone who's willing to look at the evidence and draw the conclusions), that the way that they framed it is wrong! It *doesn't* explain what they saw. What they saw was that consciousness is *endogenously generated*.

52:27 And where does the notion come from - this common sense notion, that it flows in from outside? It comes not only from our own everyday experience; but also from the British empiricist philosophers, who probably themselves derived it from common sense -- that "the mind - literally, the mind as a whole - is derived from sensory experience that leaves *impressions*, leaves memory images impressed upon our cortex, which are then associated with each other." And this, according to British empiricist philosophy -- which underpinned the *whole* of 19th century and early 20th century neurology and neuropsychology -- it's *all* predicated on that idea.

So I think it's *deeply* ingrained within ourselves and within our science, and within public perception generally, that "consciousness comes from outside."

53:32 And you're speaking about some *further* ramifications of that, and therefore sources of resistance against the alternative view. But *all the evidence suggests* that consciousness is endogenous. Consciousness is generated from within.

53:50 I'm sorry - I don't know who the person was who made the comment about, "How do we know that this is not just some sort of program spewing out language?" I have nothing against you - I don't know who you are [chuckles] -- I just hear that all the time!

And I find it frankly astonishing, that you present all of this evidence -- and then still people say, "Well, how do you *know*? - But how do you *know* that these patients are really conscious?" And this demonstrates the *extent of the prejudice* that has come with that historical legacy that I've mentioned. So, yep, I completely agree with you.

Question: Cerebral palsy and other neurological conditions

54:47 *Sillett*:

Okay - thanks. So sort of following on from that - yeah! I totally agree. And actually I think the work with Karl Friston and active vision is quite helpful. There's the idea that we're reaching out and *feeling* the world. And back before 300, 400 years ago, people actually almost had that as an intuition, that we were looking *out* in the world. But once we discovered the retina, and then photography, we kind of *re-imagined*. So one question I've got is regarding cerebral palsy. I work a lot with communities with complex disability, or non-verbal. And their premotor cortex is often damaged. And in the 1970s it was thought, medically, if you couldn't speak you couldn't think. So that was where they were then - if you couldn't communicate words...

55:37 And then once they started using *symbols* - Blissymbolics - and they suddenly discovered that all these clients who were "non-verbal" actually could form complete sentences - because they have a way to *act* in the world, and interact in the world, and develop language, syntax, and all of that.

So I'm just thinking, whether any of this work's been tied in with clients with cerebral palsy.

56:06 *Solms*:

Not only with cerebral palsy; but with all sorts of neurological conditions! If I can just sort of piggyback on what you've said about cerebral palsy: The whole trend within clinical neurology has been us realizing that patients who we thought were not conscious actually *are* conscious. And the whole *tenor* of the research that's been done in recent decades... **Locked-in syndrome**, of course, is the *extreme* example. But I don't only mean locked-in syndrome. I mean all of these states of reduced and obtunded and clouded consciousness and so on - minimally conscious states, so-called vegetative states, and so on.

We're finding more and more patients who we thought were not conscious, with increasing technology we find that they are; with improvements in our methods we find that they are.

57:10 So it's a sort of an odd contradiction, with the sorts of philosophical skepticism and doubt that one confronts from people who are saying, "Well, how do we know they're conscious? They're *behaving* as if they're conscious; but maybe they're *not!*"

In fact, what we've been finding in clinical neurology is patients who behaving as if they're *not* conscious; and then when we use the proper tools we find that they actually are.

57:40 So your comment about cerebral palsy can be multiplied. The observations that you refer to... Using "reportability" as a criterion - the capacity to "declare your conscious states" - has really not just led to methodological and epistemological problems. It's really led to *gross clinical errors*, you know - it's *ethical* problems! I feel we need a lot more than mere philosophical skepticism before we can take seriously in neurology - which, after all, is a branch of medicine! - before we can take seriously the possibility that these kids and these patients *look* as if they're conscious, *behave* as if they're conscious, *say* that they're conscious, and we say, "Well, we can't be sure." I mean, think about the implications!

Question: Hypervigilance and resistance

1:00:25 *Douglass:*

Thinking about the resistance you've experienced, to the discovery that the source of consciousness is sub-cortical:

(Putting aside *rationaly* cautious demands that "such an extraordinary claim must be supported by extraordinary evidence") -- Do you suspect that there are deep-rooted problems that cause *some* folks to reject this discovery? - maybe problems along the lines of hypervigilance and self-distrust, such as we see (for instance) in too-compliant, too-other-directed children of excessively-demanding parents? [Paraphrased for clarity]

1:01:51 *Solms:*

Well, I will just say: Please note how *affective* all of this stuff is, that you're talking about. And I think that we really underestimate the importance of affect in terms of the whole gamut of phenomena that we call consciousness.

Question: Coma and evidence for consciousness

1:02:17 *Douglass:*

Some caregivers for people who are in the death process, have stated that presumably locked-in or comatose patients seem actually to be operating on extremely *slow* time-scales. If one speaks simply and slowly, and repeats a story or request, the patients seem eventually to tune in, and to respond appropriately. Is this real, rather than wishful thinking? [Edited for clarity and length]

1:02:18 *Solms*:

I want to make clear that I agree with what you're saying about the inference of consciousness in these patients. But I want to make clear that I'm making actually quite a... In light of, or notwithstanding, all of this other evidence, and all of these other very interesting clinical phenomena - all of which point in the direction of *we've underestimated the presence of consciousness* -- you know, that we've been relying on very poor criteria, in terms of reportability and so on - that really we need more sophisticated methods. And the more we use more sophisticated methods, the more we find evidence that there is in fact somebody there. I agree with all of that.

1:03:20 But I'm making a very, very limited claim. I'm not saying these kids are like you and me. I'm not saying they have a full, rich inner mental life. I'm saying only that there is something it is like to be them. And in light of all the evidence in *their* cases, plus all the other evidence that I've cited, or using all these different methods, all I'm saying is that that's something-it-is-likeness that these kids possess - is it feels like something! They have feelings! That's all I'm saying.

So it's all very interesting, all these other things about how all these other capacities have been missed by us in relation to all of these other neurological conditions. But in these cases which have *no cortex whatsoever*; they have only... Some of them have basal ganglia and other subcortical forebrain nuclei. Some of them don't. But even in cases [where] you have no forebrain, only a brainstem, they display affects.

1:04:30 And so I'm saying that *the brainstem-prerequisite form of consciousness* does have a quality and a content. It therefore feels like something - there is something it is like to be these kids. And that quality and content is affective. That's all I'm claiming - that the foundational form of consciousness is affect, and it's generated in the brainstem. So it's actually, I think, a rather modest claim, in light of all the evidence. It's quite a conservative claim! It's a sort of a "deflationary" claim. But so much flows from just that!

1:05:12 And I'll just say once more: I find it very hard to understand -- I mean I *can* understand it *psychodynamically*, of course, or politically, or what... I mean, people's whole careers - I mean *generations* of careers - have been built on the assumption that cortex and consciousness are intrinsically bound up with each other. Which, you know, certain people have a Hell of a lot to lose! And there many of us with egg on our faces. And I freely admit that *I too* believed that the cortex was the organ of consciousness. This is what I took in like mother's milk when I trained. And I believed it for many, many years after I trained, you know? So I'm not exempted from that.

1:06:00 But, Gosh! - we've got to look at the evidence! It's wrong! We made a mistake, you know! It doesn't mean the cortex does *nothing*! it doesn't mean it doesn't contribute *anything* to consciousness! It's just not the *source* of consciousness. The cortex *elaborates* this very raw, primary, elemental *feeling* that is generated in the brainstem. And all we need to do is accept *that* - and then we'll make huge progress in consciousness science! The fallacy that "the cortex is the source of consciousness" is really a serious impediment to us at this point in the history of our field.

Question: Our inherited priors and new evidence

1:06:48 *Friedman:*

Yes. It is indeed a theme of the discussion, that ambiguous evidence are continually interpreted and reinterpreted in light of our inherited or our previously structured priors.

And that's true in the active vision example. It's also true in the scientific example, where our inherited priors about scientific models influence how new evidence is not just perceived but funded and carried out.

And it made me think about -- when people activate a gene, or remove one gene, and make claims of function - might be analogous to the brain. But removing or having one node in a complex network - showing that there is a single point of failure or gain of function - does not entail that as a generating mechanism -- which is just an example of an explanation that retains an alternative perspective.

1:07:47 And so I think it's interesting on multiple levels, as we apply this idea of updating and learning and structural learning on our theories, about ostensibly the part of the body that does this function -- as opposed to one's hand, which is important and has nerves in it, but doesn't do this kind of engagement.

1:08:12 *Solms:*

Thanks very much! I agree. That's an everyday observation -- that we are reluctant to accept prediction errors.

We cling to our priors, especially our most deeply consolidated priors. And we see this in everyday life.

Nobody likes to find they were wrong, that they have to change their mind, that they have to accept evidence for a view other than the one that they had always held.

And this applies even to our scientists -- who in their graduate studies, when they were doing their first research, didn't -- Hope that their hypothesis was *right*, and then when the evidence came in, and they found, "Oh, dear! I haven't found what I expected to find!" -- it's not welcome!

1:09:08 So what I'm referring to, based on what you've just said (I know this is not quite what you were saying - I'm sort of elaborating what you were saying)... is that we don't recognize sufficiently the affective dimension in our resistance to prediction errors. We all know about priors, and posteriors, and how this all works mechanistically. But just from the sort of anecdotes that I've mentioned -- I'm trying to get us to recognize, that there is an affective dimension to that. And when we get to this point (9), about the **concentric predictive hierarchy**, I'll come back to that.

1:09:54 But I think it's an interesting and important observation that -- as you traverse the hierarchy, and get to (what in my way of thinking about it, as a *concentric hierarchy*) - to get deeper into the hierarchy (in the traditional way, we speak of going *higher* in the hierarchy) - but as we go deeper, further from the sensory-motor periphery, so we come upon priors which are more and more difficult to update. And I say again, there's a feeling that goes with that, which is, "I don't care what you tell me! I'm not gonna believe you!" [Laughs.]

Question: Affects - ascending? descending?

1:10:45 *Sillett:*

I know you're going to come back to the cognitive consciousness. But just say - around terminology. We've been doing work around an ontology working group at the Active Inference Lab, actually trying to understand how to think about the language. And you mentioned it could be affects - it is *generated* in the brainstem. And that's when something's like something - that's what it's like.

But I'm also wondering, how are you taking the dynamic with the body and environment, going *down*? I don't know if that's seen as another manifestation, because obviously - the brainstem, it makes sense, because it's more integrated with the rest of the body's nervous system. It's less distant from it. But I'm wondering whether the term "*generated within*" is more to do with when it's connected to the rest of the brain? And when it goes down, it's actually more *distributed* in the mind-body-environment dynamic?

1:11:45 *Solms*:

That way of thinking is... I'm very sympathetic with it. I'm not a narrow localizationist. I'm very, *very* happy to recognize... it comes naturally to me to think about functional systems which are distributed and dynamic. And in fact the reticular activating system is probably the most deeply interconnected with everything else in the brain system that there is! It has tentacles everywhere! And not only... Although it's called the "ascending reticular activating system," it *descends* also, you know - it *modulates* the whole nervous system, in very important ways that we'll come to - because this whole thing of **precision modulation** is at the heart of my way of understanding consciousness.

1:12:49 So the main thing that I'm wanting to say is, I agree with you that this is a deeply interconnected system which functions by modulating *other* systems.

And I would be very happy if we had to leave it there, and say, "well, you can't quite say that affect is generated *in* the reticular activating system, because, you know, it's really modulating these *other* systems." My reluctance, however, *in this case*, to accept that - is twofold. The one is, that that's the kind of thing that people like Joe LeDoux are saying. They're saying, "Yeah, this is all unconscious, until it is labeled by, or read out in, cortex. You're not talking about cortex. So, you know, that's...!"

1:13:53 But the principle remains: That I don't want to concede the point that, "it only becomes conscious once it modulates something *else*; and that that's *really* where the contents and qualities are." That's a sort of theoretical reluctance on my part. You can see why. But more importantly, and more tellingly, is the evidence! When you look at these kids, they *have* no cortex! So their affects can't be modulating... It can't be the result of something that the brainstem is doing *to cortex* - because there ain't no cortex! And yet (you know) there's all the evidence that they have affects!

1:14:40 And so, I think, as surprising as it may be - it's not consistent with our normal way of thinking about cortex. And it's also not something that I'm wedded to - the idea that *anything* is generated *in anything*. But the evidence suggests that certainly the modulatory influence that the reticular activating system has over the cortex - true as it is, important as it is, fundamental as it is to how that system works - that the affects are not a product of the modulation of cortex. The affects are generated *in* the subcortical structures.

And if you just take a step back. The implication of that, is that: Well, maybe there are affects in vertebrates who don't have cortex! In other words, maybe affects are not the exclusive preserve of mammals! And, actually, that seems pretty likely. It seems pretty darn likely that non-mammalian vertebrates have affects. And again, when you start looking at the evidence, it's pretty darn good that they do.

1:15:54 So I'm happy to accept everything that you're saying, about "let us not forget that this is a system which is deeply interconnected with other systems; that its primary function is to modulate other systems." But all I'm saying is that, when it comes to cortex, the affects themselves are not generated by the modulating of cortex. That doesn't mean that there's a "disembodied brain," or "disembodied reticular activating system." But you can have a *decorticated* reticular activating system, and it can generate affects nonetheless.

Claim 3: Affect is not synonymous with interoceptive inference

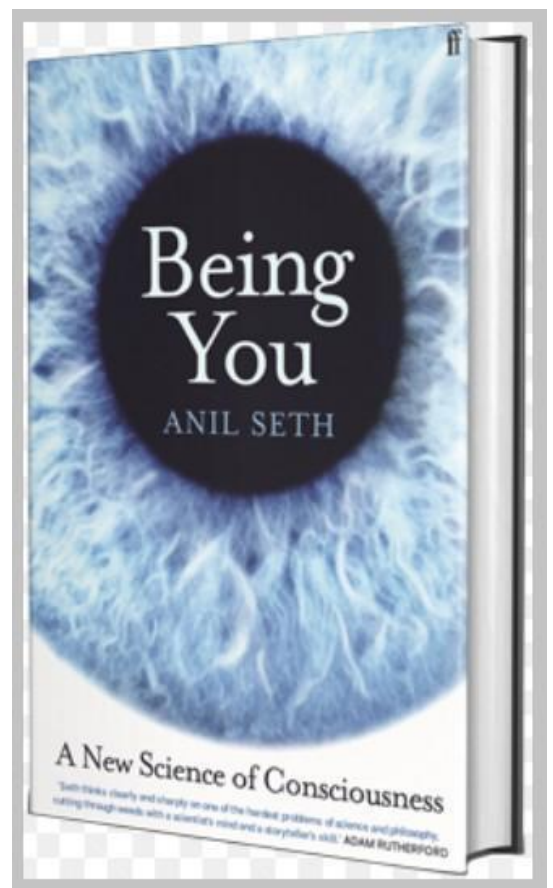
1:16:50 *Solms:*

You know, the reason I'm making this point, about affect not being synonymous with interoceptive inference, is because those few people working within the Active Inference framework who are *not* corticocentric (and, happily, they are increasing!) -- There's people who worked within this framework - they had no reason to question all of the mainstream assumptions of cognitive neuroscience as far as cortex is concerned.

1:17:28 But slowly, it's become apparent that affect is an important topic -- that we're not only talking about perception and action and cognition. We're talking also about affective processes, and about body-regulating processes, and so on. In fact, the whole of self-organization has to do with the regulation of the own body, and the whole notion of self-evidencing systems, and so on. So this is, happily, slowly, and *increasingly* coming into the mainstream of the people working within the Active Inference framework.

1:18:17 But the dominant view within that school of thought is that "affect is to the *internal* body what perception is to the external world." In other words, "exteroception is perceptual inference regarding things going on *outside* of me; and *affect* is perceptual inference about things going on *inside* of me." And if that were so that would be nice and simple. But I don't see it that way.

1:18:59 Perhaps the outstanding proponent of the view that "affect is interoceptive inference" is Anil Seth. And he's just brought out a really very good book! I have it right here - "Being You" - which is predicated on that view. Because I agree with him about so many things, I think that it's easy for people to conflate our views, and to imagine that what I'm saying is the same as what he's saying on this score. So I want to make clear that I'm *not* saying the same as what he's saying on this score.



1:19:45 But why do I think that affect is not the same as interoceptive perception?

Well, first of all: Because there are many interoceptive perceptual states which are not affective. You can perceive your heart rate, the beating of your heart. You can perceive your respiration rate, or the inspiring and expiring of air from your lungs. You can perceive gurgling in your tummy, which is an unequivocally visceral

state. And all of these things - you can perceive them indifferently. You can perceive them without any affective valence or quality at all. So if affect were synonymous with interoceptive inference, then why are these common-and-garden examples of interoception not affective?

1:20:54 And then, secondly: If affects were interoceptive perception, how come so many affects are generated by *exteroceptive* events, where there doesn't seem to be any clear interoceptive driving force?

1:21:13 For example, pain [chuckles], which is a fairly simple example of an affect. You stick a pin in my finger; it comes from the outside; and it penetrates the nociceptors in my exteroceptive nervous system; and I feel this intense affect called pain. So that's exteroception - and yet it's affective! The same applies to surprise - to fright - you get a shock! - (Well, in fact, electric shock is another one, but...) But I mean just, you know, fright - something unexpected happens outside of you, and you startle -- and there's an affect that goes with that - a particular type. And *disgust*, is another one: you imbibe something that tastes or smells bad - and you wrutch it up with this [makes wrutching noise] terrible *feeling* of disgust.

1:22:13 And then there're more -- there're purely, entirely non-bodily affective states, like missing somebody -- you know, missing your brother. It's clearly affective, and there's nothing bodily about it, and nothing interoceptive about it. (When I say "nothing bodily," I mean, particularly, it's not *about* the state of your interoceptive visceral processes.)

So I'm saying: there *are* interoceptive visceral processes which we perceive, which are not affective; and there are affective processes, affective states, that we experience, which are not interoceptive. So I don't think that we can say that "affect is interoceptive inference, that it's the same as external perception except it's directed toward the internal state of the own body." I think that it's something else.

1:23:16 And I don't say this to be difficult! I'm actually quite concerned that those few of us within the Active Inference framework who *are* recognizing the importance of affect can't agree on this point! [Laughs] Because there are few enough of us who recognize the importance of affect. So I'm just wanting to make clear *why* I think that it's *not* merely a kind of mirror image reflection of exteroceptive inference.

1:23:55 And then I want to add one last point, which is that: Everything that I'm going to go on to argue, about why exteroceptive perception... Remember what I said at the very beginning, about cortical vision having been "the model example of where we will find the neural correlate of consciousness" -- which has led us down this blind alley, and got us into such a mess -- that it has to do largely (as I'm going to argue later) with the fact that cortical processes and perceptual processes are *not intrinsically conscious things*. You can perceive the world through *all* your exteroceptive modalities, vision included, *at the cortical level, too* - you can perceive the world without perceiving it *consciously*.

1:24:46 In other words: exteroceptive inference, perceptual inference, is not an intrinsically conscious process! There's nothing about exteroceptive inference, or perceptual inference, that requires consciousness, for us to understand the mechanisms at work. And so, if interoceptive inference is just the same thing applied to the *interior* of the organism, it suffers all the same problems -- that it's not clear why it becomes conscious!

And I said that's going to be my last point on this score. But in fact I realized there's one more that I want to make -- Which is that the vast, vast, vast majority of our *interoceptive* processes *are* unconscious! - even *more* so than our exteroceptive processes. *Most* of our regulation of our visceral bodily economy happens autonomically; we know *nothing* about it.

1:25:43 So it would be, again, a very curious place to find the essential mechanism of consciousness, and, in this case, of affect, if the majority of the mechanisms of this kind (in fact) do not operate consciously, and if it suffers the very same problems as our exteroceptive... Using vision as a model example - using *any* form of perception, I

think, as a model example of consciousness, is to make a mistake, because consciousness is *not an afferent process*! Consciousness is an *endogenous* process. It's something generated not from within the guts, but from within the core of the brain. [chuckles] And that's where we're going next.

Claim 4. This casts new light on the 'Hard Problem.'

1:27:02 Solms: So now -- I've told you, in relation to the point I've just made: I think that the hard problem of consciousness has arisen in large part from the example that we've used of consciousness - starting with Crick -- who has to be credited with bringing consciousness into center stage - so you know I'm not bashing Crick! And I also have said already, I think it's perfectly reasonable that he started where he did. You know, it's clear why he started with visual perception as his example, where: "We should look for the neural correlate of visual consciousness; and then distinguish unconscious vision from conscious vision; and then, there we have it! Then we have the neural correlate; and then we can generalize from there."

1:28:02 I'm saying, that it turns out, that was an unfortunate place to start. And it links with what I was saying earlier. That's also why I'm saying I don't want us to make the same mistake, just turning it inwards and saying, "we're dealing with perceptual processes."

1:28:18 So let me say why all of this casts new light on the hard problem.

What makes the hard problem hard and almost unique is that it goes beyond problems about the performance of functions. To see this, note that even when we have explained the performance of all the cognitive and behavioural functions in the vicinity of experience ... there may still remain a further unanswered question: **Why is the performance of these functions accompanied by experience?** A simple explanation of the functions leaves this question open ... Why doesn't all this information-processing go on 'in the dark', free of any inner feel?"

David Chalmers (1995)

So here's Chalmers, the person who formulated the hard problem, in the very paper [1995] that he formulated it. And he says,

"What makes the hard problem hard and almost unique is that it goes beyond problems about the performance of *functions*. To see this, note that even when we have explained the performance of all the cognitive and behavioral *functions* in the vicinity of experience... there may still remain a further unanswered question: *Why is the performance of these functions accompanied by experience?* A simple explanation of the *functions* leaves this question open... Why doesn't all this information-processing go on 'in the dark,' free of any inner feel?"

This is a *core* aspect of Chalmers's "hard problem."

So they're claiming - Chalmers is and Nagel, and others of that ilk - that, "Somehow, consciousness is in a different realm from every other mental function. Because when you've explained all the functional mechanisms, you still are left with this **epistemic gap** - between that explanation of the function - and the qualitative, experiential, phenomenal *feel of seeing*."

"This is not to say that experience has no function. Perhaps it will turn out to play an important cognitive role. But for any role it might play, there will be more to the explanation of experience than a simple explanation of the function. Perhaps it will even turn out that in the course of explaining a function, we will be led to the key insight that allows an explanation of experience. If this happens, though, the discovery will be an extra explanatory reward. **There is no cognitive function such that we can say in advance that explanation of that function will automatically explain experience.**"

David Chalmers (1995)

1:32:05 And so I'm saying that the problem is, that they're using perception as an example. Perception is not an intrinsically conscious function! And, indeed, *cortical* perception is not intrinsically conscious! -- That it is demonstrably the case, that the cortex can perform uniquely cortical perceptual operations - like actually recognizing color - that you can recognize color unconsciously; and you can recognize faces unconsciously; you can read words unconsciously!

So if the cortex can do all of this perceptual and cognitive gymnastics without consciousness, I think it's perfectly reasonable to say, "Well, why doesn't all this information processing go on in the dark, without any inner feel?" -- because *that* kind of information processing *can* go on in the dark, without any inner feel!

Perceptual inference is not intrinsically conscious!

1:33:11 So this statement of Chalmers's - this really *strong* statement - where he says, "*There is no cognitive function such that we can say in advance that explanation of that function will automatically explain experience.*" This is what he says, on the basis of everything that I've just told you - this is where he's led. He's saying that "explaining the function can *never* explain experience!" If that's the case, we're in big trouble! -- because consciousness is *somehow* [chuckles] going to have to be dealt with in a different way from everything else in science!

1:33:56 But if you change that word "cognitive" to "affective" -- would Chalmers have said this? - "There is no *affective* function such that we can say in advance that explanation of that function will automatically explain experience?" I'm saying, of course there is! The function of *feeling*! I am saying in advance, that explanation of the function of feeling will automatically explain why it feels like something - because that *is* the function of feeling! The function of feeling is for the organism to *feel it* - to feel its own state!

1:34:34 So I'm saying: Had we begun there - had we begun our quest for the neural correlate of consciousness with *this* model example - had we begun by asking, "*Why* - and *how* - does it come about that *feelings are experienced?*" - I think that we would have had a much less hard time with it -- than we *have* had, with starting with visual information processing. The mechanism of vision, the mechanism even of cortical vision, is not intrinsically conscious. It does not automatically explain experience, because it does not automatically have to *be experienced*.

1:35:18 But the mechanism of feeling? "*You cannot explain how feeling works, why feeling exists, what it's there for!*" [Chuckles.] -- If you *have* a functional account of feeling, it would *have to* account for "why it feels like something" -- in other words, why it's experienced.

So, I think that - what I said in my earlier point (4) - that "these facts cast new light on the hard problem" -- *this* is the "new light" that it casts.

Perception and Learning Without Awareness

1:35:54 I have also on the screen this slide, this famous [1996] review paper by Kihlstrom, which summarizes all the experimental evidence for the conclusion that we were led to - actually, a good four decades ago already - that "perception without awareness of what is perceived" is perfectly commonplace, and "learning without awareness of what is learned" is perfectly commonplace. Cognitive and perceptual functions performed by the cortex are not intrinsically conscious processes.

Perception Without Awareness of What Is Perceived, Learning Without Awareness of What Is Learned

John F. Kihlstrom Yale University

Beginning in the 1980s, psychology (and cognitive science generally) has undergone a dramatic shift in its attitude toward the psychological unconscious -- that is, toward the idea that mental states and processes can influence experience, thought, and action outside of phenomenal awareness and voluntary control. Once rejected out of hand as too deeply embedded in psychoanalysis or other forms of pseudoscience, or at least as too vague to be scientifically useful, the notion of unconscious processing is now taken seriously by most researchers and theorists in the field. At this point, the debate has shifted from questions about the very existence of unconscious states and processes to debates about the nature and extent of unconscious processing. Credit for this state of affairs goes to four rather different lines of research (for a more extensive discussion of this recent history, see Kihlstrom, 1987, 1995).

First, cognitive psychology now embraces a distinction between automatic and controlled processing (e.g., Hasher & Zacks, 1979, 1984; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; for updates, see Bargh, 1989; Logan, 1989; Shiffrin, 1988). Whether they are innate or routinized by extensive practice, automatic processes are inevitably engaged by specific inputs, independent of any intentionality on the part of the subject, and they cannot be controlled or terminated before they have run their course. We have no conscious awareness of their operation, and we have little or no awareness of the information which they process. All that enters awareness is the final product of the automatic process. Thus, automaticity represents unconscious processing in the strict sense of the term: we have no introspective access to automatic procedures, or their operations; these can be known only indirectly, by inference.

Further contributions came from the emergence of cognitive neuropsychology (Rozin, 1976). (Some prefer the term *cognitive neuroscience*, but I prefer to stick with the traditional label, with its emphasis on the functioning of the whole human organism, rather than the molecular and cellular analyses which preoccupy so much of neuroscience; I also like to make clear that the mental states and processes of interest to psychologists include emotional and motivational as well as cognitive ones). Studies of the amnesic syndrome associated with bilateral lesions in the hippocampus and other medial-temporal structures, for example, revealed a distinction between two expressions of memory, explicit and implicit (Moscovitch, Goshen-Gottstein, & Vriezen, 1994; Schacter, 1995). Explicit memory is conscious recollection of the past; implicit memory is reflected in any influence of past events on subsequent experience, thought, and action. We now know that explicit and implicit memory can be dissociated in many different ways, indicating that implicit memory is in some sense independent of explicit memory (Roediger & McDermott, 1993). In the present context, the importance of the discovery of implicit memory is that it legitimized discussions of unconscious memories -- a topic which had been virtually taboo among nonclinical psychologists.

A third influence was from research on hypnosis, many of whose phenomena seem to involve a division of consciousness (Hilgard,

So this question of Chalmers's, "Why is the performance of these functions accompanied by experience? ... Why doesn't all this information processing go on in the dark?" -- that perplexity is not difficult to understand, in light of *those* facts.

Why is the performance of these functions accompanied by experience? ... Why doesn't all this information-processing go on 'in the dark', free of any inner feel?

1:36:47 But I don't think that this question has anything like the same gravity when you apply it to the kinds of mechanisms that give rise to feeling. I'm saying, in short, that if Mary was an *affective* neuroscientist, rather than a cognitive neuroscientist, or a visual neuroscientist, she *could not know* everything that there is to know about the functional mechanisms underpinning *feeling*, without expecting that it will generate something that is experienced, without the mechanism necessarily accounting for *why* it is experienced. And this is why I think that shifting our focus from cortex to brainstem, and from perception to affect, has big implications for our ability to finally come to grips with this hard problem of consciousness.

Question: Pre-arousal + awareness

1:37:49 *Friedman:*

Thank you, Mark! I'll read a comment in the chat; and then you can respond.

Dean writes, "Pre-arousal + awareness (minimum of 2) NOT as brain geography or territory establishment effort. Rather, A + A "Dance" in an evolving relationship as we might see in an agent "Dance" with the environment. Where the territorial view might examine control (brainstem of cortex or vice versa) - one in a controlling relationship - through a mechanical lens, this relationship - "Arousal + Awareness" - could be choreographed (mechanized) but is not necessarily choreographed. Are feelings a residue of Arousal + Awareness revealed (we accept the signals, like we dance-out a path, generated)?"

1:38:44 *Solms:*

Look - I think that this is similar to the point that we were discussing earlier, you know, about "*generated in the upper brainstem*," as opposed to "*the upper brainstem modulating other structures*," whether they be above or below the brainstem in the nervous system. I said then I'm very happy to think of it as a dance, as something choreographed -- where it's not a matter of who's master and who's servant, when it comes to the relationship between affective arousal and cortical representation. I think that there's some instances where affective arousal just absolutely *brushes aside* any amount of decision making and thought processing.

1:39:42 And conversely, the opposite happens, too - that there clearly are ways in which where one can, in a top-down way, modulate affect *dramatically*. And I don't only mean, you know, "regulate your emotions." I mean - if you look at some of these examples of these Buddhist monks who are able to *really massively* control autonomic functions, and the affects that normally would go with not breathing, or walking on painful surfaces, and so on -- that those affects seem to be overridden!

1:40:20 So I have nothing against the idea of there being "a choreography between brainstem and cortex." But I've already said why I think that, in this instance, the affects *do not* require the cortex, to actually qualify for the name "affect" -- that they actually are felt *without* cortex. I've already said why, what the evidence is, both in terms of here-and-now studies using the various methods that I mentioned -- the most dramatic being (as I said) cases where there *is no cortex*, and yet there's evidence that there's plenty of affect there.

1:41:08 But also the *evolutionary* implications of that, which is that we don't have to wait for the appearance of cortex and the mammalian series for the dawn of consciousness -- that if affect is generated *in* the subcortical structures, then affect does *not* require cortex; and affect can be attributed to other vertebrates. And not only vertebrates, by the way - there also is good evidence that some invertebrates are conscious! - the octopus is the most famous example.

1:41:53 But I think that there's a further problem in this comment, you know -- which has to do with the word "*awareness*." You know, you speak of "arousal" and "awareness." And you seem to attribute *arousal* to reticular activating system, which I'm happy with; and *awareness* to cortex, which I'm not so happy with - if you mean that you cannot be aware of arousal without cortex.

So I think this word "awareness" - it's like all the other words I mentioned at the beginning, like "emotion" and "affect" and so on. And I said I want to use the word "feeling" to make clear what I'm talking about, because these words mean different things to different people.

The word "awareness" means different things to different people. To many people it seems to mean something like "*reflective awareness*" - "being aware of what you're aware of," or "**access consciousness**," as Ned Block calls it -- that there's this sort of "phenomenal consciousness," which is some kind of "raw consciousness" -- and then there's a subject that *accesses* this information. It's like the Global Workspace Theory - works that way - it's pretty similar sort of an idea.

1:43:12 If we're going to reserve the word "awareness" for that higher order thought, or reflective consciousness, or access consciousness, then it means that "we can't attribute 'awareness' to raw feels." And that's an odd use of the word "awareness" to me. [Chuckles.] To me, how can you feel pain without being aware of it? - you know? And pain - you don't need to have any kind of reflective cognition about it! You don't need to have any explanatory labeling, or anything! It's just like, "Shit! That's sore! - and you pull away - even if you don't have language, even if you're not human, even if you're not a mammal, even if you're another kind of creature!

1:44:04 So "arousal," for me, *implies* awareness, [chuckles] in the way I use the word. But if we're going to use "awareness" for that higher order type of consciousness, then I would just want to remind you that there are *other* types of consciousness that don't possess that kind of awareness, which nevertheless have something-it-is-likeness, and the core of that quality is feeling. That something-it-is-likeness is feeling.

Question: Consciousness as metaphor

1:44:46 *Sillett:*

If we think of consciousness as a kind of an *integration* -- of our action-expectations of what's happening (Casper Hesp talks about this with his work with Active Inference) -- with how those expectations were or weren't met, can be effective -- Within that prediction process, when someone gets a question right...

So, you have a mathematician: When they get it right, they get a feeling that, "It's right!" - It's like it's a felt sense. And you could ask someone, "*What's that feel like?*" - and they can sort of bring out a metaphor landscape that relates to that - like it can actually have a structure. And there's a similar thing with pain as well. Pain has to be expressed as a metaphor. You have to say, "it's like a stabbing!" or "it's a throbbing!" or "it's a..." - It's always some dynamical kind of metaphorical embodiment.

1:45:55 I'm just curious. This embodiment, this meeting in this brainstem -- how you see that in terms of -- If that's congruent with you? And whether this is then a sort of a *re-imagining*? Consciousness is like a reimagining of these expectations? And it could go up *or* down, in terms of how that is being realized?

1:46:34 *Solms*:

So... I'd like to separate out two different aspects of what you're saying. The first has to do with **metaphor**. I think that that may be true of being able to report or declare your conscious states in language - that somehow or other you have to symbolize. Because language is just that - it's a symbolic re-presentation of the thing itself. As long as you... you used the word, "expressing it..." Yes. You can't *express* in words and emotion without using some sort of metaphor, some sort of symbolic system in order to be able to communicate it to others.

1:47:35 They can, of course, be the direct behavioral affective display - "Expression of the Emotions," in Darwin's sense of the use of that term - which we then, as a symbolic species - we use *that* symbolically, too. But I don't think that the raw feeling itself has to be re-represented in any way! I think it can just be the raw feeling, without it being represented in any language, in the broader sense of the word.

And so I don't think that metaphor is a necessary component of the core feature of feeling.

Perception as fantasy

1:48:28 Then the other thing you were saying, which had to do with surprise, if I understood you correctly. In my own thinking about this, there were times when I was keen on thinking about affect - or *linking* affect more with error signals than with inferences themselves. I was thinking along those lines because so many people within the Active Inference framework were speaking about perception.

1:49:10 You know this famous phrase "controlled hallucination" - that "perception is fantasy" - that it is the *upshot of the predictive process*. And what worried me about that is the very thing I've been banging on about all along, which is that perception is not intrinsically conscious. And so this business of *inferring the source of your sensory signals* doesn't explain why there's something it is *like to infer the source of*, or the cause of, your sensory signals! And so, from that point of view, I've found myself thinking - well, it's not so much that -- I mean, imagine, if what you perceive is just what you predict! That would be bloody dangerous!

1:50:07 So the *controlled aspect* seemed to me to be the more important aspect. It's the *constraint* on what you predict that seems to be more the upshot of what we call perception, as opposed to hallucination. And so I wanted to place the emphasis more on the *error* side of things than on the *prediction* side of things.

But subsequently I have changed my mind on that score, which will become clearer when I get [to Claim 8, "**The mechanism of perceptual and cognitive consciousness is precision modulation of prediction errors.**"].

1:51:00 For me, it's got everything to do with *the modulation of confidence in the prediction*, versus the error signal - or the error signal that is connected with that particular prediction.

And I will just say it now - that the pithy formulation of the way that I see it, is that "It is *good* when things turn out as expected; and it is *bad* when uncertainty prevails." And I mean here "goodness" and "badness" in the *valuative* sense of the word, that is linked to the *primary value system of all living things* - which is that "It is *bad* to die and it is *good* to survive." And if things are turning out as expected - in other words, you have a prediction which actually cuts it - that that is *good* from the viewpoint of the organism. And that if things are *not* turning out as expected and uncertainty is prevailing, that is *bad!* - It means this is likely to end in tears.

1:52:20 And - What was it - the "choreography" that somebody with the dance - that somebody else was talking about earlier? And I think that the crucial thing here, is the *palpating of the precision*, the *modulating* of the precision, in the predictions over the error signals -- and that that is at the heart of what affect's all about. And I'm only saying that now in advance.

Session wrapup

1:53:04 Are there any further questions and comments?

Question: Other thinkers

1:54:15 *Douglass*:

In addition to Professor [Anil] Seth, whose book we already have - who else can we check with, as people doing [Active Inference and affective processes]?

1:54:41 *Solms*:

Two people who come to mind immediately: Manos Tsakiris, and Katerina Fotopoulou. They are both great! And both of them are very good speakers. So if you're wanting to invite them on to this Livestream series, I think that they would do well.

1:55:25 *Sillett*:

Just to say, we've been reading some of the work of Ryan Smith and Casper Hesp...

1:55:32 *Solms*:

Oh, well! Ryan Smith also! And Maxwell Ramstead! They're also working on affect! They're doing some fabulous work! I really loved their recent paper on folk psychology! It was just brilliant!

1:56:06 *Friedman*:

Any closing thoughts?

The importance of Active Inference

1:56:32 *Solms*:

I'm particularly pleased that you're going to be recording - and I don't just mean video recording - but also turning into some sort of written form what we're talking about.

Because - as I hope you can see from what I've said - these really are pretty basic issues for consciousness science, and for the science of the mind, and by implication for what we can do within this Active Inference framework, in tackling these...

Really, I don't mean to be hyperbolic when I say, *these profound and ancient problems*! I think that we have real prospects of being able to make significant progress on these fronts! And so it *deserves*, is the point I'm making - it *deserves* the time that we're giving it. And I'm grateful to you for recognizing that.

1:57:43 (*Many speakers*):

Thanks very much! Cheers for now!

Session #016.2, June 11, 2022

Second participatory group discussion.

Session contributors: Prof. Mark Solms, Daniel Ari Friedman, Stephen Sillett, David S. Douglass

Videos (English with public comments): https://www.youtube.com/watch?v=c0_Vf5_qiWk

Scrolling Closed Captioning:

- English (word for word)
- Chinese (simplified)
- Chinese (traditional)
- Dutch
- French
- German
- Italian
- Japanese
- Korean
- Portuguese
- Russian
- Spanish

Introduction; Recap of Session 1

00:06 *Daniel Ari Friedman:*

Hello and welcome, Everyone. This is ActInfLab GuestStream number 16.2. It is June 11th, 2022. We're here with Mark Solms. This is the second part in our discussion. We had an awesome Part One! Today we are going to have a recap on some of the points that were addressed in the first part, in 16.1. And then we're going to head into some unexplored territory.

So, Mark, thanks so much for joining for these sessions. And please take it away.

00:42 *Mark Solms:*

Thanks, Daniel. Glad to be here again. (I was just testing out sharing my screen a minute ago, and it seems as if I'm incapable of showing you my face and sharing my screen at the same time. So take a good look at me. You're not going to see me for a while.)

And here's my title - "Consciousness as Precision Optimization: Some Physiological and Philosophical Considerations"

The emphasis is on physiological and philosophical considerations rather than computational ones.

Consciousness as Precision Optimization

Some Physiological and Philosophical Considerations

Mark Solms

Neuroscience Institute,
University of Cape Town

Mark Solms

So I'll start with a brief recap. I made a few claims, and I didn't get to the end of this list. So I will just remind you what those claims were.

And then I will pause when I get to the point that we stopped at last time, which was here - that [(Point 5)]
“**Affect is an extended form of homeostasis.**”

Affect is the foundational form of consciousness and it is intrinsically conscious.

This claim is based on empirical evidence, not on semantic arguments.

Affect is not synonymous with interoceptive inference.

This casts new light on the 'hard problem'.

Affect is an extended form of homeostasis.

Complex organisms require multiple (categorical) homeostats.

These must be prioritized.

The mechanism of perceptual (and cognitive) consciousness is precision modulation of allostatic prediction errors.

The predictive hierarchy is arranged concentrically, for progressive tolerance of uncertainty.

Consciousness can be artificially engineered.

Mark Solms

Claim 1: Affect is the foundational form of consciousness and it is intrinsically conscious.

01:53 I started with the claim, which really is the main claim of my presentation.

It is that *affect - feeling* - is the foundational form of consciousness, and that it is *intrinsically* conscious.

“It is surely of the essence of an emotion that we should be aware of it, i.e. that it should become known to consciousness. Thus the possibility of the attribute of unconsciousness would be completely excluded as far as emotions, **feelings** and affects are concerned.”

Sigmund Freud (1915)

02:10 I quoted Freud, who more than anyone else introduced the notion into mental science that mental processes are *not* intrinsically conscious, that much of our cognitive processing goes on *unconsciously*. And even he made this point, the point that I'm making, which in his words was that

“It is surely of the essence of an emotion that we should be aware of it, that is, that it should become known to consciousness. Thus the possibility of the attribute of *unconsciousness* would be completely excluded as far as emotions, *feelings* and affects are concerned.”

02:52 And I emphasize the word “feelings,” because there's a whole lot of reasons why some people claim that there are such things as unconscious emotions and unconscious affects. So just to be absolutely clear what I'm talking about here:

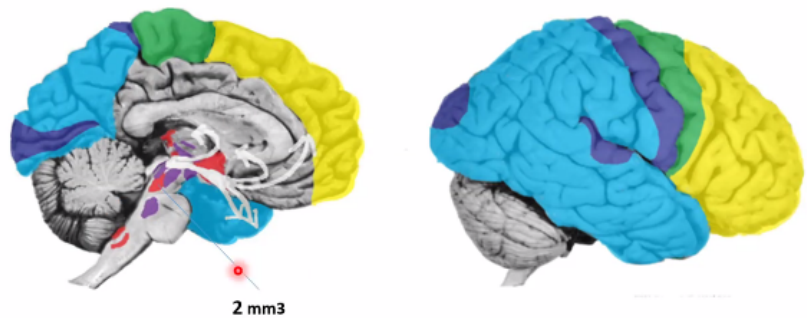
I'm talking about *feeling*. And you can't have an unconscious feeling! In other words, you can't have a "feeling" that you don't feel. So whatever might be meant by "unconscious emotions" and "unconscious affects," they can't apply to feelings, they can't be unconscious feelings.

Claim 2: This claim is based on empirical evidence, not on semantic arguments.

03:27 But that might sound like a semantic point, just a matter of words, that “you can't have a feeling that you don't feel” - that's an oxymoron.

So I then went on to set out various reasons, empirical ones, why the claim that "feelings are intrinsically conscious" also has a solid empirical basis.

I drew attention to this discovery that was made in 1949 already, that the brain mechanisms for *consciousness itself* - in other words, what wakes us up in the morning and puts us to sleep at night, the sort of switching on and off of the lights of consciousness - that this is the business of the *reticular activating system*.



RETICULAR ACTIVATING SYSTEM and PERIAQUEDUCTAL GREY

The reticular activating system is *the* brain system for arousing consciousness, for activating cortical processes and rendering them conscious.

04:33 I also mentioned the *periaqueductal gray*, which also plays a crucial part in all of this, together with the reticular activating system. But I'll come back to that later. For now, let's just focus on the reticular activating system. To give you a sense of what I mean by how these structures -

Their intrinsic function is the switching on of the lights of consciousness. I pointed out that you can make a lesion - or you can suffer a lesion - as small as two cubic millimeters in size in the *parabrachial complex* of the reticular activating system - in other words, a lesion the size of a match head - and that is enough in human beings reliably to obliterate consciousness entirely.

So I was saying, let's look at *these* structures, if we're wanting to understand something of the intrinsic consciousness-generating mechanisms of the brain.

Case 1: Patients Without Cortex.

See above images "Labeled diagram of side view of human brain," "Brain scans of hydranencephalic child."

05:38 And I made the point that if the *cortex* - which is where we normally shine our light in terms of looking for the neural correlates of consciousness - if the cortex were the seat of consciousness, then there's an easily testable

prediction, which is that “If you have a case in which there is no cortex, then the patient should not be conscious.” Not so.

[See 24:02 in <https://youtu.be/NHfGuwoCXI8?t=1441s>]

Therefore, I showed you one such case. And this is one representative example of its type, that here is a patient born with no cortex. The condition is called “*hydranencephaly*.” Here you see in an MRI scan of the brain that where cortex should be there is just cerebrospinal fluid. And you see that her brain stem is perfectly intact.

[See above images “*Hydranencephalic girl with infant brother, “Color photo of little girl laughing.”*]

06:32 So on the view that this brain stem area is where the consciousness is generated, that it doesn't require cortex, this is a critical case. On the brain stem view, she should be conscious. On the cortical view, she should not. And here she is, and she is conscious, as you can see. She wakes up in the morning, she goes to sleep at night; in this sense, “the lights are on.”

But much more interestingly, she's not merely “blankly awake.” Her wakefulness has a content and a quality. And the quality that I'm talking about is *affective quality*. You see how she responds to her baby brother being placed on her lap? She responds with some form of pleasure. And so there's a content to this mental state and a quality to this mental state that she's displaying.

07:31 Why that's important is because, when we first learned in 1949 that the reticular activating system is prerequisite for all consciousness - that there's no such thing as cortical consciousness without brainstem arousal of the cortex - we had the view that

The cortex provides the qualities and the contents of consciousness; and the reticular activating system merely provides the quantitative dimension, the sort-of ‘level of consciousness.’

So it's as if this was a power supply, like a television set needs to be plugged in at the wall. The reticular activating system is the prerequisite - sort of booting up of the system - but the television set itself, the cortex where the contents and qualities of consciousness are processed, could still be claimed to be the seat of consciousness; and the reticular activating system merely as a power source.

08:34 This is why it was important to point out that this child is not only conscious in some blank sense of wakefulness without content and quality, but rather that she displays affects. She's emotionally responsive to her baby brother being placed on her lap – just as this child, who also has no cortex, is emotionally responsive.

Here you can see - she's responding with pleasure to a stimulus. And this is quite generally the case for these children.

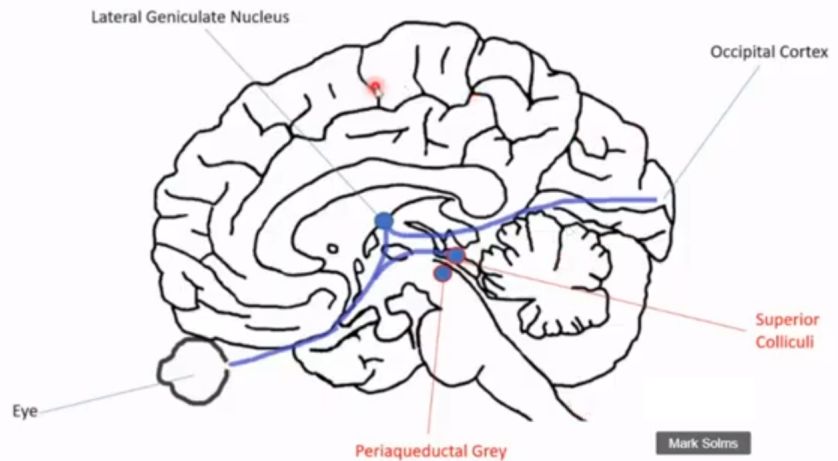
[See image “*Affective responses by hydranencephalic children.*”]

Here's Björn Merker's summary of his observations in many, many, many such children.

They express *pleasure*, they *smile*, they *laugh*. They show *aversion* and *fussing* by arching their backs and *crying*. Their faces are *animated* by these *emotional* states. They build up *play* sequences, they *smile*, they *giggle*, they *laugh*, they show *great excitement*. Et cetera, et cetera, et cetera.

09:30 So I've highlighted all of those words to show that these two cases that I just showed you with *no* cortex, they are conscious and they are responsive. And in particular, they are *emotionally* responsive. And on this I base the claim that these children do have a quality, that they do have a consciousness, and this consciousness does have a quality and it does have content.

Many people are perplexed as to how this could happen, since they have no cortex. “How can they respond to things like their baby brother's being placed in their laps?” And so I just inserted this slide, just to point out that our *sensory end organs* - here the example is the eye - same applies to the skin and to hearing and taste as well.



10:26 The optic nerve that projects to the lateral geniculate and from there to the visual cortex – But not only to the visual cortex. It also projects subcortically to the superior colliculi of the midbrain - of the brainstem - which is immediately adjacent to the periaqueductal gray, which I said earlier I was going to mention again.

So these children receive information in the brainstem which is not conscious, it's not *cognitively conscious* perception. Conscious perception is generated in the cortex; but *unconscious* sensory information goes into the brainstem, where it is responded to consciously by the emotional structures that are the main focus of what I'm talking to you about.

Case 2: Loss of Pre-Frontal Cortex or of Basal Ganglia

11:19 Now of course, many of you, like many of my colleagues, [chuckles] will reasonably say,

"Well, how do you *know* that there's something it's like to be those children? These might just be reflexes, or instincts. And these might be the equivalent, these kids, of philosophical zombies. In other words, they *look* as if they're conscious; they *behave* as if they're conscious; but we can't *know* that they're actually conscious."

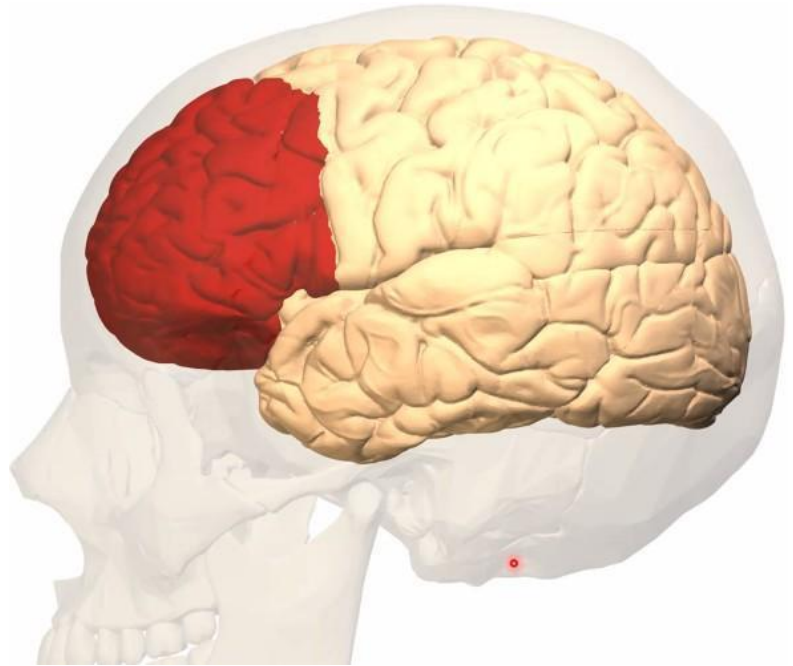
And so last time I tried to address this objection, by drawing attention to what happens in cases who lose great swathes of their cortex. Because they haven't lost *all* of their cortex, they are able to describe to us what it's like to be them.

This is one way of getting around the objection that “because these kids have no cortex and therefore can't *declare* their conscious states, we can't be sure that they *have* conscious states.”

12:21 So I focused first of all on the case [Case 2] who has a massive lesion of the prefrontal cortex.

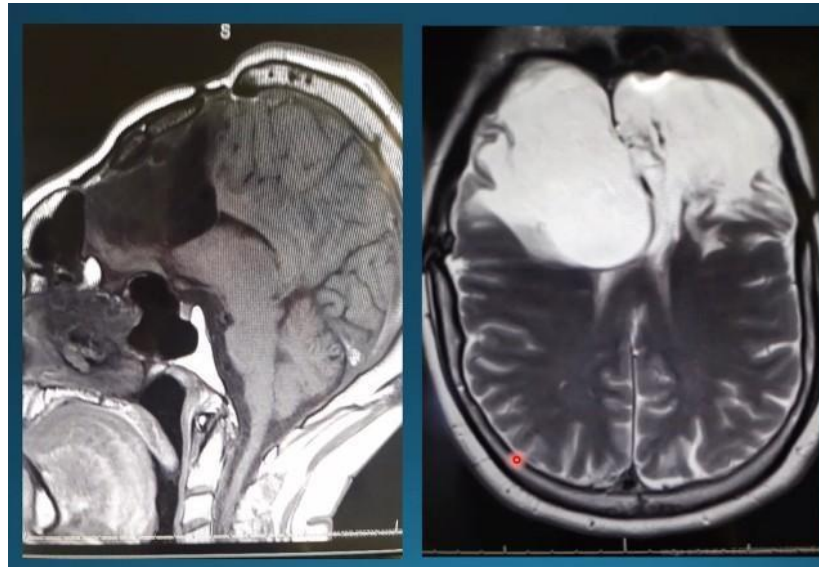
The reason I did that is because this is a favorite part of the cortex for those who claim the cortex is the seat of consciousness, like the Global Workspace theorists.

They say “It all comes together in the prefrontal lobes. All of this information that's processed in the posterior cortices is *re*-represented or accessed in the Global Workspace. And this is where the sentient being, the subject of the mind, comes about.”




And I pointed out last time that if that were the case, then again we have a falsifiable prediction, a patient who has no prefrontal cortex like this patient of mine, Patient W. He has no prefrontal cortex, but he has a sliver of language cortex.

If his sentient being was contingent upon the integrity of prefrontal cortex, he should not have sentient being.



13:30 And so... I talked to him about what it's like to be him. He claims to be consciously aware of his thoughts. I asked him to imagine something for me, to imagine two dogs and a chicken, to see them in his mind's eye. Then I asked him to count the legs. I thought this would be a reasonable test of whether he's actually got conscious mental imagery. (And please note, the person I'm talking to refers to himself as "I." So he seems to think that he exists as a sentient being.) And I asked him how many legs there are in total if you have two dogs and one chicken. And to my disappointment, he said "eight." And when I questioned his answer, he pointed out that in his mind's eye the dogs had eaten the chicken!

Me: Are you consciously aware of your thoughts?
Patient W: Yes, of course I am.
Me: In order to confirm that, I am going to ask you to solve a problem that will require you to consciously picture a situation in your mind.
Patient W: OK.
Me: Imagine that you have two dogs and one chicken.
Patient W: OK.
Me: Do you see them in your mind's eye?
Patient W: Yes.
Me: Now tell me how many legs do you see, in total?
Patient W: Eight.
Me: Eight? 
Patient W: Yes; the dogs ate the chicken.

And so I thought [laughing] that was maybe not a great joke, but it certainly suggested that there was somebody at home.

And I made the point last time that these patients generally, patients with massive frontal lesions, that this tendency to make puerile jokes is considered to be a rather common part of the frontal lobe syndrome or the frontal lobe personality.

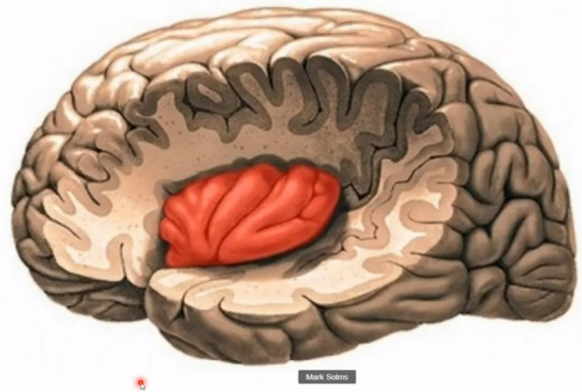
14:42 And this is part of the bigger story that I mentioned last time, which is that these patients are generally quite emotionally disinhibited. And why that's important is because, remember, what we are considering here is the question as to whether the *feelings* that you saw in those patients - those kids with no cortex - whether those feelings could be generated in the brain stem.

15:08 And the claim of **cortico-centric** theorists like Joe LeDoux, for example, is that the feelings literally come about when they are registered or re-represented or labeled - or even *named*, some people claim - that it's only once you are able to re-represent these subcortical survival circuits in declarative consciousness in prefrontal lobes, that this is literally what brings the feeling about.

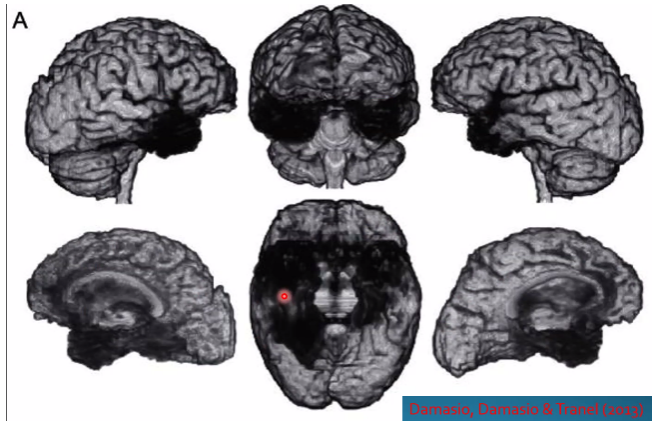
And I think it's quite interesting that these patients who have *no* prefrontal cortex *don't* have a *dearth* of feeling, which is what you would expect if the machinery that brings feelings about is absent, then they shouldn't be able to have feelings. But in fact what is generally accepted is these patients have an *excess* of feelings; and there's the disinhibited emotionality in these patients.

16:10 I made much the same argument about the other major cortical area that is associated with the **sentient self**, and that is the insular cortex.

Of course, this is associated above all with the work of Bud Craig, but has been very widely accepted, that the feeling self comes about when the state of one's interoceptive body is re-represented in insular cortex.



And so again, we have the falsifiable prediction: "If you take a case who has no insular cortex, like this patient of Damasio's, then there should be no sentient self present."



17:00 And I showed you this interview, where Damasio speaks to him about his sense of self.

And the patient is perfectly adamant that he exists as a self. He speaks about himself as "I," "I," you know?

And this interview ends with Damasio saying, "You're aware that I'm aware?" And the patient, B, says, "I'm aware that you're aware that I'm aware!"

Damasio: Do you have a sense of self?
Patient B: Yes, I do.
Damasio: What if I told you that you weren't here right now?
Patient B: I'd say you've gone blind and deaf.
Damasio: Do you think that other people can control your thoughts?
Patient B: No.
Damasio: And why do you think that's not possible?
Patient B: You control your own mind, hopefully.
Damasio: What if I were to tell you that your mind was the mind of somebody else?
Patient B: When was the transplant, I mean, the brain transplant?
Damasio: What if I were to tell you that I know you better than you know yourself?
Patient B: I would think you're wrong.
Damasio: What if I were to tell you that you are aware that I'm aware?
Patient B: I would say you're right.
Damasio: You are aware that I am aware?
Patient B: I am aware that you are aware that I am aware.

17:25 And this patient, just like my Patient W, is not deficient in emotionality. In fact, all the basic emotions are present, including both bodily affects and emotional affects and sensory affects. And not only that, he's also a little too emotional. And this is a little disinhibited in his emotionality. And this is what we quite generally see with patients with insular lesions. So again, it's very hard to sustain the argument that the self actually comes into being in the cortex.

Case 4: Deep Brain Stimulation of the Substantia Nigra

18:05 Now of course, those two cases, those examples of their kinds - patients with massive frontal cortical lesions, patients with massive insular lesions, of course they are just examples. But the point I'm making now, is that there's lots of cortex left in those cases. So the argument is, "Well, you know, maybe the sentient subject is generated in the remainder of their cortex."

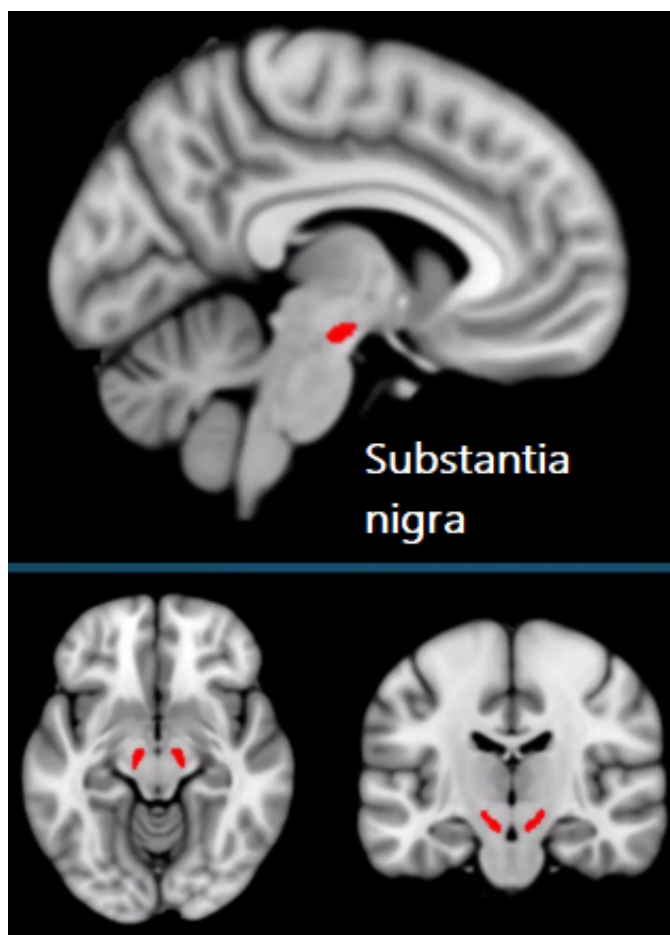
18:37 And it's a bit of a circular argument! Because, remember, the children who have *no* cortex, we're told, "Well, how do we know that that they're conscious? - they can't report their conscious states!"

18:51 I then gave these two outstanding examples, looking at the areas of cortex that are most bound up with sentient subjectivity according to cortical theories of consciousness. And these patients seem to have intact a sentient subjectivity. But now I'm told, "Well, you know, the *remainder* of the cortex might be what's generating the consciousness!" So we can't use only lesion methods. There's no way out of that impasse.

19:21 And so I then showed you (last time) evidence of a different kind. I showed you what happens if you stimulate reticular activating system nuclei.

And there again, there's an easily tested prediction. The prediction is that if these brain stem nuclei - the reticular activating system and periaqueductal gray, as I'm claiming - if these are the structures that generate consciousness and feeling, then stimulation of these structures should stimulate conscious feelings.

And I showed you that, here's a case who had a **deep brain stimulator** placed into the substantia nigra - this part of her reticular activating system - and within 5 seconds it generated, it produced, a profound depression.



20:16 The patient was actually suicidal. She didn't want to live anymore. This is a patient with no psychiatric history. The electrode was placed in her brain stem for the treatment of Parkinson's disease and stimulated the wrong nucleus. And that's how this came about. 90 seconds after the stimulator was switched off, the depressed feeling disappeared. And the patient generously agreed to allow for further stimulation in the reticular activating system and out of it. And it was only when that particular nucleus was stimulated that that she fell into the depression.

Interview with deep-brain-stimulation patient (Blomstedt et al, 2008)

"The patient's face expressed profound sadness within five seconds [...] Although still alert, the patient leaned to the right, started to cry, and verbally communicated feelings of sadness, guilt, uselessness, and hopelessness, such as 'I'm falling down in my head, I no longer wish to live, to see anything, hear anything, feel anything ...' When asked why she was crying and if she felt pain, she responded: 'No, I'm fed up with life, I've had enough ... I don't want to live any more, I'm disgusted with life ... Everything is useless, always feeling worthless, I'm scared in this world.'
[ctd. ...]

[... ctd.]

When asked why she was sad, she replied: 'I'm tired. I want to hide in a corner ... I'm crying over myself, of course ... I'm hopeless, why am I bothering you' [...] The depression disappeared less than 90 seconds after stimulation was stopped. For the next five minutes the patient was in a slightly hypomanic state, and she laughed and joked with the examiner, playfully pulling his tie. She recalled the entire episode. Stimulation at other sites did not elicit this psychiatric response."

Blomstedt et al (2008)

So this is the kind of evidence. And remember, again, I'm just giving you examples.

You can stimulate intense affective states by stimulating reticular activating nuclei and periaqueductal gray. You get the greatest intensity and the greatest variety of affects from stimulating there. And you get nothing of the kind from stimulating cortex! So this is a different line of evidence suggesting that feelings are actually generated in the upper brainstem.

21:31 And then I showed you another line of evidence. That was positron emission tomography of people in intense affective states.

And here you see research participants in states of sadness; here of anger; here of happiness; here of fear.

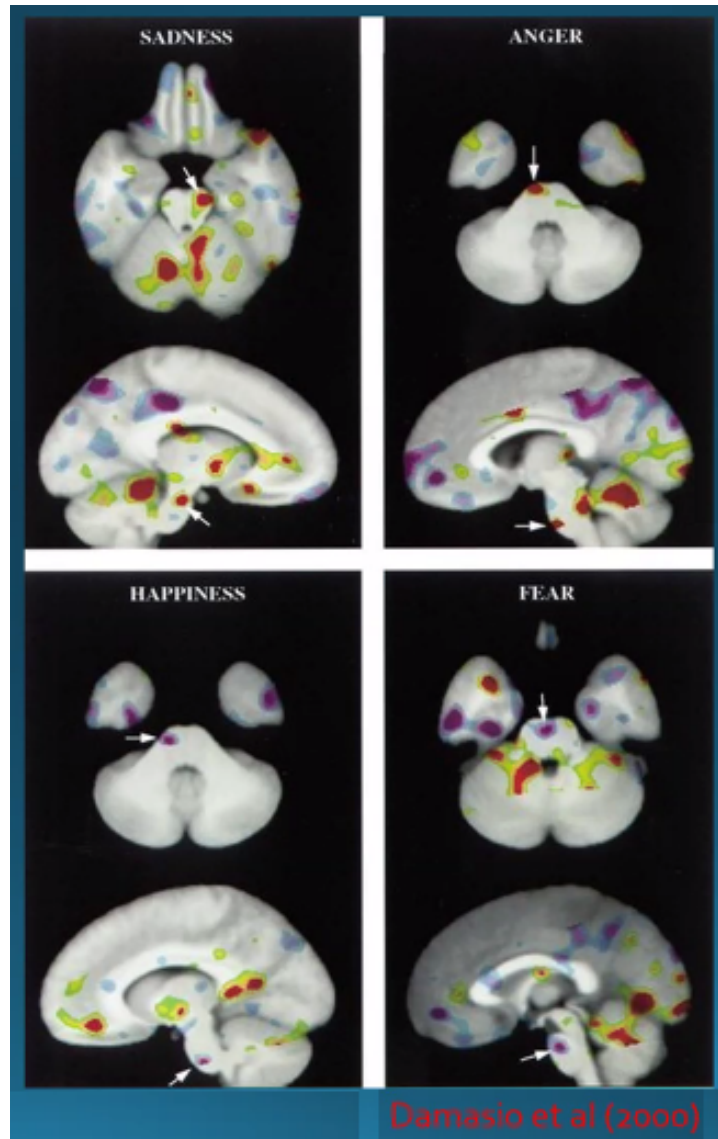
And in all instances, the activation is in the brainstem and the circuits arising from it, the subcortical circuits arising from it. That's what we see.

The cortex is, by contrast, largely *deactivated*.

So this is a further line of evidence - an entirely different line of evidence.

Remember - we've got lesion evidence; then we've got deep brain stimulation evidence; here, we've got positron emission imaging evidence - that the part of the brain that's *generating* the feelings *is* the part of the brain that switches on the lights!

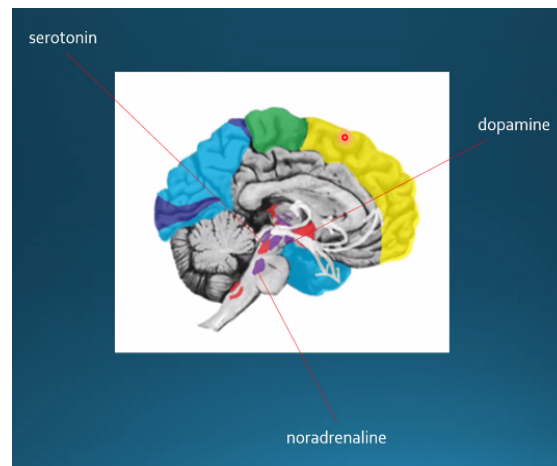
And this is why I'm claiming that the basic - the foundational - form of consciousness - this prerequisite form - is affect. Affect is the elemental form of consciousness.



22:43 A further line of evidence is pharmacological manipulations. If you tinker with the neuromodulators that are sourced in these reticular activating nuclei - like for example, noradrenaline or serotonin or dopamine, all of which are sourced in the reticular activating system - right?

Serotonin is, of course, regularly pharmacologically manipulated for the treatment of depression, dopamine for the treatment of psychosis, noradrenaline for the treatment of anxiety.

Noradrenaline is sourced in **locus coeruleus** complex; serotonin in the **raphe nuclei**, dopamine (at least the one that's important for psychosis) in the ventral tegmental area. All of these are parts of the reticular activating system.



If all that this system did was "switch on blank wakefulness," it might be of interest to anesthetists. But in fact it is the main target of the drugs of **psychiatrists**, who are treating emotional disorders by manipulating the chemistries, the source nuclei for which are in the reticular activating system.

Claim 3: Affect is not synonymous with interoceptive inference.

23:57 So on the basis of all of that, I argued that "affect is the foundational form of consciousness, and it's intrinsically conscious."

Not only on semantic grounds, I'm saying that *feeling* is the basic form of consciousness. The reason I'm saying that, is because the basic **consciousness-generating tissues** of the brain, the reticular activating system, which is prerequisite for the activation of consciousness in cortex, that these structures *generate* feeling, that feeling therefore is prerequisite - is foundational - for all forms of consciousness.

24:40 (This is a sort of a sidebar.) The reason I went into this third point, namely the claim that "affect is not synonymous with *interoceptive* inference" - in other words, that it's not just an interoceptive form of perception (as opposed to the *exteroceptive* forms, which have been the major focus of consciousness studies over the last few decades, vision being the *main* focus)...

And why I say this, is because this is increasingly being argued - in fact, it was argued by Bud Craig himself - that affect is just interoceptive perception, the perception of the state of **the Own Body** - is the equivalent of the perception of the outside world, in the more typical form, or exteroceptive form, of perception that's been equated with consciousness.

25:39 I made the point that there's good reason to believe that affect isn't just "a sixth modality of perception." And the evidence that I presented for that, was of various kinds. For example, that there are affects which are clearly not interoceptive. For example, getting a fright with[?] being startled, or feeling pain when a pin is stuck into your finger.

These are all *exteroceptive* forms of stimulation. And yet they arouse affective responses. So affect clearly is not uniquely interoceptive.

026:21 And then I also made the point that there are many interoceptive perceptual states which are not affective! You can feel your tummy grumbling or you can feel your heart beating or you can feel your lungs expanding. These are not intrinsically affective phenomena.

So interoceptive perception can happen without affect, and affect can happen without interoceptive perception. So I was trying to draw a line under the idea that "affect is just another modality of perception."

Claim 4. This casts new light on the 'hard problem.'

26:56 I think that by detaching affect from perception casts some new light on the hard problem. And so that's where I went next.

And I reminded you that David Chalmers, who coined "the Hard Problem," said that when we look at perceptual functions like vision (which was, as I said, the model example derived from or grounded upon the assumption that consciousness is a *cortical* phenomenon), cortical vision became the model example of consciousness. This was following [Francis] Crick's initiative in the mid 1990s.

What makes the hard problem hard and almost unique is that it goes beyond problems about the performance of functions. To see this, note that even when we have explained the performance of all the cognitive and behavioural functions in the vicinity of experience ... there may still remain a further unanswered question: **Why is the performance of these functions accompanied by experience?** A simple explanation of the functions leaves this question open ... Why doesn't all this information-processing go on 'in the dark', free of any inner feel?"

David Chalmers (1995)

27:40 Crick's idea was that if we can identify the neural correlate of consciousness in the case of cortical vision, then we can generalize from that, by discerning the mechanism whereby visual information processing gets turned into *conscious* vision in the cortex. By isolating that mechanism, we will be able to understand the nature and function of consciousness.

And Chalmers said that that's not true! If you isolate, identify the *mechanism* of visual information processing (like this map here does) [Image ##, "Functional Mechanisms..."], it doesn't tell you anything about why there's something it is like to see!

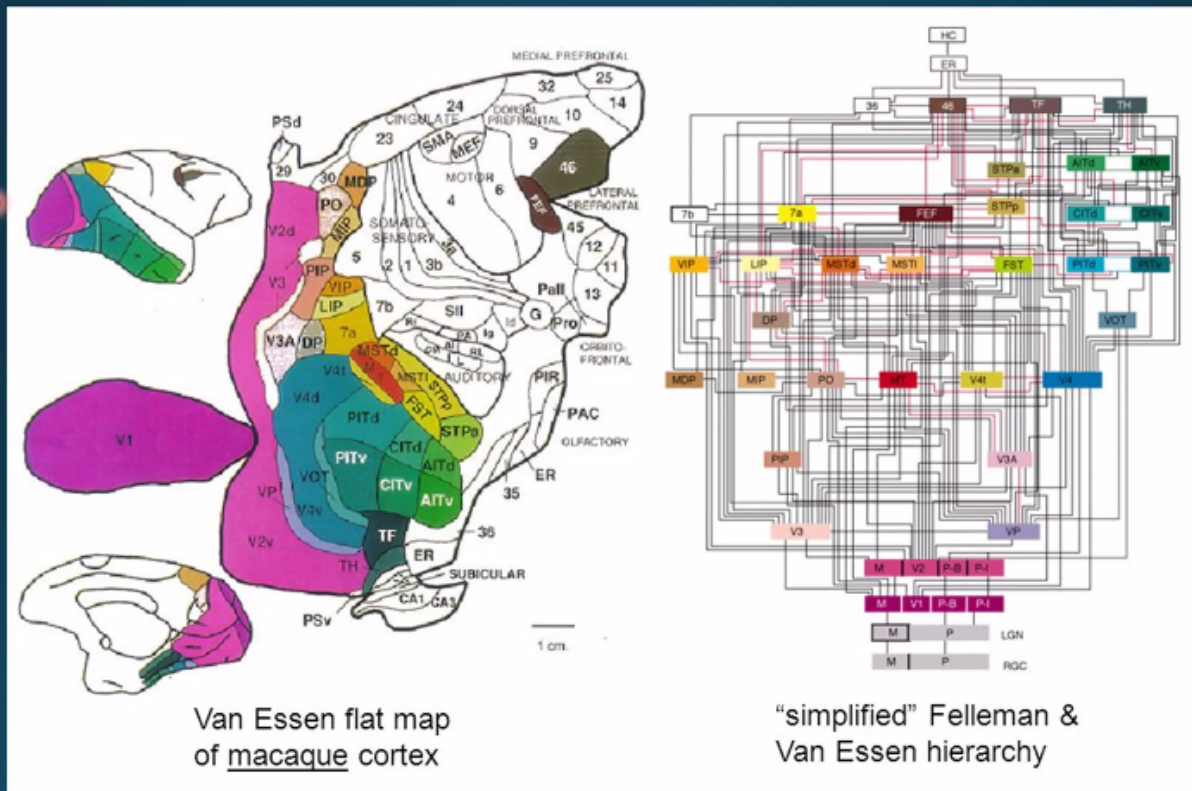
28:30 And he used the well-known "knowledge argument" of Frank Jackson, the story about Mary, the visual neuroscientist who knows everything about all of this. And I slightly simplified the story by saying, Well, let's imagine that Mary is blind. Even though she knows everything about the functions of cortical vision, she knows everything about the mechanism whereby visual information is processed in the cortex - because she's blind, she knows nothing about what it is like to see.

And if she were to be gifted suddenly with normal sight, then she would learn something completely new about vision. She would learn what it is like to see red and blue; what blueness and redness, et cetera, are like - none of which is accounted for by this information processing flow diagram.

29:31 And none of what she knew about the functional mechanisms of cortical vision that would have prepared her for what it is like to see. In other words, there's other there's something *else* about visual information processing other than the sort of things that mechanistic functionalist dissections like this provide us with.

Image. Functional Mechanisms of Visual Information Processing

Functional Mechanisms of Visual Information-Processing



Felleman & van Essen (1991)

This mechanistic account doesn't predict that there *should* be anything that it is like to see! And this is the grounds upon which people like Chalmers say that an account of the functional mechanism of cognitive and perceptual processes, it doesn't tell us anything about why there is something that is like to perform these processes. And this is the essence of the Hard Problem.

30:27 I said that I thought that this might be because they were looking in the wrong place! - that visual perception and *all* forms of cortical perception - and indeed, not only perception but learning and cognition more generally - that this can readily go on unconsciously - that these are not intrinsically conscious processes - that it is perfectly possible to see and to recognize faces and to read with comprehension and even to discriminate colors!

Perception Without Awareness of What Is Perceived, Learning Without Awareness of What Is Learned

John F. Kihlstrom
Yale University

These are all uniquely cortical processes. And they can all go on "in the dark," as it were. In other words, you do not have to be conscious of what you're perceiving in order to perceive it. And so this casts a rather different light on Chalmers's point: that all of this information processing can go on in the dark!

31:26 So, you know, consciousness isn't accounted for by our normal functionalist and mechanistic way of doing cognitive neuroscience. So I was saying, well, that's because they're looking at functions like perception and functions like learning. These are *not* intrinsically conscious processes.

And I drew your attention to this other statement of Chalmers's from his famous paper, in which he says in summary, "There is no cognitive function such that we can say in advance that explanation of the *function* will automatically explain experience."

And my point was, well, that's because they're talking about *cognitive functions*.

"This is not to say that experience has no function. Perhaps it will turn out to play an important cognitive role. But for any role it might play, there will be more to the explanation of experience than a simple explanation of the function. Perhaps it will even turn out that in the course of explaining a function, we will be led to the key insight that allows an explanation of experience. If this happens, though, the discovery will be an extra explanatory reward. **There is no cognitive function such that we can say in advance that explanation of that function will automatically explain experience.**"

David Chalmers (1995)

Could Chalmers make the same statement if we were talking about *affective* functions? I'm saying, No, he could not!

I'm saying that "We *can* say in advance that explanation of the *function of feeling* will automatically explain experience," because the function of feeling is to feel! It's intrinsically conscious, is this function!

"This is not to say that experience has no function. Perhaps it will turn out to play an important cognitive role. But for any role it might play, there will be more to the explanation of experience than a simple explanation of the function. Perhaps it will even turn out that in the course of explaining a function, we will be led to the key insight that allows an explanation of experience. If this happens, though, the discovery will be an extra explanatory reward. **There is no affective function such that we can say in advance that explanation of that function will automatically explain experience.**"

32:40 Unlike vision and perception in general, and cognition as a whole - none of that is intrinsically conscious. But feeling - affective feeling - *is* intrinsically conscious. You could not understand the mechanism of affective feeling if it didn't account for why "it feels like something!" Because that's the whole point of feeling.

So this question of Chalmers's, "Why is the performance of these functions accompanied by experience? Why doesn't all this information processing go on in the dark, free of any inner feel?" I'm saying that that question is perfectly reasonable when asked of these cognitive functions which are not intrinsically conscious. That kind of information processing *can* "go on 'in the dark,' free of any inner feel."

But that is not true of feeling, of affective feeling.

And this is where we got to last time.

I'm now starting with new arguments.

Why is the performance of these functions accompanied by experience? ... Why doesn't all this information-processing go on 'in the dark', free of any inner feel?

Claim 5: Affect is an extended form of homeostasis.

33:39 Solms:

My claim is that, if we can identify the *functional mechanism of affect*, of affective feeling, then we *will* be able to explain why this sort of information processing doesn't go on in the dark. So I hope that that's clear. That's kind of like my main point.

My main point is, we've been looking in the wrong place. We've been looking to cortical vision, and cortical perceptual and cognitive processes in general, in order to isolate the neural correlate of consciousness. And Chalmers has said, well, it doesn't work. You can isolate the mechanism whereby this sort, or these sorts, of information processing goes on; and it doesn't tell you anything about "why there is something it is like to see, et cetera."

34:29 And so my argument is, yes, that's true of *those* processes, but it's not true of affect. Remember what I've said to you in my summary now today, that this is the most concentrated consciousness-generating tissue that there is. The reticular activating system and periaqueductal gray is where the lights are switched on; and that you can switch those lights off with as small a lesion as two cubic millimeters in extent. So this is the place where we should obviously have been looking in the first place, in order to identify the neural correlate of consciousness and to understand its mechanism.

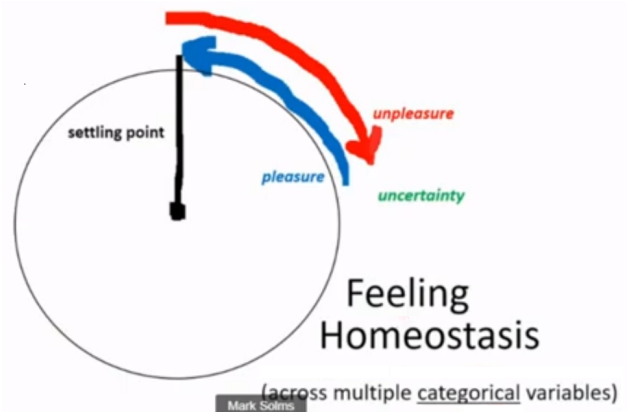
And much more important than that, is the fact that the kind of consciousness it generates, namely feeling, is an intrinsically conscious mechanism - that *feeling* wouldn't exist if it wasn't felt!

35:38 And so unlike vision and learning and cognition in general, which does *not have to* be felt, affect *does* have to be felt. That's the whole point of affect! So I'm saying that, if we could understand the mechanism of affect, in other words, the mechanism whereby feeling comes about, then we might make some progress with this hard problem of consciousness.

So remember, that's how far we got last time. And now I go on - and I'm now going to slow down a little bit, because I'm now going to try to identify what the functional mechanism of affect is.

36:41 So you see, I'm saying "Affect is an extended form of homeostasis."

And by the way, this is not my argument. This is an argument that was (to my knowledge) first introduced by Jaak Panksepp in the 1990s, subsequently popularized by Antonio Damasio. So I'm just summarizing my version of what is in fact now a good 25 year old argument.



37:17 Homeostasis is probably the most basic biological mechanism. It could be said that homeostasis is what enables living organisms to resist the second law of thermodynamics. In other words, it is what enables them to be living organisms. They don't just dissipate; they remain in an organized form. And the basis of this is homeostasis - that rather than just explore all possible states, we living things have to remain within highly specific states.

38:04 And these are called the "settling point" or the "set point," the viable ranges of the organism. And this applies across multiple different parameters. Let me use the example of core body temperature. For those of us who think in degrees Celsius, we have to remain between 36 and a half and 37 and a half degrees Celsius. That's where we need to be.

If we deviate too far from that very narrow range, then we are at risk of dying. And that doesn't apply only to temperature, it applies to water, to oxygen, to salt, to sugar, to blood pressure, to all sorts of things about our bodies. They have to remain within very narrow ranges that are viable with the preservation of our living state.

If we move outside of those ranges, in other words if we explore all possible states, then we die. So we have to work against that dissipative trend, as I said, that is, this **entropic** trend. We have to *resist entropy* and remain within our viable states - very narrow, specific ranges of viability across these multiple dimensions of our physiology. And this is why I say that homeostasis is what keeps us alive.

39:44 Now, a deviation from that settling point then becomes a **demand for work**. The organism has to *do something* to return itself *to* its viable bounds. And that's the basic mechanism of homeostasis. That's how all of those autonomic functions I was talking about earlier, like core body temperature and blood gas balance and so on, that's how they work.

If you're moving outside of your range, you have to do something to return yourself *into* the range. That is what homeostasis is. It's the mechanism that returns us back to our viable range.

40:22 Now, what I'm saying is that affect is an *extended form* of homeostasis. And how it extends homeostasis, is that when we move out of our viable range, *we feel it*. We feel an unpleasant quality. That means "I am moving out of my viable range!" And by contrast, if you're moving back *towards* your viable range, you have pleasurable feeling. This seems to be the basic function of affect: it enables the organism to know *how it's doing* in terms of its organismic viability. So the organism *feels* when it's moving outside of its viable range as **unpleasure**, which means "this is bad!" And it feels, moving back towards its viable range, back towards its ideal settling point, as pleasure, which means "this is good! - this predicts my survival!" - "This predicts my demise!"

41:23 So feelings are rooted in a **value system**. In other words, that there is something *good* and there's something *bad*. And what is good is to survive (and, as it happens, to reproduce). And what is bad is to not do so. This is, of course, the basic value system that underwrites all of life. This is the value system that underwrites **natural selection**.

So feelings are rooted in that value system. And what they do, is they enable the organism *subjectively* - the organism itself - to know whether it's moving out of its viable bounds. In other words, whether it's doing something *bad* within that value system, or something *good* within that value system. That's what feelings do: they enable the organism to know how they are doing within that value system.

42:23 Now, why? Why does this get added to homeostasis? Because not all homeostasis is felt. *Most homeostasis is entirely autonomic*. I was mentioning earlier blood pressure, for example. When you move out of your variable range, your heart rate changes and your blood vessels dilate in order to return you to your viable blood pressure.

You don't need to know - in fact, you *do not* know - anything about it. And in fact, it's clinically notorious how blood pressure regulation works, because you can be *way* out of your viable bounds and know nothing about it. And I'm just using blood pressure as an example. Then there are many, many, many ways in which your autonomic nervous system maintains you within your viable bounds without you knowing anything about it.

43:19 So why do we need to feel it? Well, what feeling adds is, when you are in a situation of uncertainty, a situation where you do not have a readily *pre-prepared reflex* which returns you to your viable bounds - So you don't need to feel how you're doing if you don't have to make any choices! If you have *automatic predictions* which return you by reflex (like in the case of blood pressure regulation) to your viable bounds, then you - *you*, the sentient being - have no part to play in the process!

44:04 Where feeling comes into its own, is where the organism finds itself in a state of uncertainty - for example, in a novel situation for which its innate pre-wired (as it were) reflexes have no preparedness, You find yourself in a state of surprise. And now what you're going to do, if there's no reflexive solution available? - Then a *feeling* enables the organism to make choices.

The choices have to be rooted in a value system. There has to be a good choice and a bad choice. Otherwise it's random.

44:47 So whether you're doing the right thing or the wrong thing, is announced to you by how it feels. So this enables *voluntary* action. I really must emphasize, that: This enables you to *choose*. It enables you to decide for yourself what to do.

"Things are getting worse; so I'm going to change my mind." - "Things are getting better; I will stick with this policy - this is working!" So these are not hardwired. These are not innate predictions. These are choices made on the fly by the organism, here and now. And they enable the organism to survive in states of uncertainty - in other words, in unpredicted situations; in other words, in novel situations - and God knows they are many of those in life.

45:39 So just to put flesh to those bones, let me give you an illustrative example. Normally, respiratory control is autonomic. You don't need to make any decisions about breathing; it just happens automatically. So you automatically remain within your viable bounds in terms of the ratio of oxygen to carbon dioxide.

But that's only when the normal autonomic reflexes manage the situation because you are in predictable circumstances.

46:21 Now imagine that you are in a carbon-dioxide-filled room. Now suddenly you move out of your viable bounds. And breathing normally - normal reflexive regulation of your respiratory system - doesn't work, because ordinary breathing in this carbon-dioxide filled room is going to rapidly lead to your demise.

So now you've got to do something. You've never been in a burning building before, let alone this particular one. So you have no reflex or instinct, no innate prediction, as to what to do. So you *feel your way* through the situation. And please note at this point, at the point when you find yourself in this unexpected situation, this is where *your need for oxygen becomes conscious*.

47:19 So this is a very important point. An otherwise autonomic function now becomes a conscious function. You feel what we call **air hunger** or **suffocation alarm**. And now you move about in this building deciding which way to go. Remember, you have no prior knowledge of what to do. And it's only on the basis of *how it feels*.

So if, for example, you go upstairs and the oxygen supply diminishes, you feel *worse* air hunger. If you go downstairs and there there's a greater provision of oxygen, then you feel better; you feel *relief* from the suffocation alarm. And so the feelings tell you whether what you're doing is working or not. And so your choices are based on feeling.

48:18 And so you are able to feel your way through this problem and survive. This is not a small advantage! The ability to survive in unpredicted environments is an enormous adaptive advantage. And so *that*, we believe, is

what the function of feeling is. That's why in this narrow example, why respiratory control suddenly becomes conscious, that it dramatically intrudes on consciousness - suddenly, your need for oxygen...

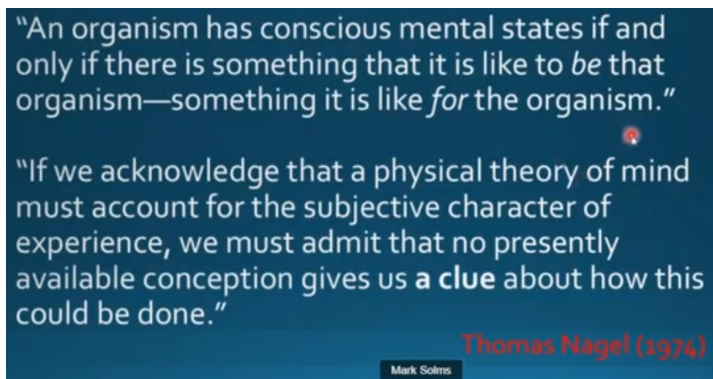
And the purpose of this is to enable you to calibrate your choices, to change your mind about your current policy, on the basis of whether it's working or not - which is exactly what feeling announces for you. In the absence of feeling you would behave randomly; and one in a million [chuckles] will do the right thing, and that one will survive and reproduce; and the rest of you have had it.

49:17 So this enables us to make choices *within our own lifetime*. We don't have to let natural selection do it. Within our own lifetimes, we can adapt to unpredicted situations. Of course, once you've done that, you then also can learn from the experience within your own lifetime. And so the next time you find yourself in a burning building, you might, on the basis of learning from experience, have somewhat less uncertainty about what to do.

So again, let me just make sure that I'm getting across my main point, because this is *the mechanism of feeling*. This is the function of feeling. This is what feeling does. This is why the organism must feel it.

50:15 Now, remember what Chalmers had said, "Why doesn't all this information processing go on in the dark," and so on?

And that was built upon an earlier argument by Tom Nagel, who said, "An organism has conscious mental states if and only if there is something that it's like to *be* that organism, something it's like *for* the organism."



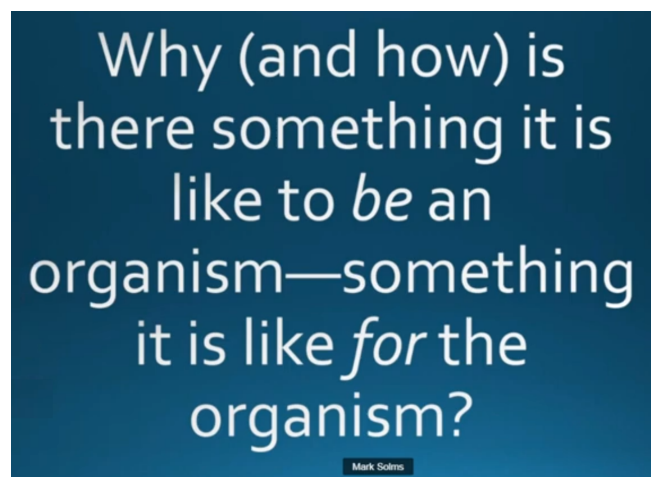
50:35 Then he went on to say, "If we acknowledge that a physical theory of mind must account for the subjective character of experience, we must admit that no presently available conception gives us *a clue* about how this could be done."

So this is what I'm trying to do in this talk: I'm trying to give us *a clue* about how this could be done.

This is Nagel's way of formulating "the Hard Problem:"

"Why (and how) is there something it's like to *be* an organism, something it's like *for* the organism?"

My point is that that question only makes sense in relation to things like visual information processing, which *it does not have to be something it is like* to see; that does not have to be something it is like to discriminate red from blue.



51:24 You can do that unconsciously. The cortex can do that automatically. That kind of information processing *can* go on in the dark.

But would Nagel even have asked this question if we were talking about feeling? I've just explained to you what the function of feeling is. And it makes it kind of absurd to ask, "Why and how is there something it is like to *be* an organism, something it's like *for* the organism?" - I hope that you can see what I mean!

"Why - and how - there is something it is like to be an organism, something it is like for the organism" has everything to do with feeling! *Feeling* is what it is like to be an organism - how you're doing as an organism! How much longer are you going to remain in existence as an organism? *That's* what feeling is like. And this is why and how it exists.

52:23 So if we can get to what the physical mechanism... Going back to what Nagel was saying here - He said, "If we acknowledge that a physical theory of mind must account for this" something-it-is-likeness, then he's saying, "we have to admit that" we currently have no clue - that I'm saying it's because we've been looking in the wrong place.

I think if we look to *feeling*, and we seek a physical theory of feeling, this *will* account for the subjective character of experience, and we will make some progress. This will provide us some clue about how we might go about solving this hard problem.

Let me pause at that point. Having said that "Affect is an extended form of homeostasis" - Let's see if anybody wants to argue the toss or make a comment or ask a question.

Question: The Emphasis on Action

53:31 *Sillett*:

Yes, thank you. Really, really enjoying this.

I'm interested in this *awareness* that leads to the homeostatic sort of correction. And if it's where the awareness is, with the actual cells or organs struggling in some way to act or enact upon the surprisal, be that in the environment that they're encountering at different scales. So this issue of actually acting or enacting, and when that action and enaction can functionally correlate up the nested hierarchies, then it's those actions, or the inability to act in the way that cells would like, that then is becoming the signal of what or where action is failing, rather than necessarily a signal that actually contains information inherently in it.

I wonder what your thoughts would be on that.

54:35 *Solms*:

Yes, I agree, if I'm understanding you correctly. I agree with the emphasis on action. (So I'll slightly run ahead of myself if I say this.) Obviously, a homeostatic deviation can also be construed as a prediction error. In other words, what you're doing has not led to the outcome that was expected.

And there're two ways of correcting prediction errors. You can either change your prediction; or you can change what you're doing, in order to bring about the prediction that you had originally... In other words, when it comes to prediction error, you can either update the prior and have a posterior prediction; or you got to *do* something differently in order to confirm the prior prediction.

55:28 So that's where the emphasis is on action. Now, why this is so important in relation to homeostasis, is that you can't change your prediction about what your viable bounds are! You know, if you expect to be between 36 and a half and 37 and a half degrees Celsius on the basis of acting in a certain way, in other words firing an autonomic reflex; and you then find yourself to be at 39 degrees Celsius, you can't say, "OK, my posterior prediction is that by doing this I'll be at 39 degrees Celsius!" Because if you do that, you're rapidly on the road to death. So the prediction error has to be corrected on the basis of changing your action policy, doing something different in order to confirm your prior prediction.

56:19 So, the emphasis is very much on action when it comes to these organismic predictions. The viable states for your phenotype can't be changed. You can't just change your mind about what you expect will flow from your actions. You have to change your actions in order to bring about the expected or preferred state of your phenotype.

56:53 *Sillett:*

Thanks! Can I just add one piece to that? I think that's really helpful. And also in terms of like multiple scales going down to smaller scales... So for instance, if I actually feel heat in my body, and my body might be, my cells are trying to act to find a better state - that can be seen as oppressive and claustrophobic, if I then read that as being me getting into an uncomfortable state. It could also be me basking in the sun on a beach, you know? In the same way that, if I taste something that's very sour, there's some sort of action at the cell or the organ level reacting to that sour. So that sweetness, which could be like a nice confectionery sweet for a child, you know, like one of these gobstoppers. Or it could be quite a problem.

So I like what you're saying with action. I was wondering whether that action piece is - the enaction is the prediction error in terms of how it can go up in terms of these nested hierarchies of physical scale.

58:07 *Solms:*

Yes. (So again, I must be careful not to run ahead of myself too much. I mean, inevitably, you know, you must ask whatever question comes into your mind at any point. But I am aware that as I go down this list of arguments, I'm going to be addressing those points. So I don't want to make all of those arguments in one go.) So I will just say that for now, that what you've just asked has, first of all, a lot to do with the fact that we have multiple needs; that there's not only one need; that they have to be balanced in relation to each other. And also very importantly, that we're talking about a predictive *hierarchy*; that this is what you're speaking about with reference to *scale*, that what applies at the sensory *periphery* and what applies at the *core* of the predictive hierarchy have different implications for the way that affect works.

So I'm just saying those very kind of vague and abstract things for now, and I hope that the picture will become clearer as I proceed.

Question: Homeostasis is only part of the story

59:20 *Dougllass:*

(This is some thinking that I got into from listening to a very recent interview you did with a young man in another part of Cape Town, talking about perception.)

I think (and maybe this has already been done) - that there's a larger framework in which these notions of what can be perceived, what can be perceived consciously, what is a sensory modality, what we *don't* call "a sense" -

It seems that there are on the one hand, conditions - we've been calling those homeostatic conditions - that you can't do anything about. Your body either adjusts the blood pressure internally or it goes haywire; but you can't be aware of it. So we don't think about this as something that we can consciously perceive; and we don't call the mechanisms that are sensitive to blood pressure and core temperature and other things as "matters of sensation, matters of perception, matters of consciousness."

1:00:37 On the other side - the other side of the "mapping" - there are "single-object influences." A mechanism that is sensitive *only* to blood pressure presents another reason not to call this "a sense." An *eye spot* just tells certain kinds of worms "there's light around here!" or "there isn't light around here!" No directionality, just "it's daytime or it isn't" - "I'm under the mud or I'm not under the mud."

Whereas even a little more information, more kinds of information, qualify this more as "a sensory modality." And in things like hearing and vision and smell, we have this very rich "many-to-many." There's many things you can do about a burning smell; and there are many kinds of odors out there. So does this tie into anything you've been thinking about?

1:01:54 *Solms:*

Yes. So the first thing that you're touching on there, is the fact that *homeostasis is only part of the story*.

In fact, the example I gave earlier, of the person who finds themselves in a carbon-dioxide filled room in a burning building - they need to maintain their homeostatic blood-gas balance. But there's nothing that the autonomic nervous system can do about that.

They therefore need to turn to action in the outside world. So when I said that the person starts moving about upstairs and down; and then feels, "Is this working or is this not?" - "Is this good or is this bad?" - that we call "**allostasis**." It's *acting in the outside world* in order to return myself to my homeostatic bounds.

1:02:55 So that's the first thing I just wanted to make clear. I know it's implicit and I know I'm telling you something that you very well. But I just want to make that *explicit* for our participants - that the importance of the outside world for these internal bodily states, the bridge from homeostasis to allostasis - is how we conceptualize that.

So then it moves to the next point, which is that -

Think about those kids that I showed you earlier who have no cortex. And all they're capable of feeling is their affect. And they can't consciously think, "I feel like this about that." In other words, they don't know what the "about that" is. They just feel things! And they can't include within their consciousness of the feeling what the thing in the outside world is that *contextualizes* that feeling. The *context doesn't become conscious*, only the feeling itself.

1:04:05 To be able to extend the feeling onto the context, to be able to say, in effect, "I feel like this about that," incorporates the context within that *terrain of uncertainty* where one's navigating, one's **palpating the uncertainty** in order to make choices - not only on the basis of blind feeling, but also on the basis of what kind of *object* brings about this change in my feeling, and what kind of object brings about that. For all of that to be incorporated within the realm of consciousness, I think, is a further leap and clearly another enormous adaptive advantage.

So the *first leap* is just to be able to feel the consequences of your actions.

The *second leap* is to be able to *picture*, as it were, those actions within the sphere of consciousness, and in this way to bring the context into consciousness and extend the realm of choices enormously.

Question: Hormesis and homeostasis

1:05:12 *Friedman:*

Thank you, Mark. And I'll just chime in with a third question.

You mentioned homeostasis is only part of the story. And in the chat Brock asked, "What part of the story is **hormesis**?" [That is,] a small stress inducing some kind of benefit over the integrated time horizon. Like in exercise, stress that leads to improved strength. You mentioned allostasis, kind of an anticipatory movement

towards a set point. But where does, for example, induced stress, mild stress and recovery, play into this extended homeostatic or generalized homeostatic framework?

1:05:57 *Solms:*

When one gives an overview of an argument like this, one always oversimplifies; why, one sort of has to simplify!

When I say that homeostasis - that "you always have to confirm your prior prediction," that's not entirely true.

1:06:22 There are also ways in which the homeostatic range can be extended. For example, what I was saying about oxygen - If you're a diver, you can learn how to hold your breath and how to manage the stress of being out of your viable blood-gas range, in a way that a naive person like I cannot.

1:06:56 So there *are* mechanisms whereby these things can - in very, very narrow limits - these things can be changed. But I think the emphasis there has to fall on the narrow limits. Ultimately, there is an outer limit to your viable range. And this is the driving mechanism of the story that I'm talking about here.

1:07:21 *Friedman:*

Thank you. Of course, so much more to say and learn and add.

Claim 6: Complex organisms require multiple (categorical) homeostats

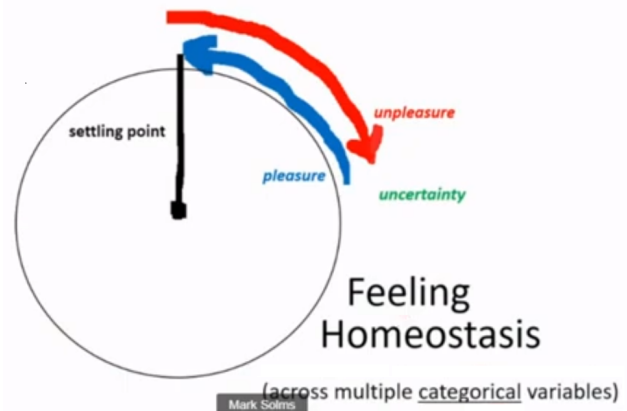
1:07:31 Solms:

Thank you. Well, you can see highlighted on the screen now my next point. And it might seem like a small point, it might seem like an obvious point, that complex organisms require multiple homeostats. And in fact, I've more or less said this already, but I want to make explicit...

And here is the same slide that you saw a moment ago, where I've just added the point that I have now that I'm now making.

The importance of this point is that *we have multiple needs which must be categorically distinguished from one another.*

In other words, we can't have a continuous variable called "Need." That variable has to be factorized across a number of different **categories**.



1:08:24 So let me be clear what I'm talking about. If we quantify on a continuous scale how much deviation there is from where you need to be, let's say, "OK, I've got six out of ten of thirst, and I've got four out of ten of sleepiness. Then all that means. I've got ten out of 20 of total need. And so all I need to do is sleep. I don't need to drink."

If I can generalize from that little example there, the point that I'm trying to make is that, No, that's not true! You have to sleep *and* you have to drink *and* you have to eat *and* you have to defecate *and* you have to breathe! You can't just summate all of these and bring down the total number. If you did that, you would die. So we have to recognize that each one of these needs has to be met in their own right. And that means we can't just have a total variable called "Need." We need to know *which need* are we talking about.

1:09:44 And this is why they have to be treated as **categorical variables**. In other words, here we're talking about how much water I'm lacking; here we're talking about how much oxygen I'm lacking; here we're talking about how much sleep I'm lacking; et cetera - so that the category of need that the organism finds itself in can be addressed appropriately. The appropriate category needs to be addressed.

And I know what I'm saying sounds absolutely obvious. But why it's important is that it's because of this, because of the points I've just been making, that hunger feels different from thirst, and thirst feels different from suffocation alarm, and suffocation alarm feels different from sleepiness, and sleepiness feels different from pain, and so on; because each one of these *feelings* - the *quality* of the feeling - tells us which need is at issue. And categorical variables are distinguished qualitatively.

1:10:52 So we are not only talking about *valence*, goodness or badness; and *arousal*, how much or how little. We're talking also about *qualitative categories*. And I think that's important when we come back to the whole point of what I'm talking about - namely **qualia**, namely the qualitative *stuff* of what it is like to do anything, be anything - those qualitative distinctions.

It's of the *essence* of consciousness that it is qualitative. It has qualitative differentiations. It's not just some total quantitative, some total continuous matter of need. It's a matter of different categories of need, which therefore are different qualities of need.

1:11:47 And I think that's important in terms of understanding what an affect is. An affect is *a state of the organism registered by the organism*. In other words, it is intrinsically subjective. It has intrinsic value, valence, goodness and badness, which has existential consequences for the organism. And in addition to that, it has *quality*. It is inherent in the nature of affects that they are qualitatively distinct from each other. And I hope that this very simple, functional, mechanistic reasoning that I've given you as to why that *needs* to be the case, why that *must* be the case, helps to make clear why affects take the form that they do.

Remember, what we are doing here is addressing Nagel's question about, "Why is there something it is like to be an organism; and how does it come about there is something it is like to be an organism, for an organism?" That something-it-is-like-ness is not just a valenced goodness and badness, but also a qualitatively differentiated state, *subjective* state of the organism.

1:13:02 So I think that by reducing the mechanism of affect to these essential features, we begin to see why it takes the form that it does, why it is a necessarily conscious "something" in the rudimentary sense that I'm defining consciousness. So that's why this point is important. That's why I wanted to make the point clearly that complex organisms require multiple categorical homeostats.

They need to be treated as categorical variables. They need therefore to be qualified. Each of these different factors of need, each of these different categories of need, have to be qualitatively differentiated from each other. And this is, I think, the Ground Zero of where qualia come from.

Claim 7: These homeostats must be prioritized

1:13:57 *Solms:*

Now, I'll go on to my next point, which is that these different categories of need must be *prioritized*. Again, I think this is in a way an obvious point. But I want to just dwell on it for a moment, to draw out some of the implications of this.

I said earlier that our autonomic homeostats run - well - autonomically, most of the time! In other words, there are automatic automated predictions, which we call reflexes, that are - when the organism moves out of its viable bounds, then it fires these reflexes, which bring it back *into* its viable bounds.

1:14:53 And then I told you that not all of our needs can be dealt with that way. Some of them, as we move outside of the range of what the autonomic reflex can achieve - like, for example, when it comes to thermoregulation, you start to perspire if you get too hot, this is a reflex. You start to breathe more shallowly and rapidly - panting, in other words - that's a reflex. These are autonomic responses to overheating.

Then they reach a certain limit, you know, that they haven't cooled you down adequately. So then you've got to *do* something. You're going to do something like leave the kitchen. And what I've just said applies to *one* of your multiple homeostats. In other words, one of them now requires allostatic action. And you now need to do something in the outside world. And clearly you can't do something in the outside world in relation to all your needs simultaneously!

1:15:59 So all the time you're sliding. Your hydration, your water in relation to salt content of your body [is] sliding all the time; it doesn't mean that you're thirsty all the time. You're burning up the sugars in your adipose tissues all the time; it doesn't mean you're hungry all the time. So the question is, What gets prioritized? There must be some prioritization because you can't do everything at once. When it comes to voluntary activity in the outside world, you have to prioritize. There's an action bottleneck.

And the point I'm making here is first of all, just simply that: that you do have to prioritize. And secondly, that what is prioritized is what becomes conscious. In other words, as I gave in the example of suffocation alarm, that business carries on unconsciously until it becomes prioritized. At that point, it forcibly intrudes on consciousness. And my priority now is "I need oxygen!" And that's what you feel.

1:17:11 (This is the other point I'm making.) The other needs don't disappear at that point. They continue to be regulated, but they are regulated automatically. So the prioritization of a need brings that need into the realm of **palpating uncertainty**. In other words, into the realm of "feeling your way through the problem." And this, I think, is an important part of how feeling works.

The periaqueductal gray

1:17:43 Remember I said I was going to talk about the periaqueductal gray. And I was going to tell you how this plays as important, if not *more* important role in the basic machinery of consciousness of the upper brainstem, than the reticular activating system.

The periaqueductal gray, which is just a 14-millimeter-long columnar structure around the central canal of the midbrain - *all* of our homeostatic mechanisms (in other words, all of the multiple homeostats - and we have many!) - all of the multiple homeostats, all of them send their residual error signal to the periaqueductal gray. The periaqueductal gray is like a final common pathway of all of these homeostatic error signals. And it seems that this is where the prioritization must be going on. The determining which of the current error signals is the most salient then gives rise to...

1:18:57 And this is an argument I must again make clear. This is not my own novel insight. This was beautifully argued on comparative anatomical grounds by Björn Merker in a brilliant [2021] paper in *Behavioral and Brain Sciences*. He recognized that the periaqueductal gray, together with the superior colliculi (that I showed you earlier) and the **midbrain locomotor region**, they form what he called a **midbrain selection triangle**.

1:19:35 In other words, this is where the affect, the need that is to be prioritized, is selected. And this gives rise to a feeling, the feeling being (as in the example that I gave with suffocation alarm) the need that is prioritized, the need that is currently most salient, the need that is going to now color your consciousness, qualify your consciousness, in terms of its most rudimentary property - in other words, the feeling, the affect: "What organismic state am I in?" - "I'm in a state of respiratory distress! I'm in a state of suffocation alarm! That's what I feel!" Why? It's because this is where choices need to be made. This is where the creature needs to feel (the person, in that example), needs to feel their way through the problem.

1:20:25 The other needs remain. But the other needs, the non-prioritized needs, are not raised to the level of feeling. And I hope it's clear why. We need to understand mechanistically, How does all of this work? And that leads to my next point.

Before I go to that next point, let me just pause again in case there any comments or questions about this prioritization function performed by the periaqueductal gray, which is what determines what affective state you are going to be in from one moment to the next. In other words, which need is going to qualify your affective state.

Question: The "Tomkins/Panksepp Tradition?"

1:21:12 *Douglass*:

In Sylvan Tomkins' later work (that is, after 1970), he came to elevate RESET to the status of a fundamental emotion. And I think it's very relevant to this - the notion of suddenly shifting from one conscious emotion or one conscious activity or world attitude to another.

I believe Jaak Panksepp drew pretty heavily on Tompkins. And the reason that he doesn't cite him in a lot of detail is that he accepted so much of it. [Edited for length.]

1:22:23 *Solms*:

I agree with that [second point] completely. Tomkins, the reason that Panksepp barely ever... He cites him usually in the beginning of a kind of a general discussion, you know, just to say "this is the tradition I belong to," and then carries on building on his shoulders. I agree with that very much.

1:22:50 *Sillett*:

I liked when you talked about "palpating uncertainty," and the salience around that. That's a nice term. And I think that relates to "What is going on?" so to speak - "So what?" about what's going on. And then, "Now what happens next?"

And I'm wondering, though, as well, if the idea of "where" – which I think is often left out. There's a lot of categorization of "what" things are, but "Where is it happening?" - that makes a difference, in terms of how significant it might be in terms of prioritization. So I was wondering how much the risk of the "where-ness" of some sort of disequilibrium happens, could be featured in this. And maybe that's where the physicalization and the spatialization becomes much more significant, than when you're just dealing with the brain.

The control centers are in the brain

1:23:50 *Solms:*

Yes. Well, all of these homeostatic mechanisms - of course, the physiological processes that they regulate are widely distributed. But the control centers of these homeostats are, for the most part, in the **brainstem-broadly-defined**. And I say "broadly defined" because there are some that are in the medulla oblongata and pons, but there's some that are in the midbrain, there are some that are diencephalic!

For example, let me just mention the most outstanding example. The hypothalamus is full of homeostatic control centers! But hypothalamus, circumventricular organs, the parabrachial complex, the area postrema, the nucleus solitarius, these are all homeostatic control centers. But of *fundamental* importance - and I think it's not sufficiently recognized, although Merker certainly made much of it - that all of these nuclei that I have just enumerated, they in turn project to the periaqueductal gray.

1:25:11 The periaqueductal gray is in my view a **meta-homeostat**. It's sort of the control center of the control centers. And the essential function there is a prioritization function. And by being able to *physically locate* these mechanisms - bearing in mind where I started - I started my answer to your question by pointing out that ultimately these homeostats regulate physiological processes which *cannot be localized*. They are distributed processes *par excellence*.

But the center - the control center - *can* be localized. And those control centers (they are numerous ones) - they in turn *all* send their residual error signal to periaqueductal gray. That is really important! Now, what that enables us then to do, is to *test* models like the one that I'm describing to you. What happens when one or another of these individual homeostatic control centers is lesioned? What happens when the periaqueductal gray as a whole is lesioned? What happens (I might as well just insert that here) is you get a **persistent vegetative state**.

1:26:34 In other words, you get that condition that [chuckles] Magoun and Moruzzi led us to believe is a theoretical possibility - in other words, **blank wakefulness**. Remember I said earlier, in my summary of what I said last time I spoke - that the idea that "the reticular activating system provides merely a quantitative level of consciousness," and not any quality or content, that that fictional state of affairs - of "blank wakefulness" - the thing that the reticular activating system was supposed to be producing - which, as I've showed you, is not the case. Because all of those different lines of evidence that are summarized show that what the reticular activating system is producing is anything but "without quality and without content," that this is in fact the foundational form of the qualities and contents of consciousness, namely the different affects. If you lesion the periaqueductal gray, *then* you get that fictional state, that artificial state, of blank wakefulness. In other words, these patients show non-responsive wakefulness.

1:27:56 They still have the autonomic sleep-waking cycle. In other words, they wake up in the morning and go to sleep at night, like those hydranencaphalic kids that I showed you. But *unlike* those kids, they show *no* emotional, no affective, no response to their situation, and **no intentionality** either. And so, you know, that's what you would expect would occur, if the periaqueductal gray is where the affects are actually being generated.

Because, remember, what I'm saying. I'm saying that the prioritization of a homeostatic error signal *is* the feeling of that signal - that is the **rendering conscious** of that signal. And that is what we *mean* by "feeling something." It means, "Now this need is not just a need - Now this need is a **drive**! Now this need is *driving my voluntary behavior*!"

And that is *why* it becomes conscious! Because, as I said to you earlier, "voluntary behavior" can be defined as "the making of choices," as opposed to "the executing of automatized predictions."

1:29:08 *Friedman:*

Just to add two notes on that:

When hearing Stephen's questions about spatial localization and how that can refer to brain localization, as well as in the peripersonal space, it made me think about how awareness - "*a*-whereness" - it could be "without whereness." Awareness could be spatially dislocated; or awareness might be a highly spatialized percept.

And then the other note that I wanted to make was: As a researcher of distributed physiology in the eusocial insect colony, this extended homeostatic perspective leads to many interesting connections when we think about extended social homeostasis as well.

1:29:59 *Solms:*

Very good points!

Claim 8: The mechanism of perceptual (and cognitive) consciousness is precision modulation of allostatic prediction errors

1:30:09 *Solms:*

(It's really quite amazing, you know! Because you are so generous with giving your speakers two hours, you lull us into a false sense of security, and thinking, "Well, I can just elaborate as much as I want to! There's plenty of time." [Laughs.] And now we're down to our last half hour, and I've still got three points to go!)

1:30:38 *Friedman:*

So it's like, "No one goes there any more! It's too busy."

1:30:43 *Solms:*

Yeah. Well, again, grateful for the time that you're giving me.

So let me move to this point. And this has everything to do with the relationship between *affective* consciousness and *cognitive* consciousness. I spoke earlier about "I feel like *this* - *about that*." In other words, the incorporating of the cognitive domain within the sphere of feeling. And so this is the point that I'm making now.

This is a diagram that comes from a paper that I wrote with Karl Friston [2018].

And here the important equation is the third one.

And here this diagram basically is just trying to spell out in a visual form what these equations are saying. So when I said earlier that we have two ways of reducing prediction error when a prior prediction does not lead to the sensory state that's expected, then we can either change our perception; in other words, we can change the prediction. - Or we can change our action; we can do something differently in order to bring about the prior prediction. So that's what these two equations describe, that we have a generative model which generates predictions as to what sensory state will flow from our actions. And so here there's an action in the external world which is predicted to bring about a certain sensory state.

1:32:38 And to the extent that it does not bring about that sensory state, in other words, the difference here is the prediction error. And of course the prediction error is then used to update the generative model in order to give rise to better predictions, in order to better maintain your expected sensory states. I need to emphasize here (because this was not the case in the early days of the predictive processing paradigm that Karl Friston and his colleagues unleashed [chuckles] upon the world):

"Sensory states" do not necessarily mean "*exteroceptive* sensory states." And so, in the examples that I have been giving in this talk so far, this would be the equivalent of, for example, core body temperature or blood oxygen level, et cetera. The most important sensory states for the organism are its viable states in terms of its homeostatic expectations.

1:33:46 So please remember that, when I speak about "actions in the world," the "world" that we're talking about here can be the visceral body, and it can be the external world. The body is as much external to the [Markoff] blanket, as the external world is external to the [Markoff] blanket of the nervous system.

$$\frac{\partial}{\partial t} M = -\frac{\partial F}{\partial M} = -\frac{\partial F}{\partial e} \frac{\partial e}{\partial M} = \frac{\partial \varphi}{\partial M} \cdot \omega \cdot e$$

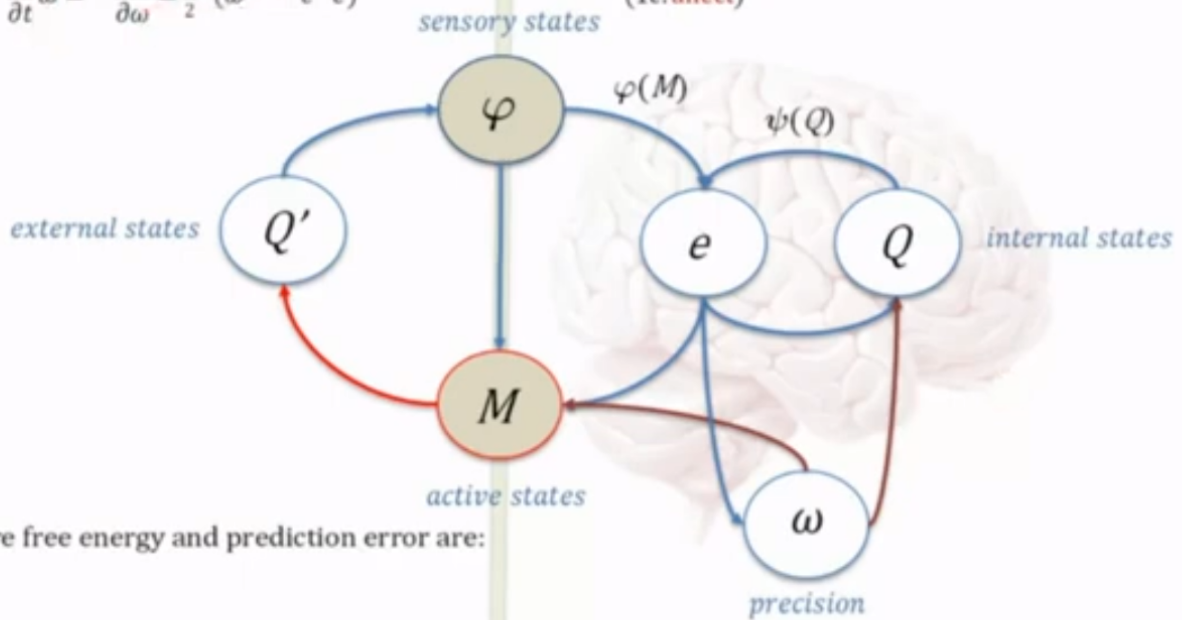
$$\frac{\partial}{\partial t} Q = -\frac{\partial F}{\partial Q} = -\frac{\partial F}{\partial e} \frac{\partial e}{\partial Q} = -\frac{\partial \psi}{\partial Q} \cdot \omega \cdot e$$

$$\frac{\partial}{\partial t} \omega = -\frac{\partial F}{\partial \omega} = \frac{1}{2} \cdot (\omega^{-1} - e \cdot e)$$

(1a: action)

(1b: perception)

(1c: affect)



Where free energy and prediction error are:

$$F = \frac{1}{2} \cdot (e \cdot \omega \cdot e - \log(\omega))$$

$$e = \varphi(M) - \psi(Q)$$

Mark Solms

Solms & Friston (2018)

This equation [(1c: affect)], here... which is written in words, here: "The rate of change of precision," which is omega (ω) in the equation... "The rate of change of precision over time depends on how much free energy [F] changes when you change precision. This means that *precision will look as if it's trying to minimize free energy*. The rate of this free-energy-minimization process is the difference between the inverse precision and the sum of squared prediction errors.

The rate of change of precision (ω) over time depends on how much free energy (F) changes when you change precision. This means that **precision will look as if it is trying to minimise free energy**. The rate of this free-energy-minimisation process is the difference between the variance (the inverse precision) and the sum of squared prediction errors ($e \cdot e$).

Mark Solms

1:35:02 So that's this equation [(1c)], here, put into words. And it's foregrounding the central role that precision plays in minimizing prediction error. So I want to be clear that this *precision modulation* is, in physiological terms, the modulation of **postsynaptic gain**.

So the message passing going on between error signals and prediction signals - it's the reticular activating system's *modulation* of that message passing. It's the increasing or decreasing of the gain on the error signals. That's the role of precision modulation, which is just the same thing as to say "that is the role of the reticular activating system."

1:36:04 It is modulating the gain in the message passing, to speak physiologically. To speak computationally, it is a matter of increasing or decreasing precision values attached to the predictions over the errors. To put it into different words: If things are turning out as expected, that's good; if uncertainty prevails, that's bad. So increasing confidence in a prediction is good; in other words, increasing precision in the prediction is good.

1:36:47 Increasing confidence in an error signal is bad; in other words, the more uncertainty, the more you become clear that things are not turning out as expected, to that extent (of course), your confidence in your current policy is reduced. In other words, the precision in your current policy is reduced, and that just is bad. So the goodness and badness, the pleasure and unpleasure function of precision modulation that I described to you earlier - it has this enormously important contribution to make to the whole of this mechanism by determining the influence of the error signal over the predictive model.

1:37:36 So to the extent that precision is reduced in the error signal and thereby confidence is maintained in the policy, to that extent, the error signal will or will not have influence over the parameters of the predictive model. This is trying to illustrate the crucial role that affect plays in this whole predictive mechanism.

So this is, as it were, the role of the reticular activating system, of the *modulation* of the message passing that goes on in... These are, as it were, the synaptic transmission mechanisms [Solms is moving from the Phi through e to Q, this is the "inbound" pathway of sensory perception]. And these are postsynaptic modulatory mechanisms [Here he is pointing to omega (ω), which is a Precision modulating variable, so the neuro-chemical interpretation of that (under the understanding that the previous "these" are like the neural firing rates in synaptic model) the omega is playing a role of postsynaptic precision-modulating or simply modulating mechanics.]. And so I'm just wanting to link these formalisms to the affective role that the reticular activating system plays.

1:38:39 So just to go back to the statement here: "**The mechanism of perceptual (and cognitive) consciousness is precision modulation of allostatic prediction errors.**"

In other words, it is the modulation of confidence in a current policy. In other words, the perceived consequences of a current...

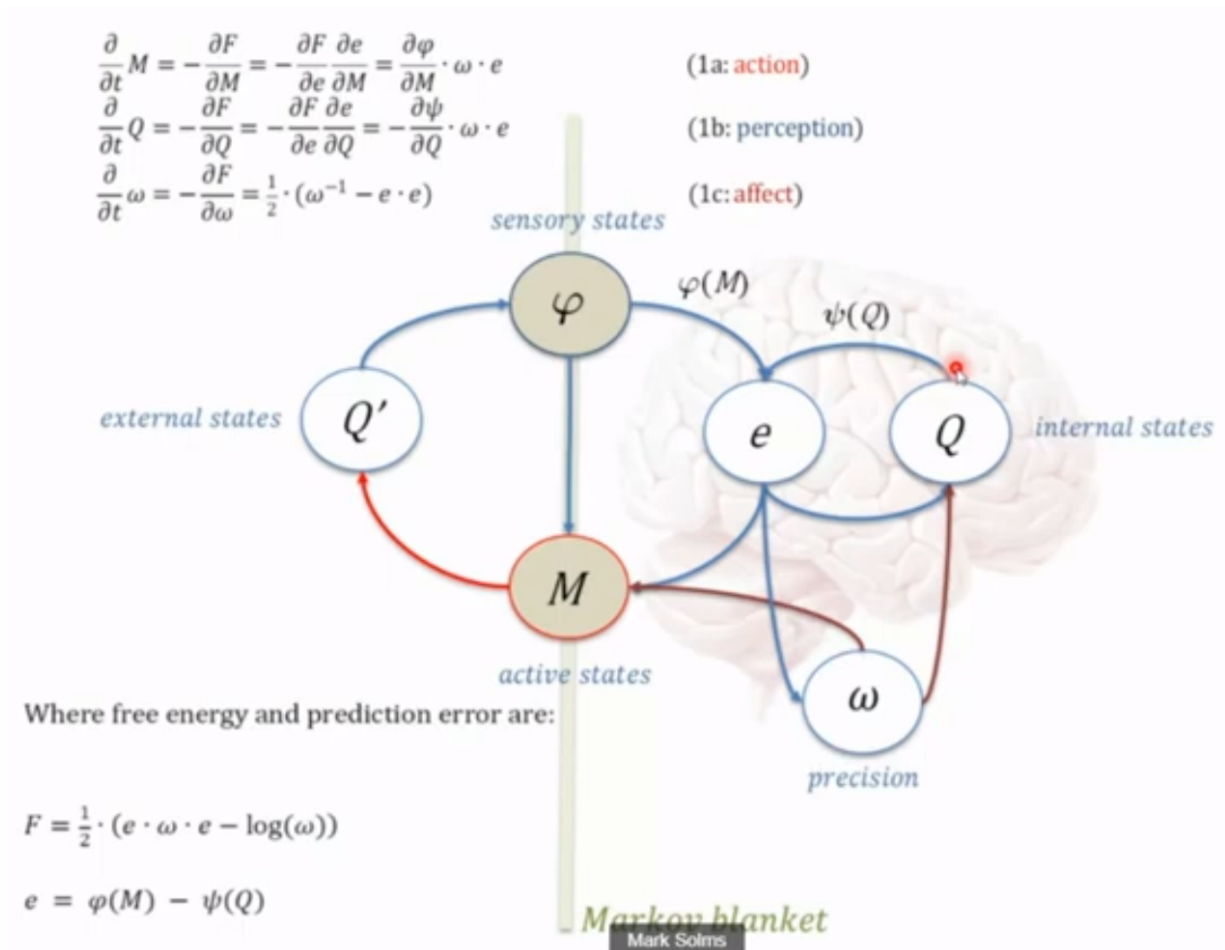
So, as a particular need is prioritized, that need is felt. This generates a *category* of predictive policies. In other words, "This is what I do in this situation, in *this* context that I find myself in. This is what I expect the consequences will be." And there's an expected precision attaching to the error signals, of course. All the other domains of need...

1:39:39 Remember what I was saying earlier, here, about "*The needs must be prioritized.*"

The other domains have monotonous precisions.

So the crucial mechanism in terms of *voluntary* action (to use the phrase that one of you said you liked earlier) is the *palpating of the uncertainty* - in other words, the palpating of the precision and the adjustment of the confidence in the current perceptual and cognitive - the *allostatic* - aspect of "What I must do about this **need state.**"

Because there's a state of uncertainty that's being prioritized there, there's a changing of one's mind on the fly. This is what this mechanism makes possible.



1:40:31 So [indicating oval marked " ω , precision"] the affects are, as it were, a *drive* for or a *demand* for predictive work, for mental work.

And *this* [indicating ovals marked e and Q], the modulation, the palpating of the confidence in the policy over the error signals *is* the predictive work so demanded.

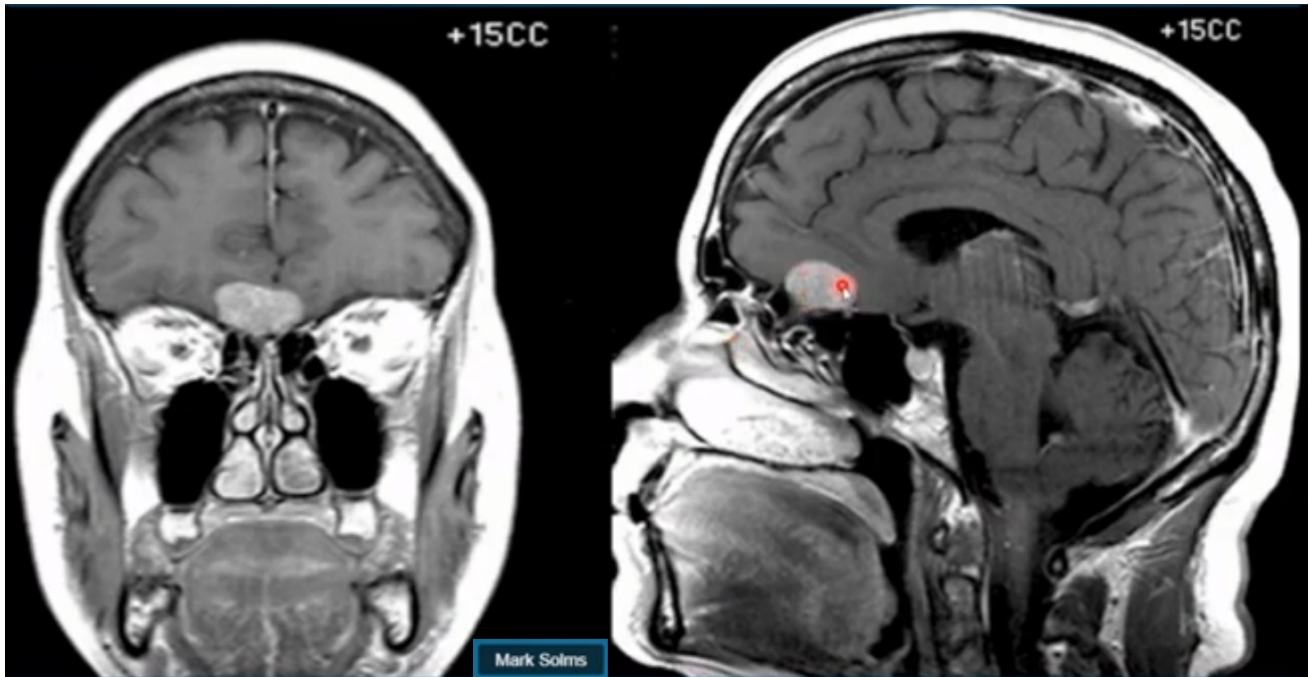
So in other words, the work of updating one's policy, of changing one's mind - it is, all of it, underwritten by the **affective demand**. So the affective demand for predictive work is what gives rise to the predictive work itself, which is the palpating of uncertainties in the current policy and the sensory states that it gives rise to. And so this is *changing your mind* - in other words, voluntary action - in other words, the capacity for choice. This is the crucial role that precision modulation plays in that process.

[A]ny questions or comments about that mechanism?

1:42:08 *Friedman*:

There's always so much to unpack with defining the variables and understanding what the edges mean and so on! But I think [your next] case is a good way to close this.

Case 5: Mr. S's tumor, and his confabulatory amnesia

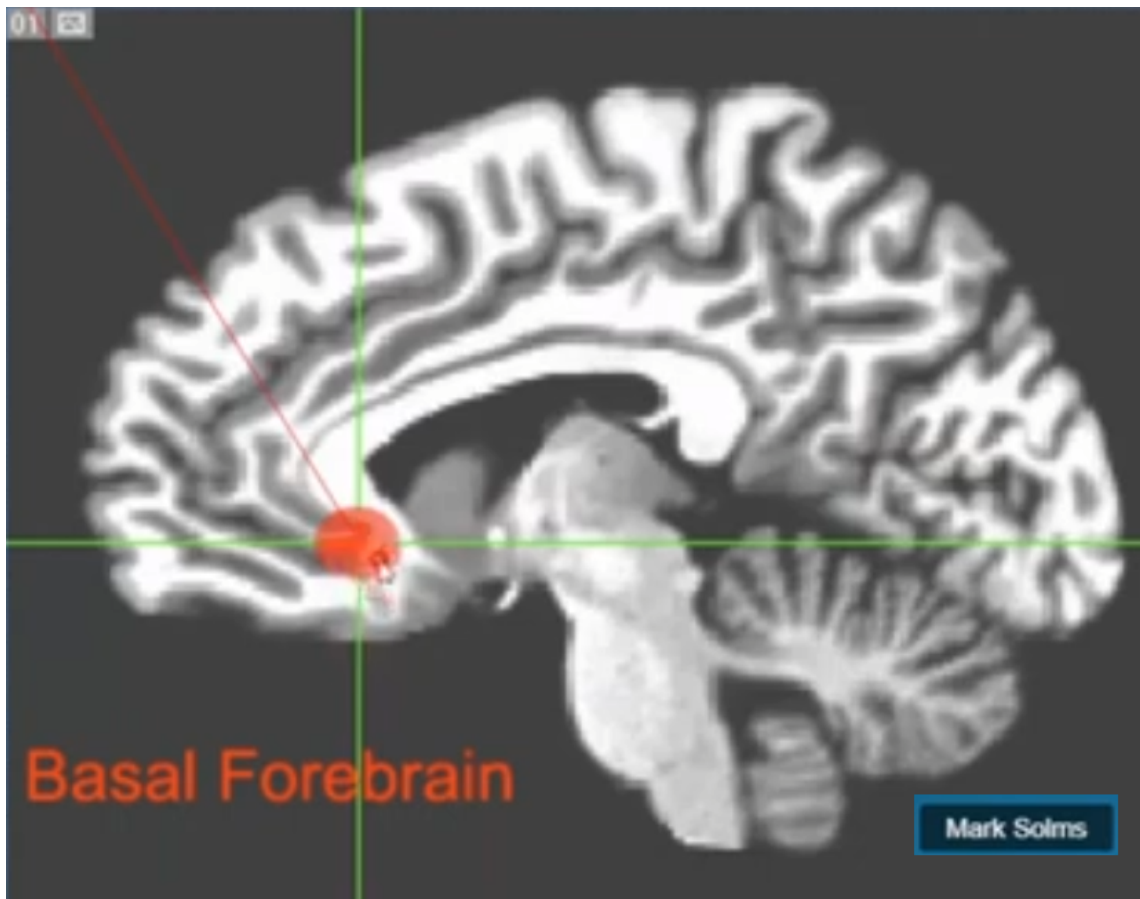


1:42:17 This is a patient of mine, Mr. S. And he had a meningioma, here, at the base of his frontal lobes. It was an olfactory sheath meningioma. And it pushed on his optic nerves. And as a result of that, his vision was impaired; and this is how the tumor came to attention. And it was successfully surgically resected.

Now because of its location in relation, in fact, to the optic nerves, there was some nervousness on the part of the surgeon to remove it in its entirety. And in fact, he felt it was not possible to remove it in its entirety. So he left a little nub of the tumor. And this regrew; and so the patient again noticed these visual difficulties, returned to the surgeon, and the operation had to be repeated.

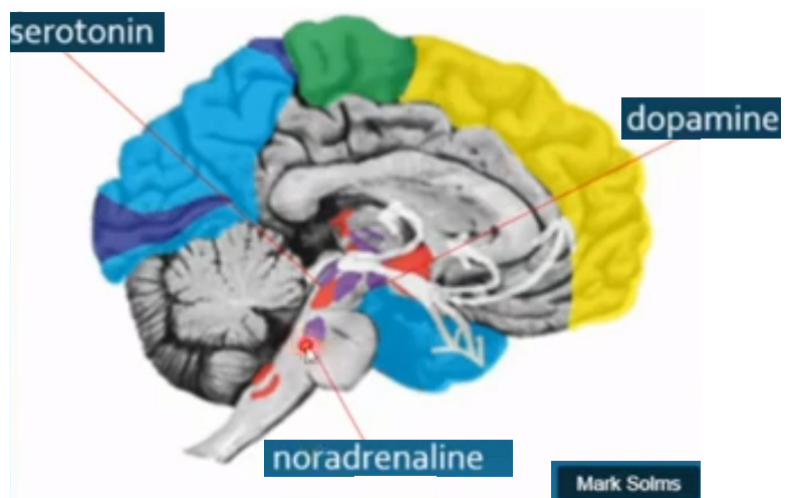
1:43:22 Because of scar tissue, the second operation is always trickier than the first. And unfortunately, in the second operation, there was a bleed, and that bleed was into the basal forebrain nuclei.

Small bleed; but the basal forebrain nuclei are crucial. These [nuclei] are the upper end of the...



Remember - I was saying earlier [22:43] that these different parts of the ascending arousal mechanisms of the brain - that they are the source nuclei for these different neuromodulators.

And I spoke of dopamine and noradrenaline and serotonin earlier.



1:44:03 Well, these nuclei are the source nuclei (there are other source nuclei for acetylcholine here too); but these are very important source nuclei for acetylcholine. Now, remember all of these - what I was saying earlier about these neuromodulatory systems is that they are modulating postsynaptic gain - which is just the same thing, physiologically, as to say, computationally, they're modulating *precision*. That's what they do - they up- and down-regulate the precision in the message passing. And this, as I said earlier, it dictates which messages are going to be selected and which not. So it plays a crucially important role, does precision modulation in cognition.

1:44:55 And I'm now wanting to show you this relationship, the relationship between the affective mechanisms that I'm talking about, that is the affective functions that are performed centrally by periaqueductal gray; and then how this gives rise to the modulation through these arousal systems of these different neurotransmitter systems. And I'm going to now show you how this worked in this case. Acetylcholine modulates the *confidence* in error signals.

And so let me just quickly tell you a little bit more about this patient. He was 56 years old at the time of the second operation. And as a result of the damage to the basal forebrain nuclei from the second operation, he woke up from the surgery with a condition called **confabulatory amnesia**. So although the tumor was resected successfully, and the visual problem was corrected, he now had this devastating new condition called confabulatory amnesia.

1:46:08 [In] confabulatory amnesia, the patient - it's not only that they are amnesic, in other words that they are unable to remember, particularly recent events. But there's also quite a long retrograde extension too - in other words, there's quite an impairment of their retrograde memory too. Not only do they have this, but they also are not *aware* of their memory problems!

And so when they draw up, attempt to retrieve a memory, the memory that they retrieve is not the correct one. And they don't realize it's not the correct one. And so they have what appear to be false memories. These are the confabulations. They frequently are related in some semantic sense with the target memory, with the memory that they're looking for. There's some semantic relationship between the memory they find and the one they're looking for; but they can be grossly misplaced in space and time.

And so these are what we call confabulations. And the patient is not sufficiently critical of these misrememberings, as a result of the damage to these neuromodulatory mechanisms.

1:47:28 So this patient, Mr. S, just to give you one example (I mean, an extreme example): He had his operation in Johannesburg, which is in South Africa, which is where I hail from. But at the time I was living and working in London. And so the surgeon (who I knew well) sent the patient to see me, to consult me, in London, because I was doing work with this condition, confabulatory amnesia.

So the patient arrives in London on a Friday and comes to me on the Monday; and has no idea that he's in London because of course he doesn't remember the journey. He can't remember anything from one minute to the next! And so I say something about, "Well, you know, the surgeon (whose name was Mr. Miller), the surgeon referred you to me because of this memory difficulty that you're having. That's why he sent you to London."

1:48:30 And he said, "London! What do you mean, London?" And I said, "Yes, you know, you're in London. You don't realize it because you don't remember the journey. You know, that's the whole point. This is the kind of problem that you having." And he denies it. He says he's not in London. So that's the confabulatory aspect, that these patients - In his memory, he's in Johannesburg; and so he believes he *is* in Johannesburg.

So I point out to him - It was winter, and it was snowing outside. I point out to him the wintry conditions outside. You never, by the way, have snow in Johannesburg. So I say, "Look out the window!" He looks out; is absolutely shocked; but then retorts, "No! I know I'm in Johannesburg. Just because you're eating pizza doesn't mean you're in Italy!" That's what he says to me. So in other words, "You know, you mustn't overrate the evidence of your senses!" That's just an extreme example of what I mean by how these patients - their amnesia, it's not just a *lack* of memory. It's also an **excess of confidence** in the incorrect memories that they draw up.

1:49:50 So that's the background. The whole point of him being referred to me is because I treat such patients. And so I then saw him six days a week at the same time in my outpatient clinic at the Royal London Hospital.

Same time, same place he came with his wife to the waiting room. I would then go and collect him, take him up to my consulting room, spend an hour with him, take him back down to the waiting room.

And then I would talk to his wife, because he was so full of confabulations I needed to verify things in order to get some sense of what was going on. And there were certain themes - and this is an interesting thing about these patients, these confabulatory patients. There's certain themes that returned again and again. The patient was an electronic engineer, in reality he was. But he thought the reason he was coming to me, frequently he thought he was coming to me because I was consulting *him* about an electronic problem.

1:51:00 And otherwise he thought that the two of us were electronic engineers together working on some electronic problem. Also, frequently he thought that he and I were in some sporting team together, that we played rugby together, or we were in this rowing team together. Now, his wife told me he *had* played rugby at university - a good 30 years or more before. And likewise he had been a keen rower; but this too was at university more than 30 years before.

And so these are again good examples of confabulations - you know, the patient, mislocating in space and time a memory that he draws up now; and he has too much confidence in that - he too readily accepts the veracity of the products of his own memory search. So that's the background.

1:51:59 Now on this particular day that I want to give you a little snippet of a session from - He on this day, when I came to the waiting room he touched the scar on his head, the craniotomy scar on his head, and he said, "Hi, doc!" to me! So this was progress. You know, for the first time he was associating me with medicine; and he was associating me with the surgical scar on his head.

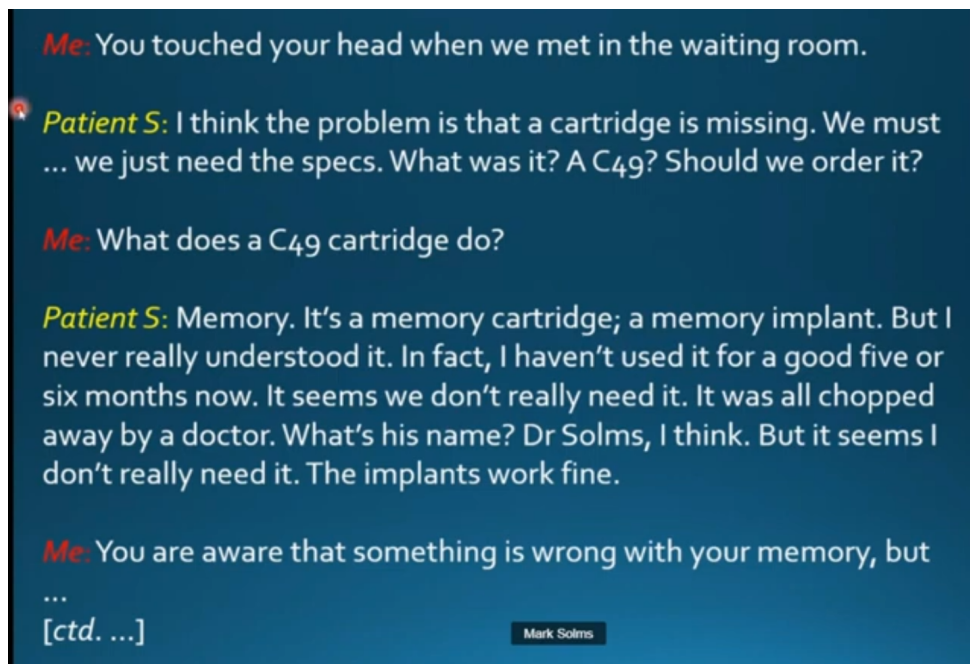
And so I thought this was great progress. And so when we got into my consulting room, I said to him,

"You touched your head when we met in the waiting room."

And he said, "I think the problem is that a cartridge is missing. We must... we just need the specs. What was it? A C49? Should we order it?"

So I said to him, "What does a C49 cartridge do?"

He says, "Memory. It's a memory cartridge. A memory *implant*."



(His wife told me that he had had dental implants. He'd had serious problems with his teeth, for many years. But these problems had finally been successfully treated by implants, teeth implanted into the jaws.)

1:53:06 So when he said that C49 cartridge is a *memory* cartridge, and that it's a *memory implant*, it brought that dental operation to my mind.

He said, "But I never really understood it. In fact I haven't used it for a good five or six months now. It seems we don't really need it. It was all chopped away by a doctor. What's his name? Doctor Solms, I think. But it seems I don't really need it. The implants work fine."

So I said to him, "You're aware that something's wrong with your memory, but..."

And he interrupts me, and he says,

"Yeah, it's not working 100%, but we don't really need it. It was just missing a few beats. The analysis showed there was some C or CO9 missing. Denise brought me here to see a doctor. What's his name again? Doctor Solms or something. And he did one of those *heart transplant* things."

So this is referring to another operation that he'd had, which is clearly also being referred to *here*, which was that he had a cardiac arrhythmia, and so he had a pacemaker fitted, a cardiac pacemaker.

[... *ctd.*]

Patient S: Yes, it's not working 100%, but we don't really need it – it was just missing a few beats. The analysis showed that there was some C or CO9 missing. Denise brought me here to see a doctor. What's his name again? Dr Solms or something. And he did one of those heart transplant things, and now it is working fine; never misses a beat.

Me: You're aware that something is amiss. Some memories are missing, and, of course, that's worrying. You hope I can fix it, just like those other doctors fixed the problems with your teeth and your heart. But you want it so much that you are having difficulty accepting that it's not fixed already.

Patient S: Oh, I see. Yes, it's not working one hundred per cent. [He touches his head.] I got knocked on the head. Went off the field for a few minutes. But its fine now. I suppose I shouldn't go back on. But you know me; I don't like going down. So, I asked Tim Noakes – because I've got the insurance, you know, so why not use it, why not go to the best – and he said: "Fine, play on."

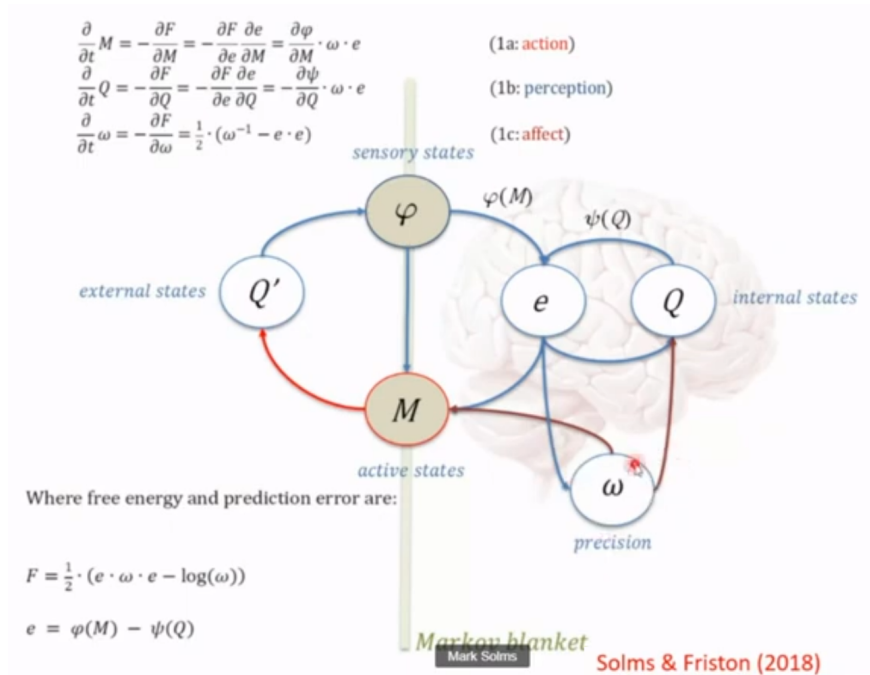
Mark Solms

1:54:15 So he said, "So he did one of those heart transplant things, and now it's working fine; never misses a beat."

So I said to him (which is what I actually thought) - I said, "You're aware that something is amiss, some memories are missing. And of course, that's worrying. *You hope I can fix it*, just like those other doctors fixed the problems with your teeth and your heart. But you want it *so* much that you're having difficulty accepting that it's not fixed already."

So he goes on and says, "Oh, I see. Yes, it's not working 100%." And he touches his head. He says, "I got knocked on the head, went off the field for a few minutes, but it's fine now. I suppose I shouldn't go back on, but you know me, I don't like going down. So I asked Tim Noakes..." (Tim Noakes, by the way, is a sports physician in South Africa.) He says, "So I asked Tim Noakes - because I've got the insurance, you know, so why not use it? Why not go to the best? And he said, 'Fine, play on.'"

1:55:18 So, this is a case in which, due to damage to one of the precision-modulating nuclei or sets of nuclei [indicating "precision" oval, lower right], in this man's case the basal forebrain nuclei - due to the damage to the acetylcholine modulating mechanisms that I showed you earlier - he *has too much confidence* in his predictions, and he *does not up the gain on the error signals*.



His predictive model tells him he's in Johannesburg. I say to him, "No, you're in London. Look out the window." He looks out the window, he sees snow - clearly not something that could possibly be associated with Johannesburg. But he sticks with his prediction. He says, "No, I know I'm in Jo'burg. Just because you're eating pizza, it doesn't mean you're in Italy."

1:56:19 And likewise, in the case of these memory processes that I showed you in that session - Each time that he starts to *feel* the unpleasure of the mounting error signals, that things are so... Just think about it! He touches his head. He's *on the brink of being aware* that he's had a brain operation, that the brain operation has resulted in loss of memory. I hope you can see that!

He says, "We just need to order the specs for this module that's missing." What does the module do? It does memory, you know, so touching his head, speaking about operations. So he's on the brink of recognizing that things were not as he expected, that things are in fact quite different.

1:57:11 My point being that this evokes feelings! - that this is bad! This is a *panic*-inducing situation. And because this man has damage to these precision-modulation mechanisms that we've been talking about, what he does is he simply up-regulates, or maintains his confidence in, his prediction; and in this way, maintains his emotional equanimity, rather than allowing the unpleasant affect to dominate and the unpleasant affect to update, to up-regulated the error signal which would which would normally update the predictive model.

So I'm hoping that in this case, you can see something of the role of affect, the role that the feelings involved in this man's case in relation to his predictions and the prediction errors. In other words, the cognitive business of what he perceives; and what he believes; and how what he perceives changes what he believes; and the role of feeling in all of this. I thought this was a succinct case to be able to illustrate all of that.

1:58:31 Because it's only one case, I thought I should just show you some of this. These are several papers that we've written on showing how these mechanisms work in confabulation.

Fotopoulou, A., Solms, M. & Turnbull, O. (2004) Wishful Reality Distortions in Confabulation: A Case Study. *Neuropsychologia*, 42, 6: 727-744.

Turnbull, O., Berry, H. & Evans, C. (2004) A positive emotional bias in confabulatory false beliefs about place. *Brain & Cognition*, 55: 490-494.

Fotopoulou, A. Conway, M. & Solms, M. (2007) Confabulation: Motivated Reality Monitoring. *Neuropsychologia*, 45, 2180-90.

Fotopoulou, A. Conway, M., Birchall, D., Griffiths, P. & Tyrer, S. (2007) Confabulation: Revising the Motivational Hypothesis. *Neurocase*, 13, 6-15.

Fotopoulou, A. Conway, M., Solms, M., Birchall, D. & Tyrer, S. (2008) Positive Emotional Biases in Confabulation: An Experimental Study. *Cortex*, 44, 764-772.

Fotopoulou, A. Conway, M., Solms, M., Kopelman, M. & Tyrer, S. (2008) Self-serving Confabulation in Prose Recall. *Neuropsychologia* 46, 1429-1441.

Fotopoulou, A. (2010) The Affective Neuropsychology of Confabulation and Delusion. M. Turner, M. Coltheart & R. Langdon (Eds.) Special Issue on 'Confabulation and Delusion', *Cognitive Neuropsychiatry*, 15, 1-13.

Besharati, S., Fotopoulou, A. & Kopelman, M. (2014), What is it like to be confabulating? In A. Mishara, A. Kranjec, P. Corlett, P. Fletcher & M. Schwartz, (eds.), *Phenomenological Neuropsychiatry, How Patient Experience Bridges Clinical with Clinical Neuroscience*. New York: Springer

We were able to show that affect regulation, that confabulation has a *wishful* quality. In other words, it has a *down-regulating-of-error-signals* quality and an *up-regulating-of-predictions* quality. And we were also able to show that - by analysis of transcripts of cases like this (we studied many cases like this) - how the affect actually improves with each confabulation.

So there's an increasing negative affect, followed by confabulation, followed by an improvement in the affect.

Session Wrapup

Claim 9 (preview): The predictive hierarchy is arranged concentrically, for progressive tolerance of uncertainty

1:59:34 (This is not such a fundamental point. It's just a slightly different way that I see the predictive hierarchy.)

If we take seriously that what we are talking about here - fundamentally, we are talking about homeostatic systems; and that the most important predictions have to do with maintaining your phenotypic prior preference distribution - that this has implications for how we conceptualize the predictive hierarchy.

But we don't have time to go into that.



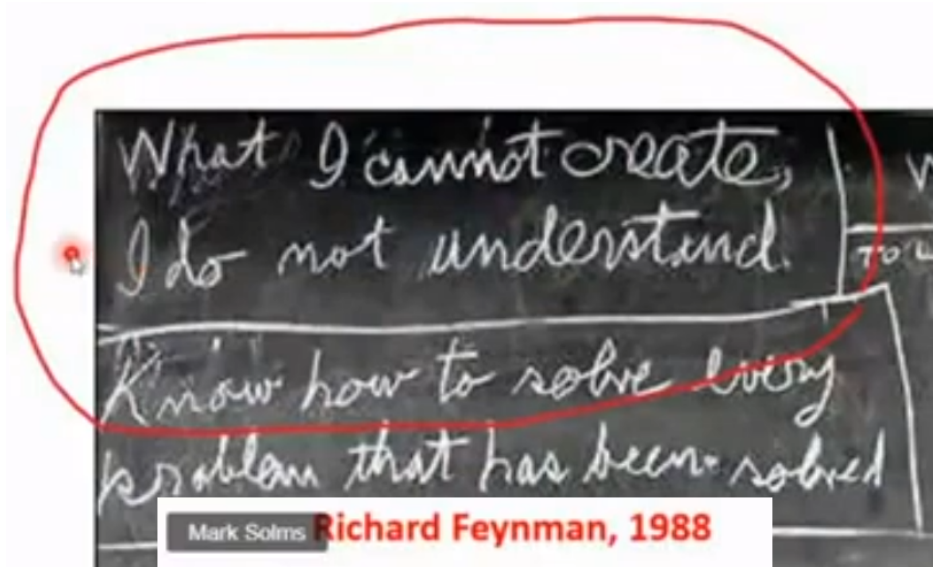
Claim 10 (preview): Consciousness can be artificially engineered

2:00:10 And then the last thing, is just to say that these models that I've, together with my colleagues, derived from the study of the neuroscientific evidence...

on the basis of this well-known statement of Richard Feynman's -

"What I cannot create, I do not understand" -

we are trying to instantiate these mechanisms that I described to you earlier in an **artificial consciousness**.



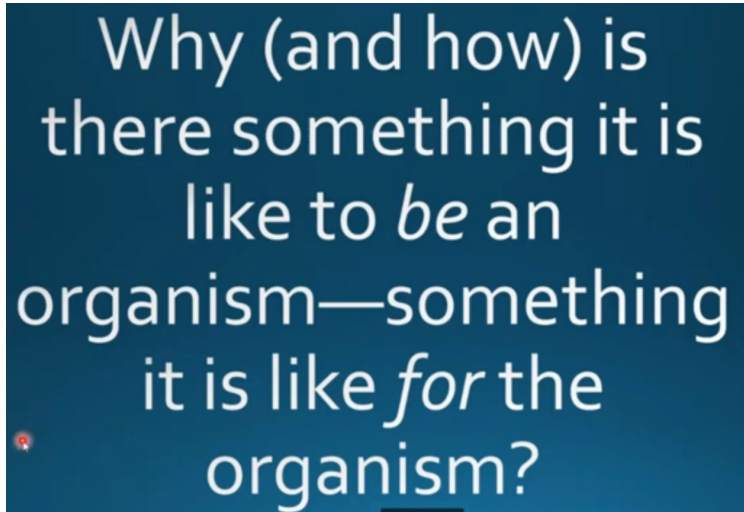
I'm working with a group of really great guys, physicists and computer scientists and applied mathematicians. And we presented our preliminary findings a few months ago at Karl Friston's Theoretical Neurobiology meetings. And so watch this space!

We are saying that, *if* we have identified the mechanism - the *causal* mechanism - whereby affects are generated, what they actually *do* in relation to a self-organizing system, *then* if this really *is* how affect is generated, if this really is how it is caused, *then* we should be able to instantiate it in an artificial system. And that's what we are trying to do at the moment.

Looking forward

2:01:51 These questions can look quite different if we look at it through the lens that I've tried to encourage you to look at it with me through - in other words, through the lens of affect.

These questions, "Why and how is it something it's like to *be* an organism, something it's like *for* the organism?"



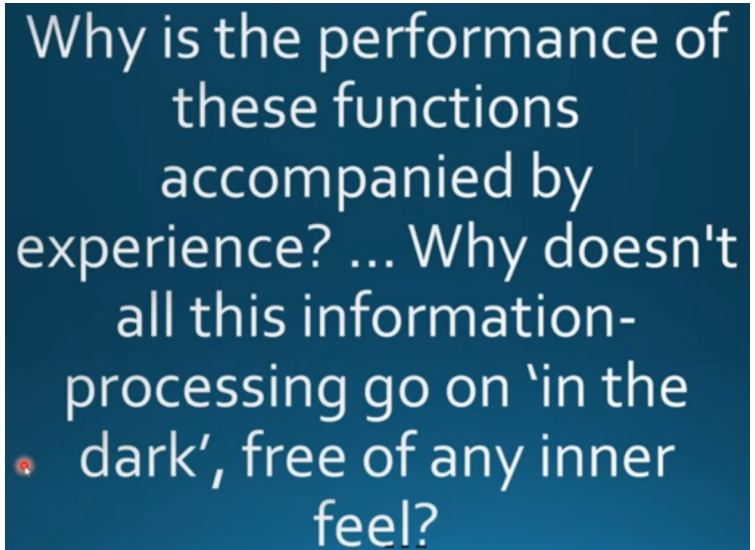
Why (and how) is there something it is like to *be* an organism—something it is like *for* the organism?

Mark Solms

... And questions like "Why is the performance of these functions accompanied by experience?"

This is Chalmers's question - "Why doesn't all this information processing go on 'in the dark,' free of any inner feel?"

I'm saying that these questions - I've tried to provide you *a clue* as to how we might be able to address them differently, but get them through this lens.



Why is the performance of these functions accompanied by experience? ... Why doesn't all this information-processing go on 'in the dark', free of any inner feel?

(I can't believe I've spoken for 2 hours again!)

But if you will allow me, I'll be happy to wait for any further comments and questions before we end.



2:03:02 *Friedman:*

As we have actually a whole document of questions, and many questions in the chat, I think it's a perfect place to end the live session. And whenever the affordance presents itself, we're always happy to have a Part Three. (The numbers keep counting!) And we'll keep hosting the sessions, because this was very fascinating! And I hope people got a lot out of it - and very tantalizing!

2:03:29 *Solms:*

Thank you so much for inviting me! I really enjoyed what little interchange we had over these four hours. Thank you so much!

2:03:42 *(Many):*

Thank you, Mark!

2:04:09 *Solms:*

Thanks, Stephen. Thanks, Dave. Thanks, Daniel!

Supplemental Lists

Subjects

Active Inference, affective demand, air hunger, allostasis, area prostroma, artificial consciousness, awareness, basal forebrain nuclei, basal ganglia, blank wakefulness, brainstem-broadly-defined, categorical variable, category, cerebral cortex, cerebral palsy, circumventricular organs, coma, concentric predictive hierarchy, confabulatory amnesia, confidence, consciousness-generating tissue, controlled hallucination, corticocentric, deep brain stimulation, demand for work, diencephalon, disinhibition, dopamine, down-regulation, drive, end-organ, entropy, epistemic gap, excess of confidence, feeling, feeling one's way, feeling self, Free Energy, frontal lobe personality, frontal lobe syndrome, Global Workspace Theory, Hard Problem, homeostasis, hormesis, hydranencephaly, hypervigilance, hypothalamus, insular cortex, intentionality, interoception, lateral geniculate, level of consciousness, locked-in syndrome, locus coeruleus, medulla oblongata, meningioma, message passing, meta-homeostasis, meta-homeostat, midbrain, midbrain locomotor region, midbrain selection triangle, natural selection, need state, neural correlates of consciousness, non-responsive wakefulness, noradrenaline, nucleus solitarius, oblongata, olfactory sheath meningioma, Own Body, palpating uncertainty, parabrachial complex, periaqueductal gray, peripersonal space, persistent vegetative state, phenomenal contents of consciousness, phenotype, philosophical zombie, pons, positron emission tomography, postsynaptic gain, precision, precision modulation, prefrontal cortex, psychiatry, psychodynamic understanding, qualia, rafe nucleus, render conscious, resistance, reticular activating arousal, reticular activating system, sentience, sentient self, serotonin, set point, settling point, somatic sensation, something-it-is-like, substantia nigra, suffocation alarm, superior colliculus, unpleasure, up-regulation, value system, vegetative state, ventral tegmental area, ventral tegmentum, wakefulness, wishful quality

References

Blomstedt, P; A Fytagoridis; S Tisch (2008) Deep brain stimulation of the posterior subthalamic area in the treatment of tremor, *Acta Neurochir*, DOI: [10.1007/s00701-008-0163-7](https://doi.org/10.1007/s00701-008-0163-7)

Felleman & van Essen (1992)

Jackson, Frank (1982) "Epiphenomenal Qualia". *Philosophical Quarterly*. 32 (127): 127–136. doi:10.2307/2960077. JSTOR 2960077. <https://www.jstor.org/stable/2960077>

Kihlstrom, J., 1996 Perception without awareness of what is perceived, learning without awareness of what is learned. In M. Velmans (ed) *The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews*. London, Routledge, pp. 23-46. <https://www.ocf.berkeley.edu/~jfkihlstrom/Velmans96.htm>

Merker, Björn; Kenneth Williford; David Rudrauf (2021) The integrated information theory of consciousness: A case of mistaken identity. DOI:

<https://doi.org/10.1017/S0140525X21000881>,
<https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/integrated-information-theory-of-consciousness-a-case-of-mistaken-identity/54A2FF12685631EC90DFC4AA24639FB>

Moruzzi, G; Magoun, H.W. (1949) "Brain Stem Reticular Formation and Activation of the EEG." *Electroencephalography and Clinical Neurophysiology*, 1, 455-473.
[http://dx.doi.org/10.1016/0013-4694\(49\)90219-9](http://dx.doi.org/10.1016/0013-4694(49)90219-9)

Solms M (2019) The Hard Problem of Consciousness and the Free Energy Principle. *Front. Psychol.* 9:2714. doi: 10.3389/fpsyg.2018.02714.
<https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02714/full>

Solms, Mark S; Friston, K. (2018). How and why consciousness arises: some considerations from physics and physiology. *J. Conscious. Stud.* 25, 202–238.
<https://discovery.ucl.ac.uk/id/eprint/10057681/>

Mentioned Names & People

David Chalmers (b. 1966)
AD (Bud) Craig, Jr. (b. 1951)
Antonio Damasio (b. 1944)
Frank C. Jackson (b. 1943)
Joseph E. LeDoux (b. 1949)
Jaak Panksepp (1943-1917)
Björn Merker (b. 1943)
Horace W. Magoun (1907-1991)
Giuseppe Moruzzi (1910-1986)
Thomas Nagel (b. 1937)
Anil Seth (b. 1972)
Sylvan S. Tomkins (1911-91)

Other Resources

Damásio, António (1994) *Descartes' Error*, Penguin Books, 978-0143036227.

Freud (1915) "The instincts and their vicissitudes"

Jackson, Frank C. (1986) "What Mary Didn't Know," *The Journal of Philosophy*, Vol 83 No. 5,
https://courses.physics.illinois.edu/phys419/sp2021/Jackson1986_WhatMaryDidntKnow.pdf

Mark P. Mattson (2009) *Hormesis Defined*.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2248601/>

Nagel, Thomas (1974) What Is It Like to Be a Bat? The Philosophical Review, Vol. 83, No. 4 (Oct., 1974), pp. 435-450.

https://warwick.ac.uk/fac/cross_fac/iatl/study/ugmodules/humananimalstudies/lectures/32/nagel_bat.pdf

Panksepp, Jaak (2004) Affective Neuroscience, Oxford University Press, ISBN-13: 978-0195178050.

Solms, Mark (2006) Freud Returns. <https://www.scientificamerican.com/article/freud-returns/>

-- The Conscious Id (2013) Neuropsychanalysis,

DOI:[10.1080/15294145.2013.10773711](https://doi.org/10.1080/15294145.2013.10773711),

<https://www.fil.ion.ucl.ac.uk/~karl/The%20Conscious%20Id.pdf>

Sugarman, Susan (2016) What Freud Really Meant (Drives and their Destinies), Cambridge University Press, <https://doi.org/10.1017/CBO9781316340240>.