

Book Review: Surfing Uncertainty

Posted on September 5, 2017 by Scott Alexander



Related to: [It's Bayes All The Way Up](#), [Why Are Transgender People Immune To Optical Illusions?](#), [Can We Link Perception And Cognition?](#)

I

Sometimes I have the fantasy of being able to glut myself on Knowledge. I imagine meeting a time traveler from 2500, who takes pity on me and gives me a book from the future where all my questions have been answered, one after another. What's consciousness? That's in Chapter 5. How did something arise out of nothing? Chapter 7. It all makes perfect intuitive sense and is fully vouched by unimpeachable authorities. I assume something like this is how everyone spends their first couple of days in Heaven, whatever it is they do for the rest of Eternity.

And every so often, my fantasy comes true. Not by time travel or divine intervention, but by failing so badly at paying attention to the literature that by the time I realize people are working on a problem it's already been investigated, experimented upon, organized into a

paradigm, tested, and then placed in a nice package and wrapped up with a pretty pink bow so I can enjoy it all at once.

The predictive processing model is one of these well-wrapped packages. Unbeknownst to me, over the past decade or so neuroscientists have come up with a real *theory* of how the brain works – a real unifying framework theory like Darwin’s or Einstein’s – and it’s beautiful and it makes complete sense.

[Surfing Uncertainty](#) isn’t pop science and isn’t easy reading. Sometimes it’s on the border of possible-at-all reading. Author Andy Clark (a professor of logic and metaphysics, of all things!) is clearly brilliant, but prone to going on long digressions about various esoteric philosophy-of-cognitive-science debates. In particular, he’s obsessed with showing how “embodied” everything is all the time. This gets kind of awkward, since the predictive processing model isn’t really a natural match for embodiment theory, and describes a brain which is pretty embodied in some ways but not-so-embodied in others. If you want a hundred pages of apologia along the lines of “this may not *look* embodied, but if you squint you’ll see how super-duper embodied it really is!”, this is your book.

It’s also your book if you want to learn about predictive processing at all, since as far as I know this is the only existing book-length treatment of the subject. And it’s comprehensive, scholarly, and very good at giving a good introduction to the theory and why it’s so important. So let’s be grateful for what we’ve got and take a look.

II

Stanislas Dehaene writes of our senses:

We never see the world as our retina sees it. In fact, it would be a pretty horrible sight: a highly distorted set of light and dark pixels, blown up toward the center of the retina, masked by blood vessels, with a massive hole at the location of the “blind spot” where cables leave for the brain; the image would constantly blur and change as our gaze moved around. What we see, instead, is a three-dimensional scene, corrected for retinal defects, mended at the blind spot, stabilized for our eye and head movements, and massively reinterpreted based on our previous experience of similar visual scenes. All these operations unfold unconsciously—although many of them are so complicated that they resist computer modeling. For instance, our visual system detects the presence of shadows in the image and removes them. At a glance, our brain unconsciously infers the sources of lights and deduces the shape, opacity, reflectance, and luminance of the objects.

Predictive processing begins by asking: how does this happen? By what process do our incomprehensible sense-data get turned into a meaningful picture of the world?

The key insight: the brain is a multi-layer prediction machine. All neural processing consists of two streams: a bottom-up stream of

sense data, and a top-down stream of predictions. These streams interface at each level of processing, comparing themselves to each other and adjusting themselves as necessary.

The bottom-up stream starts out as all that incomprehensible light and darkness and noise that we need to process. It gradually moves up all the cognitive layers that we already knew existed – the edge-detectors that resolve it into edges, the object-detectors that shape the edges into solid objects, et cetera.

The top-down stream starts with everything you know about the world, all your best heuristics, all your priors, everything that's ever happened to you before – everything from “solid objects can't pass through one another” to “ $e = mc^2$ ” to “that guy in the blue uniform is probably a policeman”. It uses its knowledge of concepts to make predictions – not in the form of verbal statements, but in the form of expected sense data. It makes some guesses about what you're going to see, hear, and feel next, and asks “Like this?” These predictions gradually move *down* all the cognitive layers to generate lower-level predictions. If that uniformed guy was a policeman, how would that affect the various objects in the scene? Given the answer to that question, how would it affect the distribution of edges in the scene? Given the answer to *that* question, how would it affect the raw-sense data received?

Both streams are probabilistic in nature. The bottom-up sensory stream has to deal with fog, static, darkness, and neural noise; it knows that whatever forms it tries to extract from this signal might or might not be real. For its part, the top-down predictive stream

knows that predicting the future is inherently difficult and its models are often flawed. So both streams contain not only data but estimates of the precision of that data. A bottom-up percept of an elephant right in front of you on a clear day might be labelled “very high precision”; one of a a vague form in a swirling mist far away might be labelled “very low precision”. A top-down prediction that water will be wet might be labelled “very high precision”; one that the stock market will go up might be labelled “very low precision”.

As these two streams move through the brain side-by-side, they continually interface with each other. Each level receives the predictions from the level above it and the sense data from the level below it. Then each level uses [Bayes' Theorem](#) to integrate these two sources of probabilistic evidence as best it can. This can end up a couple of different ways.

First, the sense data and predictions may more-or-less match. In this case, the layer stays quiet, indicating “all is well”, and the higher layers never even hear about it. The higher levels just keep predicting whatever they were predicting before.

Second, low-precision sense data might contradict high-precision predictions. The Bayesian math will conclude that the predictions are still probably right, but the sense data are wrong. The lower levels will “cook the books” – rewrite the sense data to make it look as predicted – and then continue to be quiet and signal that all is well. The higher levels continue to stick to their predictions.

Third, there might be some unresolvable conflict between high-precision sense-data and predictions. The Bayesian math will indicate that the predictions are probably wrong. The neurons involved will fire, indicating “surprisal” – a gratuitously-technical neuroscience term for surprise. The higher the degree of mismatch, and the higher the supposed precision of the data that led to the mismatch, the more surprisal – and the louder the alarm sent to the higher levels.

When the higher levels receive the alarms from the lower levels, *this is their equivalent of bottom-up sense-data*. They ask themselves: “Did the even-higher-levels predict this would happen?” If so, they themselves stay quiet. If not, they might try to change their own models that map higher-level predictions to lower-level sense data. Or they might try to cook the books themselves to smooth over the discrepancy. If none of this works, they send alarms to the even-higher-levels.

All the levels really hate hearing alarms. Their goal is to *minimize surprisal* – to become so good at predicting the world (conditional on the predictions sent by higher levels) that nothing ever surprises them. Surprise prompts a frenzy of activity adjusting the parameters of models – or deploying new models – until the surprise stops.

All of this happens several times a second. The lower levels constantly shoot sense data at the upper levels, which constantly adjust their hypotheses and shoot them down at the lower levels. When surprise is registered, the relevant levels change their hy-

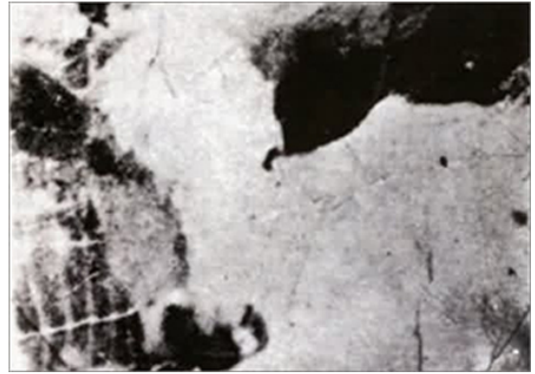
potheses or pass the buck upwards. After umpteen zillion cycles, everyone has the right hypotheses, nobody is surprised by anything, and the brain rests and moves on to the next task. As per the book:

To deal rapidly and fluently with an uncertain and noisy world, brains like ours have become masters of prediction – surfing the waves and noisy and ambiguous sensory stimulation by, in effect, trying to stay just ahead of them. A skilled surfer stays ‘in the pocket’: close to, yet just ahead of the place where the wave is breaking. This provides power and, when the wave breaks, it does not catch her. The brain’s task is not dissimilar. By constantly attempting to predict the incoming sensory signal we become able – in ways we shall soon explore in detail – to learn about the world around us and to engage that world in thought and action.

The result is perception, which the PP theory describes as “controlled hallucination”. You’re not seeing the world as it is, exactly. You’re seeing your predictions about the world, cashed out as expected sensations, then shaped/constrained by the actual sense data.

III

Enough talk. Let’s give some examples. Most of you have probably seen these before, but it never hurts to remind:



This demonstrates the degree to which the brain depends on top-down hypotheses to make sense of the bottom-up data. To most people, these two pictures start off looking like incoherent blotches of light and darkness. Once they figure out what they are ([spoiler](#)) the scene becomes obvious and coherent. According to the predictive processing model, this is how we perceive everything all the time – except usually the concepts necessary to make the scene fit together come from our higher-level predictions instead of from clicking on a spoiler link.



This demonstrates how the top-down stream's efforts to shape the bottom-up stream and make it more coherent can sometimes "cook the books" and alter sensation entirely. The real picture says "PARIS IN THE THE SPRINGTIME" (note the duplicated word "the"!). The top-down stream predicts this should be a meaningful sentence that obeys English grammar, and so replaces the the bottom-up stream with what it thinks that it *should* have said. This is a very powerful process – how many times have I repeated the the word "the" in this paragraph alone without you noticing?

You can porabbly raed tihs ptetry wlel eevn
tohguh it's all jubemld up.

A more ambiguous example of "perception as controlled hallucination". Here your experience doesn't quite *deny* the jumbled-up nature of the letters, but it superimposes a "better" and more coherent experience which appears naturally alongside.

IV

Okay. You've read a lot of words. You've looked at a lot of pictures. You've listened to "Never Gonna Give You Up" for ten hours. Time for the payoff. Let's use this theory to explain everything.

1. Attention

In PP, attention measures "the confidence interval of your predictions". Sense-data within the confidence intervals counts as a match and doesn't register surprisal. Sense-data outside the confidence intervals fails and alerts higher levels and eventually consciousness.

This modulates the balance between the top-down and bottom-up streams. High attention means that perception is mostly based on the bottom-up stream, since every little deviation is registering an error and so the overall perceptual picture is highly constrained by sensation. Low attention means that perception is mostly based on the top-down stream, and you're perceiving only a vague outline of the sensory image with your predictions filling in the rest.

There's a famous experiment which you can try below – if you're trying it, make sure to play the whole video before moving on:

see the gorilla immediately. Your confidence intervals for unusual things are razor-thin; as soon as that neuron sees the gorilla it sends alarms to higher levels, and the higher levels quickly come up with a suitable hypothesis (“there’s a guy in a gorilla suit here”) which makes sense of the new data.

There’s an interesting analogy to vision here, where the center of your vision is very clear, and the outsides are filled in in a top-down way – I have a vague sense that my water bottle is in the periphery right now, but only because I kind of already know that, and it’s more of a mental note of “water bottle here as long as you ask no further questions” than a clear image of it. The extreme version of this is [the blind spot](#), which gets filled in entirely with predicted imagery despite receiving no sensation at all.

2. Imagination, Simulation, Dreaming, Etc.

Imagine a house. Now imagine a meteor crashing into the house. Your internal mental simulation was probably pretty good. Without even thinking about it, you got it to obey accurate physical laws like “the meteor continues on a constant trajectory”, “the impact happens in a realistic way”, “the impact shatters the meteorite”, and “the meteorite doesn’t bounce back up to space like a basketball”. Think how surprising this is.

In fact, think how surprising it is that you can imagine the house at all. This really high level concept – “house” – has been transformed in your visual imaginarium into a pretty good picture of a

house, complete with various features, edges, colors, et cetera (if it hasn't, read [here](#)). This is near-miraculous. Why do our brains have this apparently useless talent?

PP says that the highest levels of our brain make predictions *in the form of sense data*. They're not just saying "I predict that guy over there is a policeman", they're generating the image of a policeman, cashing it out in terms of sense data, and colliding it against the sensory stream to see how it fits. The sensory stream gradually modulates it to fit the bottom-up evidence – a white or black policeman, a mustached or clean-shaven policeman. But the top-down stream is doing a lot of the work here. We are able to imagine the meteor, using the same machinery that would guide our perception of the meteor if we saw it up in the sky.

All of this goes double for dreaming. If "perception is controlled hallucination" caused by the top-down drivers of perception constrained by bottom-up evidence, then dreams are those top-down drivers playing around with themselves unconstrained by anything at all (or else very weakly constrained by bottom-up evidence, like when it's really cold in your bedroom and you dream you're exploring the North Pole).

A lot of people claim higher levels of this – lucid dreaming, astral projection, you name it, worlds exactly as convincing as our own but entirely imaginary. Predictive processing is very sympathetic to these accounts. The generative models that create predictions are really good; they can simulate the world well enough that it rarely surprises us. They also connect through various layers to our bot-

tom-level perceptual apparatus, cashing out their predictions in terms of the lowest-level sensory signals. Given that we've got a top-notch world-simulator plus perception-generator in our heads, it shouldn't be surprising when we occasionally perceive ourselves in simulated worlds.

3. Priming

I don't mean the weird made-up kinds of priming that don't replicate. I mean the very firmly established ones, like the one where, if you flash the word "DOCTOR" at a subject, they'll be much faster and more skillful in decoding a series of jumbled and blurred letters into the word "NURSE".

This is classic predictive processing. The top-down stream's whole job is to assist the bottom-up stream in making sense of complicated fuzzy sensory data. After it hears the word "DOCTOR", the top-down stream is already thinking "Okay, so we're talking about health care professionals". This creeps through all the lower levels as a prior for health-care related things; when the sense organs receive data that can be associated in a health-care related manner, the high prior helps increase the precision of this possibility until it immediately becomes the overwhelming leading hypothesis.

4. Learning

There's a philosophical debate – which I'm not too familiar with, so sorry if I get it wrong – about how "unsupervised learning" is possi-

ble. Supervised reinforcement learning is when an agent tries various stuff, and then someone tells the agent if it's right or wrong. Unsupervised learning is when nobody's around to tell you, and it's what humans do all the time.

PP offers a compelling explanation: we create models that generate sense data, and keep those models if the generated sense data match observation. Models that predict sense data well stick around; models that fail to predict the sense data accurately get thrown out. Because of all those lower layers adjusting out contingent features of the sensory stream, any given model is left with exactly the sense data necessary to tell it whether it's right or wrong.

PP isn't *exactly* blank slate, but it's compatible with a slate that's pretty fricking blank. Clark discusses "hyperpriors" – extremely basic assumptions about the world that we probably need to make sense of anything at all. For example, one hyperprior is sensory synchronicity – the idea that our five different senses are describing the same world, and that the stereo we see might be the source of the music we hear. Another hyperprior is object permanence – the idea that the world is divided into specific objects that stick around whether or not they're in the sensory field. Clark says that some hyperpriors *might* be innate – but says they don't have to be, since PP is strong enough to learn them on its own if it has to. For example, after enough examples of, say, seeing a stereo being smashed with a hammer at the same time that music suddenly stops, the brain can infer that connecting the visual and au-



Next up – this low-quality video of an airplane flying at night. Notice how after an instant, you start to predict the movement and characteristics of the airplane, so that you’re no longer surprised by the blinking light, the movement, the other blinking light, the camera shakiness, or anything like that – in fact, if the light *stopped* blinking, you would be surprised, even though naively nothing could be less surprising than a dark portion of the night sky staying dark. After a few seconds of this, the airplane continuing on its (pretty complicated) way just reads as “same old, same old”. Then when something else happens – like the camera panning out, or the airplane making a slight change in trajectory – you focus entirely on that, the blinking lights and movement entirely forgotten or at least packed up into “airplane continues on its blinky way”. Meanwhile, other things – like the feeling of your shirt against your skin – have been completely predicted away and blocked from consciousness, freeing you to concentrate entirely on any subtle changes in the airplane’s motion.

ditory evidence together is a useful hack that helps it to predict the sensory stream.

I can't help thinking here of [Molyneux's Problem](#), a thought experiment about a blind-from-birth person who navigates the world through touch alone. If suddenly given sight, could the blind person naturally connect the visual appearance of a cube to her own concept "cube", which she derived from the way cubes feel? In 2003, some researchers took advantage of a new cutting-edge blindness treatment to [test this out](#) ; they found that no, the link isn't intuitively obvious to them. Score one for learned hyperpriors.

But learning goes all the way from these kinds of really basic hyperpriors all the way up to normal learning like what the capital of France is – which, if nothing else, helps predict what's going to be on the other side of your geography flashcard, and which high-level systems might keep as a useful concept to help it make sense of the world and predict events.

5. Motor Behavior

About a third of *Surfing Uncertainty* is on the motor system, it mostly didn't seem that interesting to me, and I don't have time to do it justice here (I might make another post on one especially interesting point). But this has been kind of ignored so far. If the brain is mostly just in the business of making predictions, what exactly is the motor system doing?

Based on a bunch of really excellent experiments that I don't have time to describe here, Clark concludes: it's predicting action, which causes the action to happen.

This part is almost funny. Remember, the brain really hates prediction error and does its best to minimize it. With failed predictions about eg vision, there's not much you can do except change your models and try to predict better next time. But with predictions about proprioceptive sense data (ie your sense of where your joints are), there's an easy way to resolve prediction error: just move your joints so they match the prediction. So (and I'm asserting this, but see Chapters 4 and 5 of the book to hear the scientific case for this position) if you want to lift your arm, your brain just predicts *really really strongly* that your arm has been lifted, and then lets the lower levels' drive to minimize prediction error do the rest.

Under this model, the “prediction” of a movement isn't just the idle thought that a movement might occur, it's *the actual motor program*. This gets unpacked at all the various layers – joint sense, proprioception, the exact tension level of various muscles – and finally ends up in a particular fluid movement:

Friston and colleagues... suggest that precise proprioceptive predictions directly elicit motor actions. This means that motor commands have been replaced by (or as I would rather say, implemented by) proprioceptive predictions. According to active inference, the agent moves body and sensors in ways that amount to actively seeking out the sensory consequences that their brains expect. Perception, cognition, and

action – if this unifying perspective proves correct – work together to minimize sensory prediction errors by selectively sampling and actively sculpting the stimulus array. This erases any fundamental computational line between perception and the control of action. There remains [only] an obvious difference in direction of fit. Perception here matches hural hypotheses to sensory inputs... while action brings unfolding proprioceptive inputs into line with neural predictions. The difference, as Anscombe famously remarked, is akin to that between consulting a shopping list (thus letting the list determine the contents of the shopping basket) and listing some actually purchased items (thus letting the contents of the shopping basket determine the list). But despite the difference in direction of fit, the underlying form of the neural computations is now revealed as the same.

6. Tickling Yourself

One consequence of the PP model is that organisms are continually adjusting out their own actions. For example, if you're trying to predict the movement of an antelope you're chasing across the visual field, you need to adjust out the up-down motion of your own running. So one "hyperprior" that the body probably learns pretty early is that if it itself makes a motion, it should expect to feel the consequences of that motion.

There's a really interesting illusion called the force-matching task. A researcher exerts some force against a subject, then asks the subject to exert exactly that much force against something else.

Subjects' forces are usually biased upwards – they exert more force than they were supposed to – probably because their brain's prediction engines are “cancelling out” their own force. Clark describes one interesting implication:

The same pair of mechanisms (forward-model-based prediction and the dampening of resulting well-predicted sensation) have been invoked to explain the unsettling phenomenon of ‘force escalation’. In force escalation, physical exchanges (playground fights being the most common exemplar) mutually ramp up via a kind of step-ladder effect in which each person believes the other one hit them harder. Shergill et al describe experiments that suggest that in such cases each person is truthfully reporting their own sensations, but that those sensations are skewed by the attenuating effects of self-prediction. Thus, ‘self-generated forces are perceived as weaker than externally generated forces of equal magnitude.’

This also explains why you can't tickle yourself – your body predicts and adjusts away your own actions, leaving only an attenuated version.

7. The Placebo Effect

We hear a lot about “pain gating” in the spine, but the PP model does a good job of explaining what this is: adjusting pain based on top-down priors. If you believe you should be in pain, the brain will use that as a filter to interpret ambiguous low-precision pain sig-

nals. If you believe you shouldn't, the brain will be more likely to assume ambiguous low-precision pain signals are a mistake. So if you take a pill that doctors assure you will cure your pain, then your lower layers are more likely to interpret pain signals as noise, “cook the books” and prevent them from reaching your consciousness.

Psychosomatic pain is the opposite of this; see Section 7.10 of the book for a fuller explanation.

8. Asch Conformity Experiment

More speculative, and not from the book. But remember this one? A psychologist asked subjects which lines were the same length as other lines. The lines were all *kind of* similar lengths, but most subjects were still able to get the right answer. Then he put the subjects in a group with confederates; all of the confederates gave the same wrong answer. When the subject's turn came, usually they would disbelieve their eyes and give the same wrong answer as the confederates.

The bottom-up stream provided some ambiguous low-precision bottom-up evidence pointing toward one line. But in the final Bayesian computation, those were swamped by the strong top-down prediction that it would be another. So the middle layers “cooked the books” and replaced the perceived sensation with the predicted one. From Wikipedia:

Participants who conformed to the majority on at least 50% of trials reported reacting with what Asch called a “distortion of perception”. These participants, who made up a distinct minority (only 12 subjects), expressed the belief that the confederates’ answers were correct, and were apparently unaware that the majority were giving incorrect answers.

9. Neurochemistry

PP offers a way to a psychopharmacological holy grail – an explanation of what different neurotransmitters really *mean*, on a human-comprehensible level. Previous attempts to do this, like “dopamine represents reward, serotonin represents calmness”, have been so wildly inadequate that the whole question seems kind of disreputable these days.

But as per PP, the NMDA glutamatergic system mostly carries the top-down stream, the AMPA glutamatergic system mostly carries the bottom-up stream, and dopamine mostly carries something related to precision, confidence intervals, and surprisal levels. This matches a lot of observational data in a weirdly consistent way – for example, it doesn’t take a lot of imagination to think of the slow, hesitant movements of Parkinson’s disease as having “low motor confidence”.

10. Autism

Various research in the PP tradition has coalesced around the idea of autism as an unusually high reliance on bottom-up rather than top-down information, leading to “weak central coherence” and constant surprisal as the sensory data fails to fall within pathologically narrow confidence intervals.

Autistic people classically [can't stand tags on clothing](#) – they find them too scratchy and annoying. Remember the example from Part III about how you successfully predicted away the feeling of the shirt on your back, and so manage never to think about it when you're trying to concentrate on more important things? Autistic people can't do that as well. Even though they have a layer in their brain predicting “will continue to feel shirt”, the prediction is too precise; it predicts that next second, the shirt will produce *exactly* the same pattern of sensations it does now. But realistically as you move around or catch passing breezes the shirt will change ever so slightly – at which point autistic people's brains will send alarms all the way up to consciousness, and they'll perceive it as “my shirt is annoying”.

Or consider the classic autistic demand for routine, and misery as soon as the routine is disrupted. Because their brains can only make very precise predictions, the slightest disruption to routine registers as strong surprisal, strong prediction failure, and “oh no, all of my models have failed, nothing is true, anything is possible!” Compare to a neurotypical person in the same situation, who would just relax their confidence intervals a little bit and say “Okay, this is basically 99% like a normal day, whatever”. It would take something genuinely unpredictable – like being thrown on an unex-



In the same vein: this is Rick Astley's "Never Going To Give You Up" repeated again and again for ten hours (you can find some *weird* stuff on YouTube). The first hour, maybe you find yourself humming along occasionally. By the second hour, maybe it's gotten kind of annoying. By the third hour, you've completely forgotten it's even on at all.

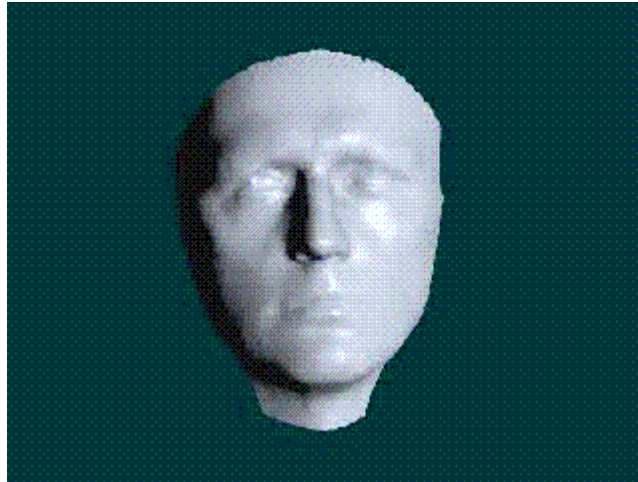
But suppose that one time, somewhere around the sixth hour, it skipped two notes – just the two syllables "never", so that Rick said "Gonna give you up." Wouldn't the silence where those two syllables should be sound as jarring as if somebody set off a bomb right beside you? Your brain, having predicted sounds consistent with "Never Gonna Give You Up" going on forever, suddenly finds its expectations violated and sends all sorts of alarms to the higher levels, where they eventually reach your consciousness and make you go "What the *heck* ?"

plored continent or something – to give these people the same feeling of surprise and unpredictability.

This model also predicts autistic people’s strengths. We know that polygenic risk for autism is [positively associated with IQ](#). This would make sense if the central feature of autism was a sort of increased mental precision. It would also help explain why autistic people seem to excel in high-need-for-precision areas like mathematics and computer programming.

11. Schizophrenia

Converging lines of research suggest this also involves weak priors, apparently at a different level to autism and with different results after various compensatory mechanisms have had their chance to kick in. One especially interesting study asked neurotypicals and schizophrenics to follow a moving light, much like the airplane video in Part III above. When the light moved in a predictable pattern, the neurotypicals were much better at tracking it; when it was a deliberately perverse video specifically designed to frustrate expectations, the schizophrenics actually did better. This suggests that neurotypicals were guided by correct top-down priors about where the light would be going; schizophrenics had very weak priors and so weren’t really guided very well, but also didn’t screw up when the light did something unpredictable. Schizophrenics are also famous for not being fooled by the “hollow mask” (below) and other illusions where top-down predictions falsely constrain bottom-up evidence. My guess is they’d be more likely to see both ‘the’s in the “PARIS IN THE THE SPRINGTIME” image above.



The exact route from this sort of thing to schizophrenia is really complicated, and anyone interested should check out Section 2.12 and the whole of Chapter 7 from the book. But the basic story is that it creates waves of anomalous prediction error and surprisal, leading to the so-called “delusions of significance” where schizophrenics believe that eg the fact that someone is wearing a hat is some sort of incredibly important cosmic message. Schizophrenics’ brains try to produce hypotheses that explain all of these prediction errors and reduce surprise – which is impossible, because the prediction errors are random. This results in incredibly weird hypotheses, and eventually in schizophrenic brains being willing to ignore the bottom-up stream entirely – hence hallucinations.

All this is treated with antipsychotics, which antagonize dopamine, which – remember – represents confidence level. So basically the medication is telling the brain “YOU CAN IGNORE ALL THIS PREDICTION ERROR, EVERYTHING YOU’RE PERCEIVING IS TOTALLY GARBAGE SPURIOUS DATA” – which turns out to be exactly the message it needs to hear.

An interesting corollary of all this – because all of schizophrenics’ predictive models are so screwy, they lose the ability to use the “adjust away the consequences of your own actions” hack discussed in Part 5 of this section. That means their own actions *don’t* get predicted out, and seem like the actions of a foreign agent. This is why they get so-called “delusions of agency”, like “the government beamed that thought into my brain” or “aliens caused my arm to move just now”. And in case you were wondering – [yes, schizophrenics can tickle themselves.](#)

12. Everything else

I can’t possibly do justice to the whole of *Surfing Uncertainty*, which includes sections in which it provides lucid and compelling PP-based explanations of hallucinations, binocular rivalry, conflict escalation, and various optical illusions. More speculatively, I can think of really interesting connections to things like phantom limbs, creativity (and its association with certain mental disorders), depression, meditation, etc, etc, etc.

The general rule in psychiatry is: if you think you’ve found a theory that explains everything, diagnose yourself with mania and check yourself into the hospital. Maybe I’m not at that point yet – for example, I don’t think PP does anything to explain what mania itself is. But I’m pretty close.

IV

This is a really poor book review of *Surfing Uncertainty*, because I only partly understood it. I'm leaving out a *lot* of stuff about the motor system, debate over philosophical concepts with names like “enactivism”, descriptions of how neurons form and unform coalitions, and of course a hundred pages of apologia along the lines of “this may not *look* embodied, but if you squint you'll see how super-duper embodied it really is!”. As I reread and hopefully come to understand some of this better, it might show up in future posts.

But speaking of philosophical debates, there's one thing that really struck me about the PP model. [Voodoo psychology](#) suggests that culture and expectation tyrannically shape our perceptions. Taken to an extreme, objective knowledge is impossible, since all our sense-data is filtered through our own bias. Taken to a *very far* extreme, we get things like [What The !@\\$ Do We Know?](#)'s claim that the Native Americans literally couldn't see Columbus' ships, because they had no concept of “caravel” and so the percept just failed to register. This sort of thing tends to end by arguing that science was invented by straight white men, and so probably just reflects straight white maleness, and so we should ignore it completely and go frolic in the forest or something.

Predictive processing is sympathetic to all this. It takes all of this stuff like priming and the placebo effect, and it predicts it handily. But it doesn't give up. It (theoretically) puts it all on a sound mathematical footing, explaining exactly how *much* our expectations should shape our reality, and in which ways our expectation should shape our reality. I feel like someone armed with predictive processing and a bit of luck should have been able to predict that

placebo effect and basic priming would work, but stereotype threat and social priming wouldn't. Maybe this is total retrodictive cheating. But I feel like it should be possible.

[illegible]

The rationalist project is overcoming bias, and that requires both an admission that bias is possible, and a hope that there's something *other* than bias which we can latch onto as a guide. Predictive processing gives us more confidence in both, and helps provide a convincing framework we can use to figure out what's going on at all levels of cognition.

selective attention test



...

...

About half of subjects, told to watch the players passing the ball, don't notice the gorilla. Their view of the ball-passing is closely constrained by the bottom-up stream; they see mostly what is there. But their view of the gorilla is mostly dependent on the top-down stream. Their confidence intervals are wide. Somewhere in your brain is a neuron saying "is that a guy in a gorilla suit?" Then it consults the top-down stream, which says "This is a basketball game, you moron", and it smooths out the anomalous perception into something that makes sense like another basketball player.

But if you watch the video with the prompt "Look for something strange happening in the midst of all this basketball-playing", you