

It's Bayes All The Way Up

Posted on September 12, 2016 by Scott Alexander



Epistemic status: *Very speculative. I am not a neuroscientist and apologize for any misinterpretation of the papers involved. Thanks to the people who posted these papers in [r/slatestarcodex](#). See also [Mysticism and Pattern-Matching](#) and [Bayes For Schizophrenics](#).*

I

Bayes' Theorem is an equation for calculating certain kinds of conditional probabilities. For something so obscure, it's attracted a surprisingly wide fanbase, including [doctors](#), [environmental scientists](#), [economists](#), [bodybuilders](#), [fen-dwellers](#), and [international smugglers](#). Eventually the hype reached the point where there was both a [Bayesian cabaret](#) and a [Bayesian choir](#), popular books using Bayes' Theorem to prove both the [existence](#) and the [nonexistence](#) of God, and even [Bayesian dating advice](#). Eventually everyone agreed to dial down their exuberance a little, and accept that Bayes' Theorem might not literally explain *absolutely* everything.

So – did you know that the neurotransmitters in the brain might represent different terms in Bayes' Theorem?

First things first: Bayes' Theorem is a mathematical framework for integrating new evidence with prior beliefs. For example, suppose you're sitting in your quiet suburban home and you hear something that sounds like a lion roaring. You have some prior beliefs that lions are unlikely to be near your house, so you figure that it's probably not a lion. Probably it's some weird machine of your neighbor's that just happens to sound like a lion, or some kids pranking you by playing lion noises, or something. You end up believing that there's probably no lion nearby, but you do have a slightly higher probability of there being a lion nearby than you had before you heard the roaring noise. Bayes' Theorem is just this kind of reasoning converted to math. You can find the long version [here](#).

This is what the brain does too: integrate new evidence with prior beliefs. Here are some examples I've used on this blog before:





All three of these are examples of top-down processing. Bottom-up processing is when you build perceptions into a model of the the world. Top-down processing is when you let your models of the world influence your perceptions. In the first image, you view the center letter of the the first word as an H and the second as an A, even though they're the the same character; your model of the world tells you that THE CAT is more likely than TAE CHT. In the second image, you read "PARIS IN THE SPRINGTIME", skimming

over the duplication of the word “the”; your model of the world tells you that the phrase should probably only have one “the” in it (just as you’ve probably skimmed over it the three times I’ve duplicated “the” in this paragraph alone!). The third image might look meaningless until you realize it’s a cow’s head; once you see the cow’s head your model of the world informs your perception and it’s almost impossible to see it as anything else.

(Teh fcat taht you can siltl raed wrods wth all the itroneir ltretrs rgranearnd is ahonter empvlae of top-dwon pssirocneg mkinag nsioy btotom-up dtaa sanp itno pacle)

But top-down processing is much more omnipresent than even these examples would suggest. Even something as simple as looking out the window and seeing a tree requires top-down processing; it may be too dark or foggy to see the tree one hundred percent clearly, the exact pattern of light and darkness on the tree might be something you’ve never seen before – but because you know what trees are and expect them to be around, the image “snaps” into the schema “tree” and you see a tree there. As usual, this process is most obvious when it goes wrong; for example, when random patterns on a wall or ceiling “snap” into the image of a face, or when the whistling of the wind “snaps” into a voice calling your name.

Most of the things you perceive when awake are generated from very limited input – by the same machinery that generates dreams with no input

— [Void Of Space \(@VoidOfSpace\)](#), September 2, 2016

[Corlett, Frith & Fletcher \(2009\)](#) (henceforth CFF) expand on this idea and speculate on the biochemical substrates of each part of the process. They view perception as a “handshake” between top-down and bottom-up processing. Top-down models predict what we’re going to see, bottom-up models perceive the real world, then they meet in the middle and compare notes to calculate a prediction error. When the prediction error is low enough, it gets smoothed over into a consensus view of reality. When the prediction error is too high, it registers as salience/surprise, and we focus our attention on the stimulus involved to try to reconcile the models. If it turns out that bottom-up was right and top-down was wrong, then we adjust our priors (ie the models used by the top-down systems) and so learning occurs.

In their model, bottom-up sensory processing involves glutamate via the AMPA receptor, and top-down sensory processing involves glutamate via the NMDA receptor. Dopamine codes for prediction error, and seem to represent the level of certainty or the “confidence interval” of a given prediction or perception. Serotonin, acetylcholine, and the others seem to modulate these systems, where “modulate” is a generic neuroscientist weasel word. They provide a lot of neurological and radiologic evidence for these correspondences, for which I highly recommend reading the paper but which I’m not going to get into here. What I found interesting was their attempts to match this system to known pharmacological and psychological processes.

CFF discuss a couple of possible disruptions of their system. Consider *increased* AMPA signaling combined with *decreased* NMDA signaling. Bottom-up processing would become more powerful, unrestrained by top-down models. The world would seem to become “noisier”, as sensory inputs took on a life of their own and failed to snap into existing categories. In extreme cases, the “handshake” between exuberant bottom-up processes and overly timid top-down processes would fail completely, which would take the form of the sudden assignment of salience to a random stimulus.

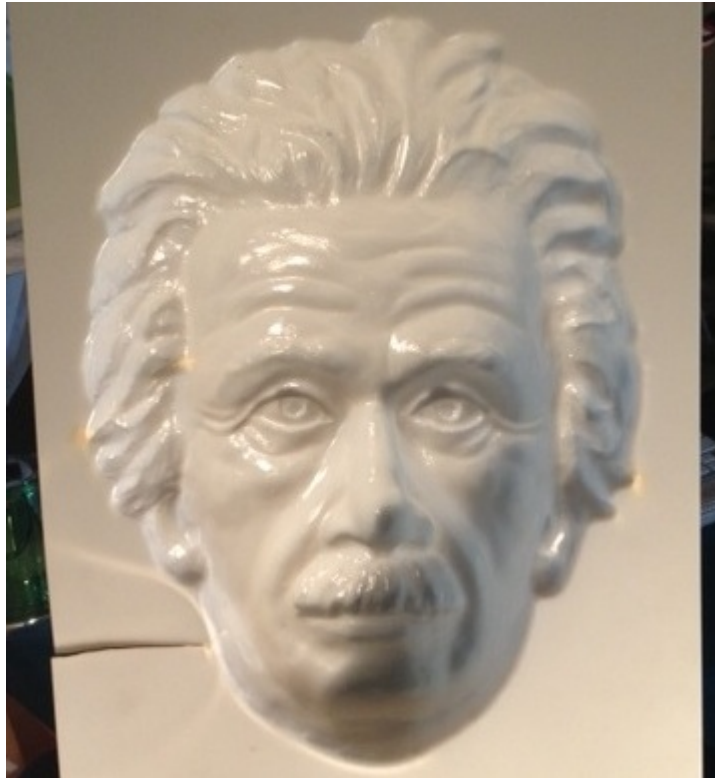
Schizophrenics are famous for “delusions of reference”, where they think a random object or phrase is deeply important for reasons they have trouble explaining. Wikipedia gives as examples:

- A feeling that people on television or radio are talking about or talking directly to them
- Believing that headlines or stories in newspapers are written especially for them
- Seeing objects or events as being set up deliberately to convey a special or particular meaning to themselves
- Thinking ‘that the slightest careless movement on the part of another person had great personal meaning... increased significance’

In CFF, these are perceptual handshake failures; even though “there’s a story about the economy in today’s newspaper” should be perfectly predictable, noisy AMPA signaling registers it as an extreme prediction failure, and it fails its perceptual handshake with overly-weak priors. Then it gets flagged as shocking and deeply important. If you’re unlucky enough to have your brain flag a random newspaper article as shocking and deeply important, maybe phenomenologically that feels like it’s a secret message for you.

And this pattern – increased AMPA signaling combined with decreased NMDA signaling – is pretty much the effect profile of the drug ketamine, and ketamine [does](#) cause a paranoid psychosis mixed with delusions of reference.

Organic psychosis like schizophrenia might involve a similar process. There’s a test called the binocular depth inversion illusion, which looks like this:



[\(source\)](#)

The mask in the picture is concave, ie the nose is furthest away from the camera. But most viewers interpret it as convex, with the nose closest to the camera. This makes sense in terms of Bayesian perception; we see right-side-in faces a whole lot more often than inside-out faces.

Schizophrenics (and people stoned on marijuana!) are more likely to properly identify the face as concave than everyone else. In CFF's system, something about schizophrenia and marijuana messes with NMDA, impairs priors, and reduces the power of top-down processing. This predicts that schizophrenics and potheads would

both have paranoia and delusions of reference, which seems about right.

Consider a slightly different distortion: *increased* AMPA signaling combined with *increased* NMDA signaling. You've still got a lot of sensory noise. But you've also got stronger priors to try to make sense of them. CFF argue these are the perfect conditions to create hallucinations. The increase in sensory noise means there's a lot of data to be explained; the increased top-down pattern-matching means that the brain is very keen to fit all of it into some grand narrative. The result is vivid, convincing hallucinations of things that are totally not there at all.

LSD is mostly serotonergic, but most things that happen in the brain bottom out in glutamate eventually, and LSD bottoms out in exactly the pattern of increased AMPA and increased NMDA that we would expect to produce hallucinations. CFF don't mention this, but I would also like to add my theory of [pattern-matching based mysticism](#). Make the top-down prior-using NMDA system strong enough, and the entire world collapses into a single narrative, a divine grand plan in which everything makes sense and you understand all of it. This is also something I associate with LSD.

If dopamine represents a confidence interval, then increased dopaminergic signaling should mean narrowed confidence intervals and increased certainty. Perceptually, this would correspond to increased sensory acuity. More abstractly, it might increase "self-confidence" as usually described. Amphetamines, which act as dopamine agonists, do both. Amphetamine users report increased

visual acuity (weirdly, they also report blurred vision sometimes; I don't understand exactly what's going on here). They also create an elevated mood and grandiose delusions, making users more sure of themselves and making them feel like they can do anything.

(something I remain confused about: elevated mood and grandiose delusions are also typical of bipolar mania. People on amphetamines and other dopamine agonists act pretty much exactly like manic people. Antidopaminergic drugs like olanzapine are very effective acute antimanics. But people don't generally think of mania as primarily dopaminergic. Why not?)

CFF end their paper with a discussion of sensory deprivation. If perception is a handshake between bottom-up sense-data and top-down priors, what happens when we turn the sense-data off entirely? Psychologists note that most people go a little crazy when placed in total sensory deprivation, but that schizophrenics actually seem to do *better* under sense-deprivation conditions. Why?

The brain filters sense-data to adjust for ambient conditions. For example, when it's very dark, your eyes gradually adjust until you can see by whatever light is present. When it's perfectly silent, you can hear the proverbial pin drop. In a state of total sensory deprivation, any attempt to adjust to a threshold where you can detect the nonexistent signal is actually just going to bring you down below the point where you're picking up noise. As with LSD, when there's too much noise the top-down systems do their best to impose structure on it, leading to hallucinations; when they fail, you

get delusions. If schizophrenics have inherently noisy perceptual systems, such that all perception comes with noise the same way a bad microphone gives off bursts of static whenever anyone tries to speak into it, then their brains will actually become *less* noisy as sense-data disappears.

(this might be a good time to remember that [no congenitally blind people ever develop schizophrenia](#) and no one knows why)



Lawson, Rees, and Friston (2014) offer [a Bayesian link to autism](#).

(there are probably a lot of links between Bayesians and autism, but this is the only one that needs a journal article)

They argue that autism is a form of *aberrant precision*. That is, confidence intervals are too low; bottom-up sense-data cannot handshake with top-down models unless they're almost-exactly the same. Since they rarely are, top-down models lose their ability to "smooth over" bottom-up information. The world is full of random noise that fails to cohere into any more general plan.

Right now I'm sitting in a room writing on a computer. A white noise machine produces white noise. A fluorescent lamp flickers overhead. My body is doing all sorts of body stuff like digesting food and pumping blood. There are a few things I need to concentrate on: this essay I'm writing, my pager if it goes off, any sorts of sud-

den dramatic pains in my body that might indicate a life-threatening illness. But I don't need to worry about the feeling of my back against the back of the chair, or the occasional flickers of the fluorescent light, or the feeling of my shirt on my skin.

A well-functioning perceptual system gates out those things I don't need to worry about. Since my shirt always feels more or less similar on my skin, my top-down model learns to predict that feeling. When the top-down model predicts the shirt on my skin, and my bottom-up sensation reports the shirt on my skin, they handshake and agree that all is well. Even if a slight change in posture makes a different part of my shirt brush against my skin than usual, the confidence intervals are wide: it is still an instance of the class "shirt on skin", it "snaps" into my shirt-on-skin schema, and the perceptual handshake goes off successfully, and all remains well. If something dramatic happens – for example my pager starts beeping really loudly – then my top-down model, which has thus far predicted silence – is rudely surprised by the sudden burst of noise. The perceptual handshake fails, and I am startled, upset, and instantly stop writing my essay as I try to figure out what to do next (hopefully answer my pager). The system works.

The autistic version works differently. The top-down model tries to predict the feeling of the shirt on my skin, but tiny changes in the position of the shirt change the feeling somewhat; bottom-up data does not *quite* match top-down prediction. In a neurotypical with wide confidence intervals, the brain would shrug off such a tiny difference, declare it good enough for government work, and (correctly) ignore it. In an autistic person, the confidence intervals are very

narrow; the top-down systems expect the feeling of shirt-on-skin, but the bottom-up systems report a *slightly different* feeling of shirt-on-skin. These fail to snap together, the perceptual handshake fails, and the brain flags it as important; the autistic person is startled, upset, and feels like stopping what they're doing in order to attend to it.

(in fact, I think the paper might be claiming that “attention” just means a localized narrowing of confidence intervals in a certain direction; for example, if I pay attention to the feeling of my shirt on my skin, then I *can* feel every little fold and micromovement. This seems like an important point with a lot of implications.)

Such handshake failures match some of the sensory symptoms of autism pretty well. Autistic people dislike environments that are (literally or metaphorically) noisy. Small sensory imperfections bother them. They literally get annoyed by scratchy clothing. They tend to seek routine, make sure everything is maximally predictable, and act as if even tiny deviations from normal are worthy of alarm.

They also stim. LRF interpret stimming as an attempt to control sensory predictive environment. If you're moving your arms in a rhythmic motion, the overwhelming majority of sensory input from your arm is from that rhythmic motion; tiny deviations get lost in the larger signal, the same way a firefly would disappear when seen against the blaze of a searchlight. The rhythmic signal which you yourself are creating and keeping maximally rhythmic is the most predictable thing possible. Even something like head-banging serves to create extremely strong sensory data – sensory data

whose production the head-banger is themselves in complete control of. If the brain is in some sense minimizing predictive error, and there's no reasonable way to minimize prediction error because your predictive system is messed up and registering *everything* as a dangerous error – then sometimes you have to take things into your own hands, bang your head against a metal wall, and say “I totally predicted all that pain”.

(the paper doesn't mention this, but it wouldn't surprise me if weighted blankets work the same way. A bunch of weights placed on top of you will predictably stay there; if they're heavy enough this is one of the strongest sensory signals you're receiving and it might “raise your average” in terms of having low predictive error)

What about all the non-sensory-gating-related symptoms of autism? LRF think that autistic people dislike social interaction because it's “the greatest uncertainty”; other people are the hardest-to-predict things we encounter. Neurotypical people are able to smooth social interaction into general categories: this person seems friendly, that person probably doesn't like me. Autistic people get the same bottom-up data: an eye-twitch here, a weird half-smile there – but it never snaps into recognizable models; it just stays weird uninterpretable clues. So:

This provides a simple explanation for the pronounced social-communication difficulties in autism; given that other agents are arguably the most difficult things to predict. In the complex world of social interactions, the many-to-one mappings between causes and sensory input are dramatical-

ly increased and difficult to learn; especially if one cannot contextualize the prediction errors that drive that learning.

They don't really address differences between autists and neurotypicals in terms of personality or skills. But a lot of people have come up with stories about how autistic people are better at tasks that require a lot of precision and less good at tasks that [require central coherence](#), which seems like sort of what this theory would predict.

LRF ends by discussing biochemical bases. They agree with CFF that top-down processing is probably related to NMDA receptors, and so suspect this is damaged in autism. Transgenic mice who lack an important NMDA receptor component seem to behave kind of like autistic humans, which they take as support for their model – although obviously a lot more research is needed. They agree that acetylcholine “modulates” all of this and suggest it might be a promising pathway for future research. They agree with CFF that dopamine may represent precision/confidence, but despite their whole spiel being that precision/confidence is messed up in autism, they don't have much to say about dopamine except that it probably modulates something, just like everything else.



All of this is fascinating and elegant. But is it elegant *enough*?

I notice that I am confused about the relative role of NMDA and AMPA in producing hallucinations and delusions. CFF say that enhanced NMDA signaling results in hallucinations as the brain tries to add excess order to experience and “overfits” the visual data. Fine. So maybe you get a tiny bit of visual noise and think you’re seeing the Devil. But shouldn’t NMDA and top-down processing also be the system that tells you there is a high prior against the Devil being in any particular visual region?

Also, once psychotics develop a delusion, that delusion usually sticks around. It might be that a stray word in a newspaper makes someone think that the FBI is after them, but once they think the FBI is after them, they fit everything into this new paradigm – for example, they might think their psychiatrist is an FBI agent sent to poison them. This sounds a lot like a new, very strong prior! Their doctor presumably isn’t doing much that seems FBI-agent-ish, but because they’re working off a narrative of the FBI coming to get them, they fit everything, including their doctor, into that story. But if psychosis is a case of attenuated priors, why should that be?

(maybe they would answer that because psychotic people also have increased dopamine, they believe in the FBI with absolute certainty? But then how come most psychotics don’t seem to be manic – that is, why aren’t they overconfident in anything except their delusions?)

LRF discuss prediction error in terms of mild surprise and annoyance; you didn’t expect a beeping noise, the beeping noise happened, so you become startled. CFF discuss prediction error as

sudden surprising salience, but then say that the attribution of salience to an odd stimulus creates a delusion of reference, a belief that it's somehow pregnant with secret messages. These are two very different views of prediction error; an autistic person wearing uncomfortable clothes might be constantly focusing on their itchiness rather than on whatever she's trying to do at the time, but she's not going to start thinking they're a sign from God. What's the difference?

Finally, although they highlighted a selection of drugs that make sense within their model, others seem not to. For example, there's some discussion of [ampakines for schizophrenia](#). But this is the opposite of what you'd want if psychosis involved overactive AMPA signaling! I'm not saying that the ampakines for schizophrenia definitely work, but they don't seem to make the schizophrenia noticeably worse either.

Probably this will end the same way most things in psychiatry end – hopelessly bogged down in complexity. Probably AMPA does one thing in one part of the brain, the opposite in other parts of the brain, and it's all nonlinear and different amounts of AMPA will have totally different effects and maybe downregulate itself somewhere else.

Still, it's neat to have at least a vague high-level overview of what *might* be going on.