Hypotheses and Hunches

1. The Atomic Bomb Considered As Hungarian High School Science Fair Project
2. It’s Bayes All The Way Up
3. Why Are Transgender People Immune To Optical Illusions?
4. The Case Of The Suffocating Woman

The Atomic Bomb Considered As Hungarian High School Science Fair Project

I.

A group of Manhattan Project physicists created a tongue-in-cheek mythology where superintelligent Martian scouts landed in Budapest in the late 19th century and stayed for about a generation, after which they decided the planet was unsuitable for their needs and disappeared. The only clue to their existence were the children they had with local women.

The joke was that this explained why the Manhattan Project was led by a group of Hungarian supergeniuses, all born in Budapest between 1890 and 1920. These included Manhattan Project founder Leo Szilard, H-bomb creator Edward Teller, Nobel-Prize-winning quantum physicist Eugene Wigner, and legendary polymath John von Neumann, namesake of the List Of Things Named After John Von Neumann.

The coincidences actually pile up beyond this. Von Neumann, Wigner, and possibly Teller all went to the same central Budapest high school at about the same time, leading a friend to joke about the atomic bomb being *basically* a Hungarian high school science fair project.

But maybe we shouldn't be joking about this so much. Suppose we learned that Beethoven, Mozart, and Bach all had the same childhood piano tutor. It sounds less like "ha ha, what a funny coincidence" and more like "wait, who was this guy, and how quickly can we make everyone else start doing what he did?"

In this case, the guy was Laszlo Ratz, legendary Budapest high school math teacher. I didn't even know people *told* legends about high school math teachers, but apparently they do, and this guy features in a lot of them. There is apparently a Laszlo Ratz Memorial Congress for high school math teachers each year, and a Laszlo Ratz medal for services to the profession. There are plaques and statues to this guy. It's pretty impressive.

A while ago I looked into the literature on teachers and concluded that they didn't have much effect overall. Similarly, Freddie deBoer writes that most claims that certain schools or programs have transformative effects on their students are the result of selection bias.

On the other hand, we have a Hungarian academy producing like half the brainpower behind 20th century physics, and Nobel laureates who literally keep a picture of their high school math teacher on the wall of their office to inspire them. Perhaps even if teachers don't explain much of the *existing* variability, there are heights of teacherdom so rare that they don't show up in the statistics, but still exist to be aspired to?

II.

I've heard this argument a few times, and I think it's wrong.

Yes, two of Ratz's students went on to become supergeniuses. But Edward Teller, another supergenius, went to the same high school but (as far as I know) was never taught by Ratz himself. That suggests that the school was good at producing supergeniuses regarldess of Ratz's personal qualities. A further point in support of this: John Harsanyi also went to the school, also wasn't directly taught by Ratz, and also went on to win a Nobel Prize and invent various important fields of mathematics. So this school – the Fasori Gymnasium – seems to have been about equally excellent for both its Ratz-taught and its non-Ratz-taught pupils.

Yet the Fasori Gymnasium *might not have even been the best high school in its neighborhood*. It competed with the Minta Gymnasium half a mile down the street, whose alumni include Manhattan Project physicists Nicholas Kurti and Theodore von Karman (von Karman went on to found the Jet Propulsion Laboratory), brilliant chemist-philosopher Michael Polanyi, economists Thomas Balogh and Nicholas Kaldor (of Kaldor-Hicks efficiency fame), and Peter Lax, who once said "You don't have to be Hungarian to be a mathematician – but it helps". There are also some contradictory sources suggesting Teller attended this school and not Fasori; for all I know he might have attended both. Once again, most of these people were born in the 1890-1910 period when the Martian scouts were supposedly in Budapest.

Worse, I'm not even sure that the best high school in early 20th-century Hungary was *either* of the two mentioned above. The Berzsenyi Gymnasium, a two mile walk down Gyorgy Street from the others, boasts alumni including multizillionaire George Soros, Intel founder Andrew Grove, BASIC inventor John Kemeny, leading cancer biologist George Klein, great mathematician George Polya, and Nobel Prize winning physicist Dennis Gabor.

Given that the Fasori Gymnasium wasn't obviously better than either of these others, is it possible that the excellence was at a higher level – neither excellent teachers nor excellent principals, but some kind of generally excellent Hungarian culture of education?

This is definitely what the Hungarians want us to think. According to *Cultures of Creativity*:

> What's so special about Budapest's schools? A certain elitism and a spirit of competition partly explains the successes of their students. For example, annual competitions in mathematics and physics have been held since 1894. The instruction the students receive as well as these contests are an expression of a special pedagogy and a striving to encourage creativity. Mor Karman, founder of the Minta school, believed that everything should be taught by showing its relation to everyday life. Instead of learning rules by heart from books, students tried to formulate the rules themselves.

This paper on "The Hungarian Phenomenon" makes similar claims, but adds a few more details:

> The Eotvos Contests were a powerful mean for the stimulation of mathematics on a large scale and were used to motivate mathematical culture in the society. It also provided a channel to search for talented youths. The contests, which have been open to Hungarian high school students in their last year since 1894, played a remarkable role in the development of mathematics.

Okay. But I want to challenge this. During this era, formal education in Hungary began at age 10. By age ten, John von Neumann, greatest of the Hungarian supergeniuses,

already spoke English, French, German, Italian, and Ancient Greek, knew integral and differential calculus, and could multiply and divide 8-digit numbers in his head. Wikipedia notes that on his first meeting with his math teacher, the math teacher "was so astounded with the boy's mathematical talent that he was brought to tears". This doesn't sound like a guy whose potential was kindled by formal education. This sounds like a guy who would have become one of history's great mathematicians even if his teachers had slept through his entire high school career.

Likewise, the book above notes that Dennis Gabor, the Hungarian inventor of holography, "developed his passion for physics during his youth, but did so for the most part on his own". His biography notes that "During his childhood in Budapest, Gabor and his brother would often duplicate the experiments they read about in scientific journals in their home laboratory."

Likewise, consider Paul Erdos, a brilliant mathematician born in Budapest around this time. As per his Wikipedia page, "Left to his own devices, he taught himself to read through mathematics texts that his parents left around their home. By the age of four, given a person's age, he could calculate, in his head, how many seconds they had lived."

I have no knock-down proof that Hungary's clearly excellent education system didn't contribute to this phenomenon. A lot of child prodigies burn out, and maybe Hungary was unusually good at making sure that didn't happen. But it sure seems like they had a lot of child prodigies to work with.

So what's going on? Should we just accept the Manhattan Project consensus that there was a superintelligent Martian scout force in early 20th-century Budapest?

III.

Here's something interesting: every single person I mentioned above is of Jewish descent. *Every single one*. This isn't some clever setup where I only selected Jewish-Hungarians in order to spring this on you later. I selected all the interesting Hungarians I could find, then went back and checked, and every one of them was Jewish.

This puts the excellence of the Hungarian education system in a different light. Hungarian schools totally failed to work their magic on Gentiles. You can talk all you want about "elitism and a spirit of competition" and "striving to encourage creativity", yet for some reason this worked on exactly one of Hungary's many ethnic groups.

This reduces the difficult question of Hungarian intellectual achievement to the easier question of Jewish intellectual achievement.

I say "easier question" because I find the solution by Cochran, Hardy, and Harpending really compelling. Their paper is called A Natural History Of Ashkenazi Intelligence ("Ashkenazi" means Eastern European Jew) and they start by expressing the extent of the issue:

> Ashkenazi Jews have the highest average IQ of any ethnic group for which there are reliable data. They score 0.75 to 1.0 standard deviations above the general European average, corresponding to an IQ 112 – 115. This fact has social significance because IQ (as measured by IQ tests) is the best predictor we have of success in academic subjects and most jobs. Ashkenazi Jews are just as successful as their tested IQ would predict, and they are hugely overrepresented in

> occupations and fields with the highest cognitive demands. During the 20th century, they made up about 3% of the US population but won 27% of the US Nobel science prizes and 25% of the Turing Awards [in computer science]. They account for more than half of world chess champions.

This doesn't seem to be due to any advantage in material privilege; Ashkenazi Jews frequently did well even in countries where they were persecuted. Nor is it obviously linked to Jewish culture; Jews from other regions of the world show no such advantage. So what's going on?

Doctors have long noted that Ashkenazi Jews are uniquely susceptible to various genetic diseases. For example, they're about a hundred times more likely to have Gaucher's Disease, a hundred times more likely to get Tay-Sachs Disease, ten times more likely to have torsion dystonia, et cetera. Genetic diseases are so common in this population that the are official recommendation is that *all* Ashkenazi Jewish couples get screened for genetic disease before marriage. I'm Ashkenazi Jewish, I got screened, and I turn out to be a carrier for Riley-Day syndrome – three hundred times as common in Ashkenazi Jews as in anyone else.

Evolution usually gets rid of genetic diseases pretty quickly. If they stick around, it's because they're doing something to earn their keep. One common pattern is "heterozygote advantage" – two copies of the gene cause a disease, but one copy does something good. For example, people with two copies of the sickle cell gene get sickle cell anaemia, but people with one copy get some protection against malaria. In Africa, where malaria is relatively common, the tradeoff is worth it – so people of African descent have high rates of the sickle cell gene and correspondingly high rates of sickle cell anaemia. In other places, where malaria is relatively uncommon, the tradeoff isn't worth it and evolution eliminates the sickle cell gene. That's why sickle cell is about a hundred times more common in US blacks than US whites.

The moral of the story is: populations can have genetic diseases if they also provide a useful advantage to carriers. And if those genetic diseases are limited to a single group, we expect them to provide a useful advantage for that group, but not others. Might the Jewish genetic diseases provide some advantage? And why would that advantage be limited to Jews?

Most of the Jewish genetic diseases cluster into two biological systems – the sphingolipid system and the DNA repair system. This is suspicious. It suggests that they're not just random. They're doing something specific. Both of these systems are related to neural growth and neural branching. Might they be doing something to the brain?

Gaucher's disease, one of the Ashkenazi genetic diseases, appears to increase IQ. CHH obtained a list of all of the Gaucher's patients in Israel. They were about 15 times more likely than the Israeli average to be in high-IQ occupations like scientist or engineer; CHH calculate the probability that this is a coincidence to be 4×10^-19.

Torsion dystonia, another Ashkenazi genetic disease, shows a similar pattern. CHH find ten reports in the literature where doctors comment on unusual levels of intelligence in their torsion dystonia patients. Eldridge, Harlan, Cooper, and Riklan tested 14 torsion dystonia patients and found an average IQ of 121; another similar study found an average of 117. Torsion dystonia is pretty horrendous, but sufferers will at least get the consolation prize of being really, really smart.

Moving from medicine to history, we find that Ashkenazi Jews were persecuted for the better part of a millennium, and the particular form of this persecution was locking them out of various jobs until the main career opportunities open to them were things like banker, merchant, and doctor. CHH write:

> For 800 to 900 years, from roughly 800 AD to 1650 or 1700 AD, the great majority of the Ashkenazi Jews had managerial and financial jobs, jobs of high complexity, and were neither farmers nor craftsmen. In this they differed from all other settled peoples of which we have knowledge.

They continue:

> Jews who were particularly good at these jobs enjoyed increased reproductive success. Weinryb (1972, see also Hundert 1992) comments: "More children survived to adulthood in affluent families than in less affluent ones. A number of genealogies of business leaders, prominent rabbis, community leaders, and the like – generally belonging to the more affluent classes – show that such people often had four, six, sometimes even eight or nine children who reached adulthood. On the other hands, there are some indications that poorer families tended to be small ones…as an example, in a census of the town of Brody in 1764 homeowner households had 1.2 children per adult member while tenant households had 0.6.

Now we can start to sketch out the theory in full. Due to persecution, Jews were pushed into cognitively-demanding occupations like banker or merchant and forced to sink or swim. The ones who swam – people who were intellectually up to the challenge – had more kids than the ones who sank, producing an evolutionary pressure in favor of intelligence greater than that in any other ethnic group. Just as Africans experiencing evolutionary pressure for malaria resistance developed the sickle cell gene, so Ashkenazim experiencing evolutionary pressure for intelligence developed a bunch of genes which increased heterozygotes' IQ but caused serious genetic disease in homozygotes. As a result, Ashkenazi ended up somewhat more intelligent – and somewhat more prone to genetic disease – than the rest of the European population.

If true, this would explain the 27% of Nobel Prizes and 50% of world chess champions thing. But one still has to ask – everywhere had Jews. Why Hungary in particular? What was so special about Budapest in the early 1900s?

IV.

Okay, sure, everywhere had Jews. But it's surprising exactly how *many* Jews were in early 1900s Hungary.

The modern United States is about 2% Jewish. Hungary in 1900 was about 5%. The most Jewish city in America, New York, is about 15% Jewish. Budapest in 1900 was 25%. It was one of the most Jewish large cities anywhere in history, excepting only Israel itself. According to Wikipedia, the city's late 19th-century nickname was "Judapest".

So is it possible that all the Jews were winning Nobel Prizes, and Hungary just had more Jews and so more Nobelists?

No. This doesn't seem right. The 1933 European Jewish Population By Country site lists the following size for each country's Jewish communities:

Poland: 3 million
Russia: 2.5 million
Romania: 750,000
Germany: 500,000
Hungary: 500,000
Britain: 300,000
France: 250,000
Austria: 200,000

It's hard to find a good list of all famous Manhattan Project physicists, but I tried this article and got the following number of famous Jewish Manhattan Project physicists per country of origin:

Hungary: 4
Germany: 2
Poland: 2
Austria: 2
Italy: 1
Netherlands: 1
Switzerland: 1

Here's an alternative source with a different definition of "famous", broken down the same way:

Germany: 5
Hungary: 4
Poland: 3
Italy: 2
Austria: 2

The main point seems to be disproportionately many people from Central European countries like Hungary and Germany, compared to either Eastern European countries like Poland and Russia or Western European countries like France and Britain.

The Central European advantage over Western Europe is unsurprising; the Western European Jews probably weren't Ashkenazim, and so didn't have the advantage mentioned in the CHH paper above. But is there any reason to think that Central European Jews were more intelligent than Polish and Russian Jews?

I'm not really sure what to think about this. This paper finds that the sphingolipidoses and other Jewish genetic diseases are about twice as common in Central European Jews as in Eastern European Jews, but I have very low confidence in these results. Intra-Jewish gossip points out the Lithuanians as the geniuses among world Jewry, but doesn't have any similar suggestions about Hungarians. And torsion dystonia, maybe the most clearly IQ-linked disease, is unique to Lithuanians and absent in Hungarians.

Probably much more promising is just to focus on the obvious facts of the social situation. Early`1900s Hungary was a great nation and a prosperous center of learning. Remember, we're talking about the age of the Austro-Hungarian Empire, one of the most industrialized and dynamic economies of the time. It might have had advantages that Poland, Romania, and Russia didn't. My list of historical national GDPs per capita is very unimpressed by the difference between Hungarian and Polish GDPs in 1900, but maybe it's wrong, or maybe Budapest was an especially modern part of Hungary, or maybe there's something else I'm missing.

Also, there could have been a difference in the position of Jews in these countries. Russia was still experiencing frequent anti-Jewish pogroms in 1900; in Hungary, Jews were among the country's most noble families. Actually, the extent of Jewish wealth and influence in Hungary sort of defies belief. According to Wikipedia, in 1920 Jews were 60% of Hungarian doctors, 50% of lawyers, 40% of engineers and chemists, and 90% of currency brokers and stock exchange members. "In interwar Hungary, more than half and perhaps as much as 90 percent of Hungarian industry was owned or operated by a few closely related Jewish banking families."

So Central European Jews – the Jews in Hungary and Germany – had a unique combination of intellectual and financial advantages. This means Hungary's only real rival here is Germany. Since they were rich, industrialized, and pretty liberal about Jewish rights at the beginning of the 20th century – and since they had just as many Jews as Hungary – we should expect to see the same phenomenon there too.

And we kind of do. Germany produced its share of Jewish geniuses. Hans Bethe worked for the Manhattan Project and won a Nobel Prize. Max Born helped develop quantum mechanics and also won a Nobel Prize. James Franck, more quantum physics, another Nobel Prize. Otto Stern, even *more* quantum physics, yet *another* Nobel Prize. John Polanyi, chemical kinetics, Nobel Prize (although he was half-Hungarian). And of course we probably shouldn't forget about that Einstein guy. All of these people were born in the same 1880 – 1920 window as the Martians in Hungary.

I think what's going on is this: Germany and Hungary had about the same Jewish population. And they produced about the same number of genius physicists in the same window. But we think of Germany as a big rich country, and Hungary as a small poor country. And the German Jews were spread over a bunch of different cities, whereas the Hungarian Jews were all crammed into Budapest. So when we hear "there were X Nobel Prize winning German physicists in the early 1900s", it sounds only mildly impressive. But when we hear "there were X Nobel Prize winning physicists from Budapest in the early 1900s", it sounds kind of shocking. But the denominator isn't the number of Germans vs. Hungarians, it's the number of German Jews vs. Hungarian Jews, which is about the same.

V.

This still leaves one question: why the period 1880 to 1920?

On further reflection, this isn't much of a mystery. The emancipation of the Jews in Eastern Europe was a difficult process that took place throughout the 19th century. Even when it happened, it took a while for the first generation of Jews to get rich enough that their children could afford to go to fancy schools and fritter away their lives on impractical subjects like physics and chemistry. In much of Eastern Europe, the Jews born around 1880 were the first generation that was free to pursue what they wanted and seek their own lot in the world.

The end date around 1920 is more depressing: any Jew born after this time probably wasn't old enough to escape the Nazis. Almost all the famous Hungarian Jews became physics professors in Europe, fled to America during WWII using channels open to famous physicists, and then made most of their achievements on this side of the Atlantic. There are a couple of stragglers born after 1920 who survived – George Soros' family lived because they bought identity documents saying they were Christian; Andrew Grove lived because he was hidden by righteous Gentiles. But in general Jews born in Europe after 1920 didn't have a great life expectancy.

All of this suggests a pretty reasonable explanation of the Martian phenomenon. For the reasons suggested by Cochran, Hardy, and Harpending, Ashkenazi Jews had the potential for very high intelligence. They were mostly too poor and discriminated against to take advantage of it. Around 1880, this changed in a few advanced Central European economies like Germany, Austria, and Hungary. Austria didn't have many Jews. Germany had a lot of Jews, but it was a big country, so nobody really noticed. Hungary had a lot of Jews, all concentrated in Budapest, and so it was really surprising when all of a sudden everyone from Budapest started winning Nobel Prizes around the same time. This continued until World War II, and then all anyone remembered was "Hey, wasn't it funny that so many smart people were born in Budapest between 1880 and 1920?"

And this story is really, really, gloomy.

For centuries, Europe was sitting on this vast untapped resource of potential geniuses. Around 1880, in a few countries only, economic and political conditions finally became ripe for the potential to be realized. The result was one of the greatest spurts of progress in scientific history, bringing us relativity, quantum mechanics, nuclear bombs, dazzling new mathematical systems, the foundations of digital computing, and various other abstruse ideas I don't even pretend to understand. This lasted for approximately one generation, after which a psychopath with a stupid mustache killed everyone involved.

I certainly can't claim that the Jews were the only people being crazy smart in Central Europe around this time. This was the age of Bohr, Schrodinger, Planck, Curie, etc. But part of me wonders even here. If you have one physicist in a town, he sits in an armchair and thinks. If you have five physicists in a town, they meet and talk and try to help each other with their theories. If you have fifty physicists in a town, they can get funding and start a university department. If you have a hundred, maybe some of them can go into teaching or administration and help support the others. Having this extra concentration of talent in central Europe during this period might have helped Jews and Gentiles alike.

I wonder about this because of a sentiment I hear a lot, from people who know more about physics than I do, that we just don't get people like John von Neumann or Leo Szilard anymore. That there was some weird magical productivity to the early 20th century, especially in Central Europe and Central European immigrants to the United States, that we're no longer really able to match. This can't be a pure numbers game – the Ashkenazi population has mostly recovered since the Holocaust, and people from all over the world are coming to American and European universities and providing more of a concentration of talent than ever. And even though it's impossible to measure, there's still a feeling that it's not enough.

I started down this particular research rabbit hole because a friend challenged me to explain what was so magical about early 20th century Hungary. I think the Jewish population calculations above explain a lot of the story. I'm not sure whether there's a missing ingredient, or, if so, what it might be. Maybe it really was better education. Maybe it really was math competitions and talent searches.

Or maybe it was superintelligent Martian scouts with an Earthling fetish.

It’s Bayes All The Way Up

[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]

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:


THE
CAT


I
LOVE
PARIS IN THE
THE SPRINGTIME

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 wtih all the itroneir ltretrs rgraneanrd is ahonter empxlae 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

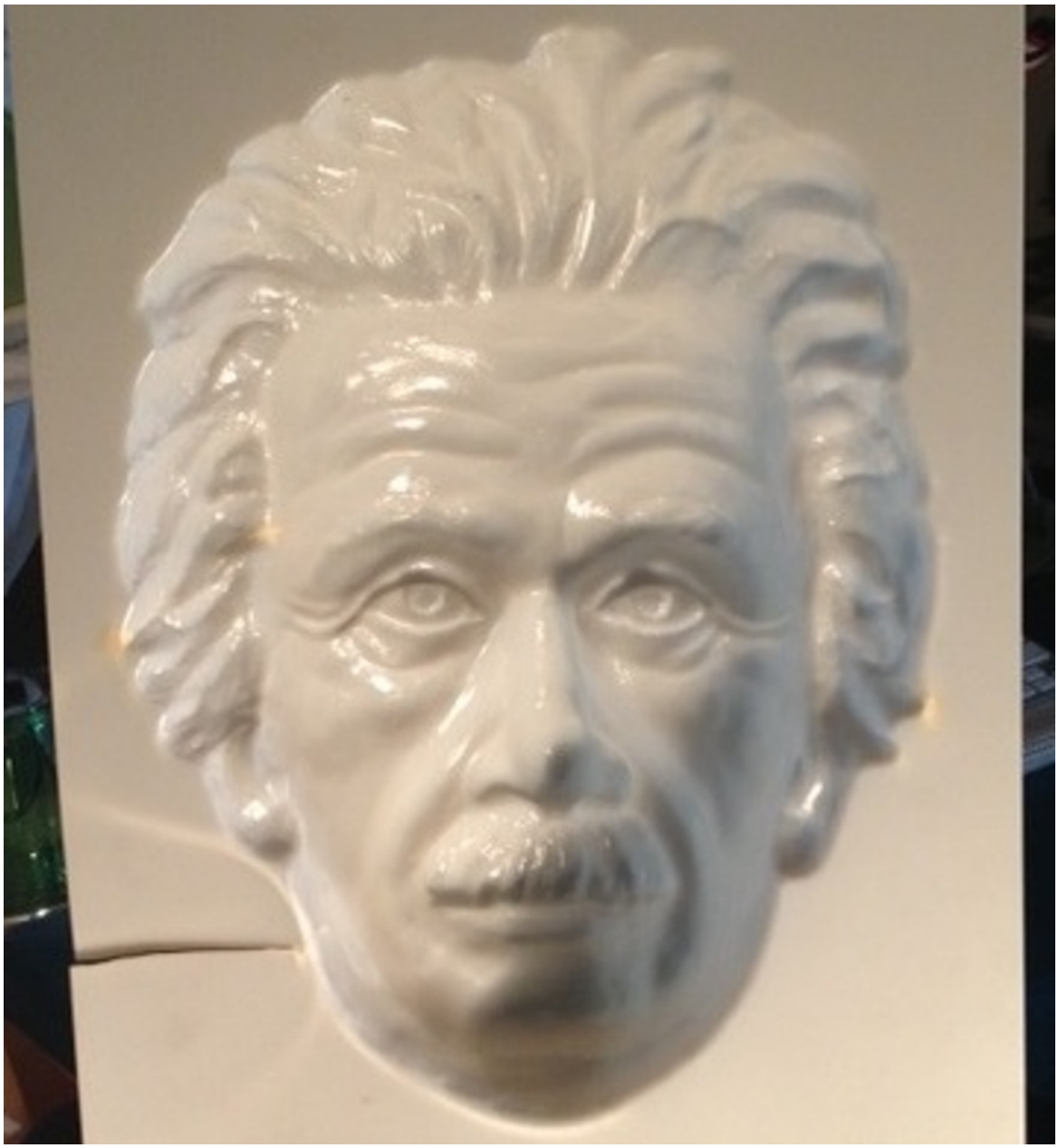
– Seeing objects or events as being set up deliberately to convey a special or particular meaning to themselves

– 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 congentally blind people ever develop schizophrenia and no one knows why)

II.

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 sudden 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 fo 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 dramatically 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.

III.

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 autist 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.

Why Are Transgender People Immune To Optical Illusions?

[Epistemic status: So, so speculative. Don't take any of this seriously until it's replicated and endorsed by other people.]

I.

If you've ever wanted to see a glitch in the Matrix, watch this spinning mask:

Source: http://hearingthevoice.org/2013/11/14/predictive-coding-masterclass/

Did you see it? As the face started to turn away from you, your brain did...*something*, and then you were seeing a normal frontwards-facing mask again. It turns out your visual system has really strong views about whether faces should be inside-out or not, and it's willing to execute a hard override on perception if it doesn't like what it sees.

But not always. Some people get glitchier glitches than others; a few seem almost immune. Studies find schizophrenics and autistic people to be consistently less glitchy than the rest of us. The correlation's not perfect. But it's definitely there. Something about these people's different cognitive processing styles lets them see through the illusion.

I wanted to replicate this result myself. So a few months ago, when I surveyed readers of my blog, I included some questions about perceptual illusions (including a static version of the hollow mask). I got five thousand responses, including a few from schizophrenic and autistic readers. Sure enough, the effect was there.

Schizophrenic readers were about twice as likely to report a weak reaction to the mask illusion as non-schizophrenics (28% vs. 14%, p = 0.04). They were also more likely to have a weak reaction to a similar illusion, the Spinning Dancer (58% vs. 81%, p = 0.01). Readers with a family history of schizophrenia landed in between schizophrenics and healthy controls (16% for mask, 63% for dancer, ns).

Autistic readers were only slightly more likely to report a weak reaction to the mask illusion than neurotypicals (17% vs. 14%), but thanks to our big sample size we could be pretty confident that this was a meaningful difference (p = 0.004). There was no different between autists and neurotypicals on the Spinning Dancer, not even a weak trend (58% vs. 60%, p = 0.4).

Looking deeper, I found a few other anomalies on illusion perception. Most were small and inconsistent. But one stood out: transgender people had an altered response pattern on both illusions, stronger than the alteration for autism and almost as strong as the one for schizophrenia (mask: cis 14% vs. trans 21%, p = 0.003; dancer: cis 58% vs. trans 71%, p = 0.001). These results are very tentative, and need replication. My mass survey isn't a very sensitive instrument, and I place low confidence in any of this until other people can confirm.

But for now, it sure looks like a signal. Something seems off about transgender people's perception, something deep enough to alter the lowest-level components of visual processing. If it's real, what could it be?

II.

A few days ago, trans blogger Zinnia Jones asked me if there might be any neurochemical reason trans people dissociate so much.

Dissociation is a vague psychiatric symptom where you feel like you're not real, or the world isn't real, or you're detached from the world, or something like that. It sounds weird, but if you explain it to someone who's had it, they'll say "Oh yeah, *that* thing!" It's usually unpleasant, and tends to occur in PTSD, borderline personality, and extreme stress.

And in transgender people. The only formal study I can find on this describes it as "greatly prevalent", and suggests that up to 30% of trans people may have dissociative conditions (compared to less than 1% of the general population). This matches trans people's self-reports (1, 2, 3, 4, 5). Anecdotally (according to Zinnia's impression of the trans community) and formally (see Costa & Colizzi 2016) hormone replacement therapy is an effective treatment for dissociative problems.

Intuitively this makes sense. Trans people feel like they're "trapped in the wrong body", so of course they feel detached from their bodies / like their bodies aren't real / like their bodies aren't theirs. Hormone therapy helps solve the "wrong body" problem, so it also solves the dissociative symptoms.

We aim to bridge psychosocial and biological levels of explanation. We can say that someone is stressed out because their boss overworks them, but *also* because they're secreting high levels of cortisol. We can say that someone is depressed because they broke up with their boyfriend, but *also* because they have decreased synaptogenesis in their hippocampus. Causation gets tricky, and this is a philosophical minefield for sure, but overall these two levels should be complementary rather than competitive. So what's the biological correlate to trans people having dissociation problems?

Practically all searches for the biological basis of dissociation end up at the NMDA glutamate receptor, one of the many neurotransmitter systems in the brain. Even though its cousins dopamine and serotonin usually get top billing, glutamate is probably the brain's most important neurotransmitter, and NMDA glutamate receptors in particular are involved in all sorts of interesting things.

Drugs that block NMDA receptors cause dissociation. The most famous dissociative anaesthetic, ketamine, is an NMDA antagonist. So is DXM, a recreational drug that causes dissociation in abusers. Wikipedia's list of dissociative drugs is basically just fifty-five NMDA antagonists in a row. The only other category they list are kappa opioid agonists, and kappa opioid agonism probably – you guessed it – antagonize NMDA. If we take this result seriously, every substance we know of that causes dissociation is an NMDA antagonist in some way.

Does anything improve NMDA function – an effect we might expect to alleviate dissociation? Yes, and among a list of intimidating research chemicals called things like "aminocyclopropanecarboxylic acid" is one familiar name: estrogen. See eg El-Bakri et al, which finds that "estrogen modulates NMDA receptors function in the brain…enhancing NMDA function". McEwen et al: "One of the long-term effects of estradiol [estrogen] is to induce NMDA receptor binding sites in the CA1 region of the hippocampus." Bi et al: "17-B-estradiol [estrogen] enhances NMDA receptor phosphorylation and function." I don't fully understand this research, but it seems to point to estrogen promoting NMDA activity in *some way*.

So transgender people dissociate a lot, a state usually associated with hypofunctioning NMDA receptors. And trans women get better when they take estrogen, a hormone that improves NMDA function. That's interesting. But what does this have to do with those optical illusions?

III.

The Hollow Mask illusion and its cousins may depend on NMDA function.

To oversimplify: the brain interprets the world through Bayesian calculations. In Corlett et al's model, it communicates top-down priors (ie assumptions based on previous knowledge about the world) through NMDA receptors and bottom-up new evidence through AMPA receptors. They write:

> In a hierarchical cortical system in which representations become more abstract with increasing distance from the primary input, higher levels of the hierarchy specify top-down predictions through NMDA receptor signaling and any mismatches between expectancy and experience are conveyed upward through the hierarchy via rapid AMPA and GABA signaling

When you see a hollow mask, the brute facts of how the mask looks are your bottom-up sensory evidence. Your top-down prior is that every other face you've seen for your entire life has been normal, not inside-out. Given the strength of the prior, the prior wins, and your brain interprets the mask as a normal face.

Unless your brain is bad at applying priors, ie its NMDA receptors aren't working that well. Then it just sticks with the bottom-up sensory evidence showing that the mask is hollow.

Schizophrenia and autism both probably involve decreased NMDA function in different ways. For schizophrenia, see eg Olney, NMDA receptor hypofunction model of schizophrenia, and Coyle, NMDA receptor and schizophrenia: a brief history. Ketamine seems to replicate the symptoms of schizophrenia pretty well and is commonly used as a model for the disorder. For autism, see eg Lee, NMDA receptor dysfunction in autism spectrum disorders and this study where screwing with NMDA receptors in mice seems to turn them autistic.

From this we would predict that estrogen would help treat schizophrenia and autism. It does. Schizophrenia is more common and more severe in men than women, with researchers noting that "gonadal steroids may play a role in buffering females against the development of schizophrenia". Women are known to sometimes get schizophrenia triggered by menopause when their estrogen levels decrease. Estrogen supplementation is an effective schizophrenia treatment, and there's some interest in developing estrogen receptor modulators that can help schizophrenic men without making them grow breasts. Meanwhile, autism continues to be about four times more common in men than women, autistic women tend to have more "male-typical brains", and although it's considered unethical to treat autistic boys with estrogen, it works in mice and fish. Once again, doctors are looking into estrogen analogues that don't turn people female as possible autism treatments.

We might also predict that estrogen would increase glitching on the hollow mask. I can't study this directly, but on the survey, 15% of biological males had weak reactions to the illusion, compared with only 11% of biological females, $p = 0.01$. Since women have more estrogen, that looks good for the theory.

Transgender people have higher rates of autism and schizophrenia. The Atlantic actually had a good article about this recently : The Link Between Autism And Trans Identity. They cite one study showing 8% autism rate in trans people (compared to 1-2% in the general population), and another showing that autistic people were 7.5x more likely to express "gender variance". Apparently a lot of trans people have problems getting hormone therapy because their doctors think the gender issues are "just" because of their autism. Some might say that denying people estrogen because they have a condition which studies suggest estrogen can successfully treat is a bit, I don't know, crazy and evil, but I guess people get really weird around this stuff.

My survey broadly confirms these numbers. Autism rates were sky-high in every category – it's almost as if the sorts of people who like reading blogs about how gender is all just NMDA receptors skew more autistic than average – but there was a remarkable difference across gender identities. 15% of cisgender people were autistic, but a full 52% of trans people were.

The survey also finds that about 4% of non-schizophrenic people were transgender, compared to 21% of schizophrenics and self-suspected schizophrenics. Other people have noticed the same connection, and I've met more schizophrenic transgender people than I would expect by chance given the very low rates of both conditions.

If this is right, we end up with this rich set of connections between schizophrenics, autistics, ketamine, dissociative experiences, estrogen, gender identity, and the hollow mask. Anything that decreases NMDA function – schizophrenia, autism, ketamine – will potentially cause dissociative experiences and decreased glitching on the mask illusion. Estrogen will improve NMDA function, treat dissociative experiences, and bring back hollow-mask glitching.

So I wonder: is NMDA hypofunction related to transgender? That would explain the autism and schizophrenia connections. It would explain the hollow mask numbers. It would explain the dissociation. It would explain why estrogen helps the dissociation. And it would explain a lot of internal connections between all of these different conditions and factors.

IV.

I'm going to stop here, even though there's a lot more worth saying on this, because I've already gotten so far into Speculation Land that trying to chain any more conclusions on would probably be premature. So let's switch to some reasons for skepticism.

First, the research into NMDA receptors is *too interesting*. People argue that NMDA is key to depression, key to anxiety, OCD, chronic pain, and borderline personality (my guess is the depression claims are mostly overblown, the borderline claims are 100% absolutely right and revelatory, and I'm agnostic on the others). On the one hand, explaining everything sounds sort of good. On the other hand, it also sounds like what would happen if a field was getting kind of overhyped and slipping into methodology loose enough to prove anything it wanted. Maybe a vague link between a receptor which is literally everywhere in the brain and some psychiatric disease isn't that interesting. A theory that can explain absolutely everything should always cause suspicion.

Second, I'm still not sure what to make of the Hollow Mask results on my survey. Although the transgender results were unusually strong, I did get mildly statistically significant results on about half of the thirty-or-so things I looked at, including

seemingly-unrelated items like political affiliation. Some might argue that this means something is wrong with my survey. Others might argue that hey, we know political attitudes are about 50% genetic, and the last time people tried trace the genes involved all the strongest results were genes for NMDA receptors. Have I mentioned that NMDA receptors are *really interesting?*

Third, I included a second illusion I asked about on the survey, the Spinning Dancer. It also had an odd response pattern among transgender people. But it didn't correlate at all with the Hollow Mask Illusion, and it doesn't seem to be elevated among autists. I don't know what's going on here, and the whole thing makes me more suspicious that all of this is some weird artifact.

Fourth, all this predicts that ketamine will cause reduced glitching on the Hollow Mask. It doesn't. Corlett argues that this is because chronic, but not acute, NMDA dysfunction is required to stop the hollow mask glitches "because [keatmine] has a predominant impact on bottom-up AMPA signaling". I don't really understand this and it seems like a prediction failure to me. On the other hand, chronic marijuana use *does* prevent mask glitching, which might be because of marijuana causing NMDA hypofunction over time, which I guess is a point in favor of the chronicity theory.

Fifth, although trans women dissociate less when they take estrogen, trans men dissociate less when they take testosterone. I can't find whether testosterone has similar NMDA-promoting properties in the brain, although it sometimes gets aromatized to estrogen so that might be relevant. Also, I've never heard of any trans woman taking testosterone or trans man taking estrogen. If that makes dissociation worse – and from the psychosocial perspective it probably should – then that would be a strike against this theory.

Sixth, although I played up the transgender/autism and transgender/schizophrenia links, the truth is that transgender people have higher rates of every mental illness, to the point where it may just be some general factor. I think I'm justified in focusing on these two results because transgender people's higher rates of depression and anxiety are probably just related to being transgender being depressing and anxiety-provoking in this society. But schizophrenia and autism are 80+% genetic, and so harder to explain away like that. Still, somebody could question the relevance of worrying about these two conditions in particular.

I hope some of this can be sorted out in the near future. A first step would be for someone official to replicate the transgender Hollow Mask pattern and prove that it's not just confounded by autism and schizophrenia rates in that population. A very tentative second step would be to investigate whether chronic use of the supplements that improve NMDA function in schizophrenia – like glycine, d-serine, and especially sarcosine – can augment estrogen in improving gender dysphoria. Remember to consult your doctor before trying any weird supplements since they may cause unintended side effects, like becoming a Republican.

It could also be worth trying to understand more explicitly *why* gender identity and NMDA should be linked. This post is long enough already, but I might write more on this in the future. If you want a preview, check out The Role Of Neonatal NMDA Receptor Activation In Defeminization And Masculinization Of Sex Behavior In The Rat and draw the obvious conclusions.

The Case Of The Suffocating Woman

[Content warning: panic, suffocation]

I.

I recently presented this case at a conference and I figured you guys might want to hear it too. Various details have been obfuscated or changed around to protect confidentiality of the people involved.

A 20-something year old woman comes into the emergency room complaining that she can't breathe. The emergency doctors note that she's breathing perfectly normally. She says okay, fine, she's breathing normally *now*, but she's certain she's about to suffocate. She's having constant panic attacks, gasping for breath, feels like she can't get any air into her lungs, been awake 96 hours straight because she's afraid she'll stop breathing in her sleep. She accepts voluntary admission to the psychiatric unit with a diagnosis of panic disorder.

We take a full history in the psych ward and there's not much of interest. She's never had any psychiatric conditions in the past. She's never used any psychiatric medication. She's never had any serious diseases. One month ago, she gave birth to a healthy baby girl, and she's been very busy with all the new baby-related issues, but she doesn't think it's stressed her out unreasonably much.

We start her on an SSRI with (as usual) little immediate effect. On the ward, she continues to have panic attacks, which look like her gasping for breath and being utterly convinced that she is about to die; these last from a few minutes to a few hours. In between these she's reasonable and cooperative but still very worried about her breathing. There are no other psychiatric symptoms. She isn't delusional – when we tell her that our tests show her breathing is fine, she's willing to admit we're probably right – she just *feels* on a gut level like she can't breathe. I'm still not really sure what's going on.

So at this point, I do what any good psychiatrist would: I Google "how do you treat a patient who thinks she's suffocating?" And I stumble onto one of the first convincing explanations I've ever seen of the pathophysiology of a psychiatric disorder.

II.

Panic disorder is a DSM-approved psychiatric condition affecting about 3% of the population. It's marked by "panic attacks", short (minutes to hours) episodes where patients experience extreme terror, increased heart rate, gasping for breath, feeling of impending doom, choking, chest pain, faintness, et cetera. These episodes can happen either after a particular stressor (for example, a claustrophobic patient getting stuck in a small room) or randomly for no reason at all when everything is fine. In a few cases, they even happen when patients are asleep and they wake up halfway through. The attacks rise to the level of a full disorder when they interfere with daily life – for example, a patient can't do her job because she's afraid of having panic attacks while engaged in sensitive activities like driving.

The standard model of panic disorder involves somatosensory feedback loops. Your body is always monitoring itself to make sure that nothing's wrong. Any major organ dysfunction is going to produce a variety of abnormalities – pain, blockage of normal

activities like digestion and circulation, change in chemical composition of the blood, etc. If your body notices enough of these things, it'll go into alarm mode and activate the stress response – increased heart rate, sweating, etc – to make sure you're sufficiently concerned.

In the feedback model of panic disorder, this response begins too early and recurses too heavily. So maybe you have an itch on your back. Your body notices this unusual sensation and falsely interprets it as the sort of abnormality that might indicate major dysfunction. It increases heart rate, starts sweating, et cetera. Then, because it's stupid, it notices the increased heart rate and the sweating that it just caused, and decides this is *definitely* the sort of abnormality that indicates major dysfunction, and there's nothing to do except activate even *more* stress response, which of course it interprets as even *more* organ dysfunction, and so on. At some point your body just maxes out on its stress response, your heart is beating as fast as it can possibly go and your brain is full of as many terror-related chemicals as you can produce on short notice, and then after a while of that it plateaus and returns to normal. So panic disorder sufferers are people who are overly prone to have the stress response, and overly prone to interpret their own stress response as further evidence of dysfunction.

This is probably part genetic and part learned – I have a panic disorder patient who has a bunch of really bad allergies, whose body would shut down in horrifying ways every time he accidentally ate a crumb of the wrong thing, and this seems to have "sensitized" him into having panic attacks; that is, his body has learned that worrying sensations *often* foretell a health crisis, and lowered its threshold accordingly to the point where random noise can easily set it off. I've done a lot of work with this guy, but none of it has been "just ignore your panic attacks, you'll be fine". His body knows what it's doing, and we've got to work from a position of respecting it while also teaching it not to be quite so overzealous.

So this is where my understanding of panic disorder stood until I Googled "how do you treat a patient who thinks she's suffocating?" and came across Donald Klein's theory of panic as false suffocation alarm. You might want to read the full paper, as it's got far too many fascinating things to list here, including a theory of sighing. But I'll try to go over the basics.

Klein is a professor of psychiatry who studies the delightful field of "experimental panicogens", ie chemicals that cause panic attacks if you inject them in someone. These include lactate, bicarbonate, and carbon dioxide, all of which naturally occur in the body under conditions of decreased respiration.

But this is actually confusing. All of these chemicals naturally occur in the body under conditions of decreased respiration. But they don't cause panic attacks then. During exercise, for example, your body has much higher oxygen demand but (no matter how much you pant while running) only a little bit higher oxygen supply, so at the muscle level you don't have enough oxygen and start forming lactate. But exercise doesn't make people panic. Even deliberately holding your breath doesn't make you panic, although it's about the fastest way possible to increase levels of those chemicals. So it looks like your body is actively predicting how much lactate/bicarbonate/CO2 you *should* have, and only getting concerned if there's more than it expects.

So Klein theorized that the brain has a "suffocation alarm", which does some pretty complicated calculations to determine whether you're suffocating or not. Its inputs are anything from blood CO2 level to very high-level cognitions like noticing that you're in space and your spacesuit just ruptured. If, after considering all of this, *and* taking into

account confounding factors like whether you're exercising or voluntarily holding your breath, it decides that you're suffocating, it activates your body's natural suffocation response.

And the body's natural suffocation response seems a lot like panic attacks. Increased heart rate? Check. Gasping for breath? Check. Feeling of impending doom? Check. Choking? Check. Chest pain? Check. Faintness? Check. Some of this makes more sense if you remember that the brain works on Bayesian process combining top-down and bottom-up information, so that your brain can predict that "suffocation implies choking" just as easily as "choking implies suffocation".

A quick digression into medieval French mythology. Once upon a time there was a nymph named Ondine whose lover was unfaithful to her, as so often happens in mythology and in France. She placed a very creative curse on him: she cursed him not to be able to breathe automatically. He freaked out and kept trying to remember to breathe in, now breathe out, now breathe in, now breathe out, but at some point he had to fall asleep, at which point he stopped breathing and died.

So when people discovered a condition that limits the ability to breathe automatically, some very imaginative doctor named the condition Ondine's Curse (some much less imaginative doctors provided its alternate name, central hypoventilation syndrome). People with Ondine's curse don't *exactly* not breathe automatically. But if for some reason they stop breathing, they don't notice. Needless to say, this condition is very, *very* fatal. The usual method of death is that somebody stops breathing at night (ie sleep apnea, very common among the ordinary population, but not immediately dangerous since your body notices the problem and makes you start breathing again) and just never starts again.

Klein says that this proves the existence of the suffocation alarm: Ondine's Curse is an underactive suffocation alarm – and thus the opposite of panic disorder, which is an overactive suffocation alarm. In Ondine's Curse, patients don't feel like they're suffocating even when they are; in panic disorder, patients feel like they're suffocating even when they're not.

This picture has since gotten some pretty powerful confirmation, like the discovery that panic disorder is associated with *ACCN2*, a gene involved in carbon dioxide detection in the amygdala. If you're looking for something that causes you to panic when you're suffocating, a carbon dioxide detector in the amygdala is a pretty impressive fit.

I don't think this is necessarily a replacement for the somatosensory feedback loop theory. I think it ties into it pretty nicely. The suffocation alarm is one of the many monitors watching the body and seeing whether something is dysfunctional, maybe the most important such monitor. It goes through some kind of Bayesian learning process to constantly have a prior probability of suffocation and update with incoming evidence. Let me give two examples.

First, my patient with the bad allergies. Every time he eats the wrong thing, he goes into anaphylactic shock, which prevents respiration and brings him to the edge of suffocating. His suffocation alarm becomes sensitized to this condition, increases its prior probability of suffocation, and so drops its threshold so low that it can be set off by random noise.

Second, claustrophobics. There's a clear analogy between being crammed into a tiny space, and suffocating – think of people who are buried alive. For claustrophobics, for

some reason that link is especially strong, and just being in an elevator is enough to set off their suffocation alarm and start a panic attack. Now, why *agoraphobics* get panic attacks I'm not sure. Maybe fear makes them feel woozy and hyperventilate, and the suffocation alarm treats wooziness and hyperventilation as signs of suffocation and then gets stuck in a feedback loop? I don't know.

III.

Bandelow et al find that you're about a hundred times more likely to develop a new case of panic disorder during the postpartum period than usual.

This can be contrasted with two equally marked trends. Panic attacks decrease markedly during pregnancy, and disappear entirely during childbirth. This last is really remarkable. People get panic attacks at any conceivable time. When they're driving, when they're walking, when they're tired, when they're asleep. Just not, apparently, when they're giving birth. Childbirth is one of the scariest things you can imagine, your body's getting all sorts of painful sensations it's never felt before, and it's a very dangerous period in terms of increased mortality risk. But in terms of panic attack, it's one of the rare times when you are truly and completely protected.

Maternal And Fetal Acid-Base Chemistry: A Major Determinant Of Perinatal Outcomes notes that:

> There is a substantial reduction in the partial pressure of carbon dioxide in pregnancy…this fall is found to reach a mean level of 30-32 mmHg and is associated with a 21% increase in oxygen uptake. The physiological hyperventilation of pregnancy is due to the hormonal effect of progesterone on the respiratory center.

In other words, you're breathing more, you have more blood oxygen, you have less blood CO2, and you're further away from suffocation. This nicely matches the observation that there's fewer panic attacks.

According to Klein, "There is a period of extreme hyperventilation during delivery, which drops the blood carbon dioxide to the minimum recorded under nonpathological conditions". This explains the extreme protective effect of labor against panic disorder, despite labor's seeming panic-inducing properties. When your CO2 is that low, even an oversensitive suffocation alarm is very far from a position where it might be set off.

(source)

Then you give birth, and progesterone – the hormone that was increasing respiratory drive – falls off a cliff. Your body, which for nine months has been doing very nicely with far more oxygen than it could ever need, suddenly finds itself breathing much less than usual and having a normal CO2/oxygen balance. This explains the hundredfold increased risk of developing panic disorder! Somebody who's previously never had any reason to think they're suffocating finds themselves with much less air than they expect (though still the physiologically correct amount of air they need), and if they've got any sensitivity at all, their suffocation alarm interprets this as possible suffocation and freaks out.

This can go one of two directions: either it eventually fully readjusts to your new position and becomes comfortable with a merely normal level of oxygen. Or the constant panic and suffocation feelings sensitize it – the same way that my allergy patient's constant anaphylaxis sensitized him – the alarm develops a higher prior on suffocation and a lower threshold, and the patient gets a chronic panic disorder.

The reason my patient was so interesting was that she was kind of in the middle of this process and had what must have been unusually good introspective ability. Instead of saying "I feel panic", she said "I feel like I'm suffocating". This is pretty interesting. It's like a heart attack patient coming in, and instead of saying "I feel chest pain", they say "I feel like I have a thrombus in my left coronary artery". You're like "Huh, good job".

So I explained all of this to her, and since she didn't know I used Google I probably looked very smart. I told her that she wasn't suffocating, that this was a natural albeit unusual side effect of childbirth, and that with luck it would go away soon. I told her if it didn't go away soon then she might develop panic disorder, which was unfortunate, but that there were lots of good therapies for panic disorder which she would be able to try. This calmed her down a lot and we were able to send her home with some benzodiazepines for acute exacerbation and some SSRIs which she would stay on for a while to see if they helped. She's scheduled to see an outpatient psychiatrist for followup and hopefully he will monitor her panic attacks to see if they eventually get better.

IV.

I realize that case reports are usually supposed to include a part where the doctor does something interesting and heroic and tries an experimental new medication that saves the day. And I realize there wasn't much of that here. But I think that in psychiatry, a good explanation can sometimes be half the battle.

Consider Schachter and Singer (1962). They injected patients with adrenaline (a drug which among other things makes people physiologically agitated) or a placebo. Half the patients were told that the drug would make them agitated. The other half were told it was just some test drug to improve their eyesight. Then a confederate came and did some annoying stuff, and they monitored how angry the patients got. The patients who knew that the drug was supposed to make them angry got less angry than the ones who didn't. The researchers theorized that both groups experienced physiological changes related to anger, but the patients who knew it was because of the drug sort of mentally adjusted for them, and the ones who didn't took them seriously and interpreted them as their own emotion.

We can think of this as the brain making a statistical calculation to try to figure out its own level of anger. It has a certain prior. It gets certain evidence, like the body's physiological state and how annoying the confederate is being. And it controls for certain confounders, like being injected with an arousal-inducing drug. Eventually it makes its best guess, and that's how angry you feel.

In the same way, the suffocation monitor is taking all of its evidence about suffocation – from very low-level stuff like how much CO2 is in the blood to very high-level stuff like what situation you seem to be in – and then adjusting for confounders like whether you're exercising. And I wonder whether telling a patient "You're not actually suffocating, your panic comes from a known physiologic process and here are the hormones that control it" is the equivalent of telling them "You're not really angry,

your agitation comes from us giving you a drug that's known to produce agitation". It tells the suffocation alarm computer that this is a confounder to be controlled for rather than evidence on which to update.

I can't claim to really understand this at a level where it makes sense to me. There are a lot of things that very directly increase CO2 but don't increase panic, or vice versa. Hyperventilation can either cause or prevent panic depending on the situation. There seems to be something going on where the suffocation monitor controls for some things but not others, but this is an obvious cop-out that allows me to avoid making real predictions or narrowing hypothesis-space.

For example, this theory would seem to predict that waterboarding shouldn't work. After all, its whole deal is artificially inducing the feeling of suffocation in a situation where the victim presumably knows that the interrogators aren't going to let him suffocate. You would think that eventually the alarm realizes that "is being waterboarded" is a confounder to control for, but this doesn't seem to be true.

(on the other hand, the inability to condition yourself seems relevant here. It seems like the brain might be not be controlling for whether something is reasonable, but only for whether something is *produced by yourself*. So maybe exercise counts because it's under your control, but waterboarding doesn't count because it isn't. I wonder if anyone has ever tried letting someone waterboard themselves and giving them the on-off switch for the waterboarding device. Was Hitchens' experience close enough to this to count? Why would this be different from letting someone hold their breath, which doesn't produce the same level of panic?)

But overall I find Klein's evidence pretty convincing and feel like this must be at least part of the story. And I think that giving this kind of explanation to somebody can comfort them, reassure them, and (maybe) even improve their condition.