



Psychological Reality: Why We See the World the Way We Do

Audio Transcript

What can we learn from hallucinations? A conversation with psychiatrist Prof. Dr. Philipp Sterzer

0:04

Fabian Mueller: So, welcome, Philipp. You are a professor of Translational Psychiatry at the University of Basel and Chief Physician at the UPK. Before moving to Basel in 2022, you were professor of Psychiatry and Computational Neuroscience at the Charité in Berlin. Your research combines neuroimaging, psychophysics and computational modelling to investigate how predictive processes shape perception and how alterations in these processes may contribute to delusions, hallucinations and psychosis. Philipp, thank you very much for being here with us today. It's an absolute pleasure to have you.

0:38

Philipp Sterzer: Hi, thank you.

0:39

FM: So, before we go into the specific questions, I was wondering whether you could briefly give us a high-level explanation of Bayesian inference and predictive processing. In a nutshell, what's the core idea?

0:52

PS: Sure, the core idea is that the brain uses an internal model of the world to generate predictions about the sensory data that are coming in by our sensory organs. Now, these predictions are, if you will, compared to the sensory input at the level of, for example, primary sensory cortices, so the regions in the brain that primarily process sensory information. And if the incoming sensory information doesn't match these predictions, then an error signal is generated, a so-called prediction error signal. And this prediction error signal is, or can be, used to update the internal model of the world. So, if, for example, I have a certain prediction about, for example, a glass standing in front of me having a certain weight because I assume a certain material, and then if I pick up this glass and it's much lighter than I assumed, then I may update my model about the world with regard to this specific object, updating my prediction towards a different expected weight.

2:02

FM: And so, how is that different? Is that more Bayesian inference, or is that already specifically predictive processing, or how can we really discriminate these two ideas, or what comes first?

2:14

PS: Well, this is the general idea on a conceptual level. Then, of course, the question is how this is implemented. Predictive processing, or more narrowly predictive coding, makes assumptions of how certain types of neurons are involved in this, and the Bayesian idea in this is that the brain uses certain types of algorithmic computations that basically determine how strongly these predictions and prediction errors, or predictions and sensory evidence, are weighted.

2:50

FM: What do you find particularly compelling about predictive processing, and how do you think



we can use this lens to really explain how disruptions in normal processing may contribute to the development of phenomena like hallucinations?

3:04

PS: What I find very compelling on a conceptual level is the idea that the brain constructs or generates our perceptions or everything we perceive and everything we think about the world, because previously, also with regard to models of psychosis, we had these ideas of more like a passive process. Psychosis was often regarded as an altered filtering process or something like that, which assumes, which has the basic assumption, that perception is something more passive, you know, data being processed, being filtered and then being projected somehow into our brains.

Now, what is compelling on the conceptual level is the idea that the brain uses this predictive model, which is, of course, based on learning, primarily, right? It's always updated by prediction errors. So, basically, this model is a result of what we have experienced, what we have learned in our lives, and since this model is so decisive, so important in determining what we perceive, this model, or the idea of predictive processing, provides, well, a good grasp of the differences, the individual differences that we are confronted with, for example, people having very different ideas but also very different perceptions of the world based on their individual predictive models. So, the idea that there is an individual predictive model in each individual brain that comes to different predictions and then also to different conclusions about the sensory input helps us to understand why ideas, convictions, beliefs, but also perception can be profoundly different between individuals.

4:48

FM: So, then, if I understand you correctly, you would sort of say that it's the same processes that we have in health but also here in, in relation to psychosis, but they just seem to be functioning a little bit differently, or how would you, how would you sketch that?

5:03

PS: Yes, precisely. It's the same fundamental mechanisms that lead to what we would call false beliefs, like delusions, or false perceptions, like hallucinations. It's basically always the brain trying to make sense of the, of the sensory data, of what's going on in our sensory cortices.

5:22

FM: Predictive processing accounts distinguish between bottom-up sensory information and top-down expectations or prior beliefs. In psychosis, some accounts emphasize overly strong prior beliefs, whereas others emphasize noisy or excessively weighted sensory prediction errors. Could you help us better understand the central ideas behind these two perspectives?

5:45

PS: Coming from an empirical angle, I have to say we have evidence for both these accounts, like a strong prior or weak prior account, as they are often briefly called or referred to. I don't want to go into the details of the experiments, but early on, like in the early days when the term predictive processing didn't even exist yet, in the eighties, nineties, Bayesian ideas already existed and models existed about, you know, for example, how predictions influence what we sense and what we perceive. And very early models, for example of hallucinations, assumed that there is a deficit in predicting the consequences of one's own actions. These were models mostly proposed by Chris Frith and others. According to these models, you would say, well, priors are too weak in a Bayesian sense, so the predictions are too weak to actually predict the consequences of our own actions, including our own thoughts and also something like inner speech. So, according to this account, the neural activity that gives rise to inner speech, right,



when I'm just thinking to myself or speaking to myself without speaking aloud, the neural signals are generated, you know, in some areas that are involved in speech processing, and this can be regarded as a kind of action, even though it's not visible as an action from outside. Now, if the predictive mechanisms that help us to predict what will be going on if I have these thoughts fail to predict the consequences in sensory cortices, then there may be something like a prediction error signal, right? So there is activity going on that is not adequately predicted, and this could give rise then to the, to the experience that this speech is not generated by myself but by somebody else. That's a very plausible explanation for something that I couldn't predict, right? So this was one of the earlier models of how hallucinations would come about, but then there are also ideas that also have their own plausibility, saying that hallucinations must be a consequence of too strong top-down priors. This could be the case, for example, if I have very strong delusional beliefs, for example, that people in my environment are basically trying to harm me or that they are against me, and if I have very strong beliefs of that kind, that will obviously also influence the way I perceive the world, including what's going on in my sensory cortices, even in the absence of sensory stimuli. So that would explain how a very strong prior belief could actually lead to hallucinations.

Now, to disambiguate between these two accounts: What can we do?

We can do experimentation, we can do computational modelling, but so far we were not able to really clearly disambiguate between these two accounts. Probably both ideas are somehow true, and I think we need to refine our models, maybe abandon this simplistic thinking of either strong or weak priors and define models and then test these models, provide the evidence for these models that give us a more nuanced understanding of how these strong or weak predictions or priors actually can work to produce a symptom like hallucinations.

9:10

FM: Now, there are several accounts of how predictive processing is interpreted, but the framework has also been criticized for being a rather flexible account. We know that the brain is a complex system, and potentially different mechanisms might play a role in different subsystems. With predictive processing, we tend to come from perception and try to make it sort of a more general principle of how the brain might work. Do you think this is defensible, or it's even a good idea to look for this sort of unifying approach to explaining the brain?

9:41

PS: Of course, any unifying approach also runs the risk of being too unifying, too simplistic in a way, and trying to explain everything under one concept, under one umbrella, so we have to be very careful with that. What I think is important to point out is that predictive processing is not one kind of unified model. It's more something like a framework. It is a framework that gives us, like, a conceptual idea. It's based on a normative idea of brain function, and it is, of course, flexible in a way to provide explanations for different types of evidence, maybe too flexible, but that's also a consequence of it not being like a rigid model that is really determined in what it predicts, but it's, as I said, more of a framework. Importantly, however, it is a computationally based framework, so, based on this framework, we can actually come up with concrete formalistic models, mathematical models that make very precise predictions about certain processes, and these predictions are testable by experimentation in combination with modelling.

So, I think this is really important to distinguish between these two levels, the framework level that needs to be flexible to accommodate maybe, you know, various new findings that come to light, and then the actual computational models that are derived from this framework or are embedded in this framework that then make very precise predictions and that can be confirmed or contradicted.



11:16

FM: Beautiful. That's a perfect way of putting it. Thank you. So, looking ahead, what do you think is the most important unanswered question in the field? What needs to happen for predictive processing research to really make a meaningful difference for clinical assessment or treatment?

11:32

PS: That's a big question. But it's a good question, and I think predictive processing already has made a meaningful contribution. It has helped us think differently about psychosis on a more conceptual level, to give us an understanding of how psychotic symptoms may come about without basically relying on neurobiological deficits as an explanation, right, framing these types of experiences as embedding them in a normative framework of brain function. I think that's very helpful.

But I think we need to, we need to dig deeper, so far also referring to the levels of observation in the brain, right, according to David Marr. Clearly, predictive processing operates on the computational level and the algorithmic level, but so far very little on the implementational level, and I think we really need to learn more about how predictive processes are implemented at the neural level and, especially, how alterations of perception and beliefs, hallucinations and delusions, are related to predictive processing alterations that are based on certain neurobiological changes, so, for example, on the neural circuit level: which transmitter systems, which neural circuits, in particular cell types in the cortex, for example, are responsible for the predictive coding or predictive processing alterations that we observe, because this will be important also for the development of new interventions on this basis. This, I think, is one of the big questions.

13:19

FM: And maybe to wrap things up: So, after a decade of research at this intersection of psychiatry and mechanistic explanations of mental illness: What have you personally learned? What is sort of your take-home message from your research in the last few years?

13:35

PS: For me, the most important take-home message is, as I have already alluded to, basically a new understanding of psychosis as an alteration, not necessarily a deficit, but an alteration of inferences that the brain needs to make, has to make, and viewing these symptoms and viewing the experiences of patients like that helps me to, I would say, even to better empathize with the people who have these experiences, because it gives me a new angle of actually understanding how these experiences come about: What is the function of these experiences? How do they sometimes also help people to make more sense of their experiences?

So, we know that, for example, delusional beliefs also have positive effects on people, because the uncertainty that they sometimes experience—everything is unclear, everything is uncertain—that is associated with a lot of anxiety. But if we have an explanation for that now from the perspective of a patient, even if it's a delusional explanation, that may be tremendously helpful in reducing this uncertainty and, you know, also relaxing a little bit, because once you have an explanation, things at least appear easier.

That does not mean that these delusions per se are healthy or something, I don't know. We don't have to make any statements regarding the value of these, of these explanations, but in a way they help the individual, and I think predictive processing, this framework, this understanding of the brain, helps us to also view these symptoms in a different way, and also to see their functionality and sometimes even benefit for people.



University
of Basel

15:18

FM: Thank you so much for being here. It's always very inspirational to talk to you. I think you have made great contributions to the field, and the way you think and talk about these things is extremely helpful for us doing the research, and I hope that it will also ultimately benefit the patients. Thank you so much for being here with us today, Philipp, and I wish you have a wonderful day.

15:37

PS: Thank you very much.