HHS Public Access

Author manuscript

Trends Cogn Sci. Author manuscript; available in PMC 2020 February 01.

Published in final edited form as:

Trends Cogn Sci. 2019 February ; 23(2): 114–127. doi:10.1016/j.tics.2018.12.001.

Hallucinations and Strong Priors

Philip R. Corlett[1,*], **Guillermo Horga**[2], **Paul C. Fletcher**[3], **Ben Alderson-Day**[4], **Katharina Schmack**[5], and **Albert R. Powers III**[1]

[1]Department of Psychiatry, Yale University, New Haven, CT, USA

[2]Department of Psychiatry, Columbia University, New York, NY, USA

[3]Department of Psychiatry, University of Cambridge, Cambridge, UK. The Cambridgeshire and Peteborough NHS Foundation Trust, UK

[4]Department of Psychology, Durham University, Durham, UK

[5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

Abstract

Hallucinations – perceptions in the absence of objectively identifiable stimuli – illustrate the constructive nature of perception. Here we highlight the role of prior beliefs as a critical elicitor of hallucinations. Recent empirical work from independent laboratories shows strong, overly precise, priors can engender hallucinations in healthy subjects, and that individuals who hallucinate in the real world are more susceptible to these laboratory phenomena. We consider these observations in light of work demonstrating apparently weak, or imprecise, priors in psychosis. Appreciating the interactions within and between hierarchies of inference can reconcile this apparent disconnect. Data from neural networks, human behavior and neuroimaging support this contention. This work underlines the continuum from normal to aberrant perception, encouraging a more empathic approach to clinical hallucinations.

Hallucinations and Perception

> "Instead of saying that an hallucination is a false exterior percept, one should say that the external percept is a true hallucination." [1]

Hallucinations (see Glossary) are percepts without corresponding external stimuli [2]. They characterize many serious mental illnesses like schizophrenia and post-traumatic stress disorder [3]. They occur in the context of Alzheimer's and Parkinson's diseases and epilepsy [4], hearing loss [5] and eye disease [6]. But they frequently occur in the absence of any detectable illness, as isolated experiences in up to 50% of people (following bereavement, for example) and in between 2 and 10% of the population on a daily basis [7]. They occur in all sensory modalities, though auditory and visual are most commonly reported. There has

*Corresponding author. philip.corlett@yale.edu.

been a long and growing appreciation of the constructive nature of perception [1]: it is more than mere receipt of sensory information, instead involving a synthetic process, based upon prior expectancies (henceforth **priors**) [8]. Whilst this is efficient with regards to information processing [9], it can be prone to error. In particular, this renders perception similar to hallucination [1] [10]. Here we are concerned with the role of priors in generating hallucinations [11].

We focus on auditory verbal hallucinations (AVHs) or **voices**, though the principles that we outline do, we believe, extend beyond audition. AVHs occur in around 80% of patients diagnosed with schizophrenia but they present us with profound clinical problems: since they are so common, their status as diagnostic markers for mental illness is very uncertain [12]. Moreover, when treatment is indicated, they can, in a significant proportion of people, prove persistent [13]. Thus, there is a real need to understand AVHs. We review work using formal computational models of perception to understand the basis of AVH [14] and suggest that hallucinations arise when prior beliefs exert an inordinate influence on perception [11] [15].

According to **predictive coding** accounts, perception involves adopting a hypothesis that explains what is causing our current sensation [16], an inference that is optimized by prior knowledge about probable candidates [16]. Those priors are compared to incoming sensation and **prediction errors** are computed. If priors are more precise than sensory inputs, we consider them strong. They will dominate inference and prediction errors will be ignored [17–20]. On the other hand, relatively more precise prediction errors will dominate priors and drive belief updating (changing ones' priors for subsequent inference). The powerful contribution of expectation to perception leads us to speculate that hallucinations might arise when prior predictions exert an inordinate influence over perceptual inferences, creating percepts with no corresponding stimuli at all [11] [15].

Data consistent with the strong prior account

That prior experiences could engender hallucinations in healthy volunteers was first reported by Carl Seashore, working at the Yale Psychological Laboratory in 1895. Seashore reported hallucinations in one modality (touch, sight, taste) conditional on learned cues in another sensory modality, and engendered hallucinations by mere suggestion [21]. A more recent example of hallucination-by-suggestion: When healthy volunteers were instructed they would hear the song *White Christmas*, but instead they were played white noise, 5% reported experiencing Bing Crosby's voice [22]. People who hear hallucinatory voices in their everyday lives are more susceptible to this effect [23].

Pavlov's assertion that such inferences were similar to the conditioned reflex [24] has garnered empirical support. Ellson in the 1940s embraced the conditioning paradigm and employed more controlled procedures for consistent stimulus delivery. The illumination of a tungsten filament bulb was presented as a predictor of the presentation of a near threshold 1kHz tone. After learning this association, the tone could be omitted and participants would continue to report tones, conditional on the illumination of the bulb [25]. These studies underline the role of learning in these hallucinations. The predicting stimulus needn't be

visual and the hallucination needn't be auditory. Tones can likewise be established as predictors of visual stimuli (e.g. dimly illuminated triangles), such that a 1khz tone presentation can engender a visual hallucination (of triangles that were never presented). These experiences even transfer out of the laboratory; subjects report seeing triangles on their television screen, conditional on hearing a 1khz tone, even when no triangles were presented [26].

More recently, visual-auditory conditioning has been employed to demonstrate that voice-hearing patients are significantly more susceptible to this effect than patients without hallucinations and controls [27]. However, these observations (and those that preceded them) could have been driven by the demand characteristics perceived from the experimental situation. More contemporary research has sought to mitigate this concern. Prior knowledge of a visual scene confers an advantage in recognizing a degraded version of that image [28]. Patients at risk for **psychosis**—and, by extension, voice-hearing—were particularly susceptible to this advantage and its magnitude correlated with hallucination-like percepts. Similarly, there is a version of this effect in audition; voice-hearing participants appear to have an enhanced prior for speech in degraded auditory stimuli even when not explicitly instructed. Indeed, they were able to detect that degraded speech earlier than control participants [29]. It is difficult to attribute these findings to demand characteristics.

A bias toward top-down information appears to be the basis for "sensory conditioning" [27, 30–33], wherein modern psychophysics was brought to bear on the Ellson paradigm, establishing a visual stimulus as a predictor of a difficult-to-detect, near-threshold, auditory stimulus. Participants begin to report auditory stimuli that were not presented on the basis of the visual cue. This effect is amplified in individuals who hallucinate. We recently showed that this effect is mediated by precise prior beliefs, that those priors are more precise in people who hallucinate, and that people with a diagnosed psychotic illness are less likely to update those prior beliefs in light of new evidence [34]. Critically, the neural circuit underlying these conditioned phenomena—including superior temporal gyrus and insula—largely overlapped with the circuit engaged when patients report hearing voices in the scanner [34, 35].

Individuals with more striatal dopamine (itself a marker for psychosis risk [36]) are more susceptible to this impact of prior expectation on auditory perception: those with higher dopamine are more likely to perceive target auditory stimuli embedded in a stream consistent with their expectations rather than what was actually presented to them [37]. Visual hallucinations in Parkinson's disease have likewise been related to precise priors [38], which may be driven by dopaminergic, cholinergic, and serotonergic perturbations[39]. Whilst we focus here on AVH, we suggest that precise priors in may underwrite hallucinations in other modalities and illnesses. The challenge will be characterizing them and discerning why each specific underlying pathology culminates in voices rather than visions, hallucinations rather than delusions (see Box 1).

Taken together, it appears that across sensory modalities prior experiences can induce hallucinatory experiences even in healthy participants. The circuits and neurotransmitters involved appear to align with those implicated in psychotic illness and indeed, people with

hallucinations in their everyday lives are more susceptible to these prior-driven effects in the laboratory. The source of those priors and their relationship to the **precision** of sensory evidence (i.e., prediction errors) in that modality or other modalities (including sensations related to one's own actions) is critical to understand. We attempt to do so in what follows.

Data that challenge the strong prior account

There are at least two lines of empirical work that challenge our account (i) the apparent immunity to visual **illusions** and (ii) the failure of **corollary discharge** in schizophrenia. (Corollary discharges are predictions of the sensory consequences of our actions and, are thought by some, to underwrite our inference about **agency** for events [40] - a type of prior [41]). Both are considered indicative of less precise priors. We suggest that these data may actually be consistent with the strong prior account – if we allow that these precise priors arise as a mollifying response to the uncertainty engendered by less precise priors elsewhere in the system: either at lower hierarchical levels – in the case of illusions, or in parallel hierarchies in a centrifugal network, as in corollary discharge.

Visual Illusions

Illusory perception departs from actual sensation. For example, artists can create illusory depth on a canvas by mimicking the images that would be formed on the retina by a truly three-dimensional environment [42]. Illusions are thought to be mediated by top-down priors [43]. That is, our prior belief about some feature of sensation sculpts our ultimate percept so it conforms to the belief, rather than the sensation. There are empirical data that demonstrate a reduced sensitivity to some illusions in patients with schizophrenia [44, 45], which tends to suggest that their priors are less precise. However, such effects rarely correlate with symptom severity and when they do, they do not relate to hallucinations specifically. Furthermore, other illusions are enhanced in people with schizophrenia (for example, the three flash illusion – in a manner that correlates with positive symptoms [46]). Hierarchy may be crucial here – since the three flash illusion involves temporal expectations, it is likely to engage more complex and, we would argue, higher-level processing than other more static illusions [47].

It is possible that illusions could fail at low levels of processing and hallucinations could be generated as a response, higher in the hierarchy, as suggested by the work with deep neural networks (see below). In Figure 1, for example, the precision of priors and prediction errors at the sensory level are relatively balanced. When illusions are not perceived by patients with schizophrenia, it could be that they fail to attenuate sensory precision, enabling prediction errors to ascend the hierarchy to induce belief updating. We propose that these un-attenuated (or un-explained away) prediction errors induce a particular sort of high-level prior belief that is the hallucination. There are even data suggesting that psychotic individuals with and without hallucinations utilize different priors to different extents in the same task. People with hallucinations have strong/precise perceptual priors that are not present in psychotic patients who did not hallucinate and who, indeed, may have weak/imprecise priors [34].

Corollary Discharge

The suggestion that a disturbance of corollary discharge, leading to erroneous attribution of inner speech to an external source, can explain voice-hearing [48–53], also appears incompatible with our account. Critically, the theory is that this inappropriate external attribution arises from a failure of prediction of the sensory consequences of one's own actions, which appears to contradict the idea that hallucinations arise from overly-precise priors.

Corollary discharges are predictions of the sensory consequences of our actions. Normally, they cancel out the expected consequences of actions. Unexpected consequences are more readily attributed to external causes: the more prediction error is generated, the more likely an action (or perhaps a thought) has external (rather than internal) causes [51]. These predictions in the form of corollary discharges have been increasingly referred to as priors in the literature. Their failure in people with schizophrenia has been associated with the severity of psychotic symptoms, though not always hallucinations specifically [48–53]. These data trouble our model because they seem to represent instances of less precise priors causing psychotic symptoms. How can this be?

Much like prediction errors from lower in the hierarchy can be suppressed by updating or increasing the precision of higher level priors, we propose that a similar compensation may occur. According to predictive processing theory, for any given sensory experience, there is a cause to be hypothesized. This cause may be agentic or non-agentic in origin (if non-agentic, the problem is simply one of perceptual identification). If agentic, there is an additional problem: was it self or an external agent? Here, the winning hypothesis will be determined by the relative precision of our predictions about internal and external cues. We suggest that the unattenuated prediction errors arising from the poorly predicted consequences of self-generated actions portend a misattribution of agency. This then leads to an appeal to hypotheses or beliefs that these prediction errors were generated extrinsically: "I am not sure it is me so it must be you".

Egocentric processing, in the form of corollary discharges, has been detected in multiple brain regions [40]. Its impairment in schizophrenia and psychosis in particular has been demonstrated in eye-movement [54] and force matching tasks [55]– in each case, impaired corollary discharge correlates with the severity of positive symptoms. Some inferences are more impacted by egocentric processing than others [56]. Indeed, visual processing, whilst integrated, possesses dissociable streams for inferences about exteroceptive, interoceptive, and proprioceptive states (i.e. actions). Likewise, speech has reception and motor areas [57]. The auditory cortex receives a range of inputs from these regions, within and between hemispheres, and across different timescales, mediating specific speech predictions and thus specific contents [58], and we would argue agency attributions for perceived speech. It would be interesting to map impairments in each specific predictive mechanism to the contents of particular hallucinations.

Intuitively, imagine that I failed to generate an accurate prediction of my inner speech. This would induce large amplitude prediction errors over the consequences of that speech. Now, these large amplitude prediction errors themselves provide evidence that it could not have

been me speaking; otherwise, they would have been resolved by my accurate predictions. Therefore, the only plausible hypothesis that accounts for these prediction errors is that someone else was speaking. Note that this formulation is consistent with both sides of the story. On the one hand, a failure to suppress (covert or overt) sensory prediction errors in this setting can be due to a failure of sensory attenuation - or due to a failure to generate accurate predictions (i.e., corollary discharge). The consequence of these failures is the false inference that I was not the agent of these sensations. Crucially however, this conclusion rests upon precise prior beliefs about sensory precision or the amplitude of prediction errors; namely, "*large amplitude prediction errors can only be generated by things I can't predict*". It is worth noting here that such beliefs mean that prediction errors are informative not just about one's degree of surprise but also about the origin or nature of an input: perceptions generated by external sources are more surprising than those generated by oneself and perceptions generated by external agents are more unpredictable than those that are non-agentic (for a fuller discussion of this see [59]). Technically, these are known as hyperpriors —in other words, prior beliefs about precision as opposed to content. In short, one can have both an abnormality of sensory precision and prior precision that, together, underwrite the Bayes optimal but false inference that someone else caused my sensations.

Bayesian models of multi-sensory integration emphasize —and empirical work confirms — that the prioritization of streams of information during inference depends on their relative reliability or precision [60–63]. This case has also been made for reinforcement learning [64]. The rubber-hand illusion [65] — to which people with psychosis are more sensitive [66] — is likewise explained as Bayesian integration of visual and proprioceptive sources of information under the prior assumption of a common source [67]. People tend to infer that common cause and thus conclude that a prop hand belongs to their body when they see the hand stroked in synchrony with what they feel (they report feeling ownership of the hand, they show fear responses when the prop hand is threatened [68] and they mount an immune response to that actual hand [68]), suggesting a malleability of egocentric agency mechanisms. People with psychosis even experience the illusion in the asynchronous control condition [66], perhaps due to an expanded binding window of touch and vision [69], under the influence of a stronger prior on the precision of **exteroceptive**/**allocentric** sources relative to **interoceptive**/**egocentric** sources [70].

Audiovisual integration of sensory cues is critically dependent upon the relative precisions of the information to be integrated [62]. Thus, if we assume similar integration of external and self-related sources of information in overall inference, when corollary discharge fails, precision of self-processing decreases, and as such weighting shifts towards external sources and reliance on external priors should increase. Thus, weak/imprecise and strong/precise prior effects can co-exist. And failures of corollary discharge can, in theory, give rise to the high precision prior expectation that a voice will be heard – as has been observed [29].

Perturbed integration of internal and external cues will also disrupt sense of agency. For example, internal action cues (experience of an intention, sensory feedback) may be imprecise while the external cues (outcome of action for example) may be precise. This would lead some thoughts and actions to be experienced as unintended and thus externally generated [71, 72]. At the same time, the binding of action and outcome may be increased

(due to the relative precision of outcome processing), which has been empirically demonstrated in psychosis, ketamine infusion, and psychosis proneness [73–77].

The co-existence of precise priors for external perception and imprecise for self-agency speaks to the way in which hierarchies of inference are organized, compartmentalized and integrated. When we think about more than one hierarchy, say for inferences about internal states, external states and actions [78], there are – as we ascend each hierarchy – more and more points of intersection, such that, at their apex, there is an amodal model that integrates them into a coherent percept [79]. It is our contention that the relative impact of each hierarchy to ongoing experience is proportional to the precision of its predictions, and, when one element is imprecise or noisy, other hierarchies minimize the precision-weighted prediction error.

Thus, both within and between hierarchies, imprecise priors can subtend more prcise ones and the apparent contradiction between datasets and theories of hallucination is reconciled. We predict that corollary discharge deficits, as measured with the force matching task or two-step saccade task [54], should correlate with strong/precise priors in the tasks reviewed presently.

Strong Priors & Inner Speech

If precise priors do drive hallucinations, then the relationship between inner speech and hallucinations still needs to be explained. Why is the theory that when people hear voices they are hearing their own inner speech so compelling? And how do we explain the functional imaging data derived from symptom capture? When people report hearing voices in the scanner, their speech production and reception circuitry, including Wernicke's area (within the superior temporal gyrus) and Broca's areas, are engaged [61]. Non-speech perceptual predictions seem to engage those same brain regions, including superior temporal gyrus [80] [34], which is itself tuned to the statistics of human speech [81]. The activity of speech circuitry during hallucinations may reflect the imposition of priors rather than the presence of inner speech *per se*. Thus, we can explain the consistent engagement of speech (or we would posit prediction) regions during hallucinations and their specific relationship to hallucination severity. Voices seem likely to be the dominant (though not exclusive) content of auditory hallucinations because our auditory apparatus is tuned to (i.e., has precise priors for) the natural statistics of speech. And they may be experienced as agents communicating, because we believe (based on our overwhelming experience of the contiguity between voices and agents) that voices are typically attached to an agent [82]. The higher-level prior that emerges from paranoia (that one is the cynosure of others' preoccupations, thoughts and actions) provides a ready-made expectation that there will be agents and communication. In other words, at one level, the voice is an expected characteristic of auditory processing and, in addition, the agency/paranoia prior increases the voice prior (since voices typically emanate from agents). The experience of the voice of course reinforces the higher-level prior. Voices are distressing because they are non-consensual and they engage the highest levels of our inferential hierarchy—those levels that contain our narratives about ourselves [83] - but also because they are omniscient, since the content and not just the form is

constructed by expectation, it has the potential to touch on all that preoccupies and worries the person.

Basic Preclinical Mechanisms

Expanding the explanation of hallucinations beyond an inner speech focus proffers a number of advantages. First, and most importantly, it allows a more inclusive definition of voice-hearing to incorporate sensed presences, voices without speech, and the rest of the rich range of hallucination phenomenology that characterizes the lived experience of voices [84]. In proposing a mechanism such as this, it reminds us that beliefs are part of the perceptual process. This is crucial because it highlights that we cannot treat a voice as an isolated symptom or aberration but must include a consideration of their corpus of experience and belief. Indeed, it encourages a much more nuanced view of the gradation from aberrant perceptions to aberrant beliefs [85].

At a more reductionist level of analysis, our account suggests the exciting possibility of modeling hallucinations in species other than humans. Whilst simple conditioning processes may be involved in human language development [86], and clearly other species have language-like processes, these are not as developed and intricate as human speech. Animals cannot report their subjective experiences in a manner that we can understand, unlike humans who hear voices. Prior work on animal hallucinations has tended to focus on objective behavioral responses to psychotomimetic drugs, like head twitch responses under serotonergic hallucinogens and stereotypies induced by amphetamines [87]. Despite the neurochemical relevance (and the potential of these behaviors to guide antipsychotic drug development) [87], to call these hallucinations entails an unacceptable degree of anthropomorphism (to us at least).

Our account of hallucinations avoids this. It is grounded in much earlier work in associative learning theory, wherein learned predictive associative links between stimuli can, on presentation of one of the associates, be so precise that the other is evoked and richly experienced, perhaps as a hallucination. This possibility was first acknowledged by Jerzy Konorski in 1967, who alluded to a *'perceptual hunger'* that gave rise to hallucinations [88], a sentiment echoed more recently by the suggestion that aberrant prediction errors induce a *'hunger for priors'*, a drive to reconcile errors in expectation that may manifest as a bias towards prior experiences and ultimately hallucinations [89]. With regards to animal conditioning and hallucinations, Konorski argued that conditioned motor responses to an external stimulus in a dog could be elicited by predictors of those stimuli and used to infer that those stimuli were '*hallucinated*' [88]. This line of inquiry was further developed and formalized by Peter Holland and colleagues [90]. When a hungry rat is presented with a tone and subsequently a sweet sugar solution, the rat learns after only a few trials that the tone predicts sugar. By making the rat nauseous in the presence of the tone, it can be demonstrated that the tone evokes a highly realistic, sensory representation of the sugar, which the rat has trouble distinguishing from reality – because he transfers the nausea associated with the tone to the sugar and subsequently consumes less sugar [91]. With extended training, rats stop having these cue-induced hallucinations [91], but not in animal models that recapitulate the biology of psychosis [92, 93]. Recent work suggests that these

representation-mediated phenomena are dopaminergically mediated. For example, behavioral over-expectations based on the rats' associatively learned model of reward contingencies are driven by dopamine signaling [94–96].

When human participants' sense of control over their environment is intentionally decreased (with spurious feedback), they tend toward illusory pattern perception, seeing nonexistent signal in noise and detecting illusory trends in the stock market [97]. This uncertainty may also be signaled in part by dopamine [98]. As described above, people who weighted their priors more strongly during perceptual inference have higher levels of striatal dopamine including those who hallucinate outside of the laboratory [37].

The deep hierarchical **neural networks** described below may be uniquely suited to modeling these representation-mediated over-expectation phenomena [99], since they are capable of capturing the underlying complex, contextually sensitive, associative relationships that drive these effects. We argue the same may be true of hallucinations. The debate continues in artificial intelligence as to whether there should be any innate prescribed structures within deep neural networks [100]. With regards to our account of hallucinations, we do not believe that humans are born with their models of the world pre-configured; we, like neuro-constructivists [101], believe that humans infer their parameter values through experience. But, those values can be innately constrained. For example, the expectation that a caregiver would protect us may be relatively hardwired. If that expectation is violated, we may develop a world model—and a set of social expectations—that color our perceptual inferences in a maladaptive manner [102].

Neural Network Models of Hallucinations

In some of the earliest computational psychiatry, the late Ralph Hoffman implemented an artificial neural network model (a Hopfield network) that could learn outputs given particular input patterns [103]. By pruning the allowable connections Hoffman made the network 'hallucinate' spurious outputs [103]. State of the art AI, 'Deep Learning' [104] involves stacks of Hopfield-like networks, separated by hidden layers, that learn representations of data, each stage in the hierarchy learns to generate or reconstruct the activation patterns in the stage below. One such-network, the **Deep Boltzmann machine** (DBM), utilizes both feedforward and feedback processing [6]. Each hidden layer receives input from below, and from above (a kind of prior). Its units learn latent variable representations of the input data. Thus, it can synthesize input data even in the absence of such data [105], which makes DBMs ideal to study hallucinations. Indeed, a DBM to model the genesis of visual hallucinations in **Charles Bonnet Syndrome** [106], an illness characterized by vision loss accompanied by visual hallucinations. They showed that, in response to low-level impairments, homeostatic mechanisms stabilize network activity levels, such that hallucinations arise when input is lacking, consistent with the observation that de-afferented cortex becomes hyper excitable – we argue this reflects the imposition of explanatory priors on noisy inputs. They assayed the role of hierarchical structure. A DBM trained without the topmost layer failed to learn a generative model and thus failed to show complex hallucinations; hence, no high-level priors, no hallucinations (see Figure 2). They speculate that the balance between feedforward and feedback, between priors and prediction

errors, is mediated by acetylcholine [107–109]. Intriguingly, administering scopolamine, a muscarinic cholinergic antagonist, increases conditioned hallucinations in healthy volunteers [31].

These synthetic examples of aberrant perceptual inference fit comfortably with the predictive coding formulation - in the sense that most psychedelics act upon serotonergic neuromodulatory receptors in deep pyramidal cells. It is these cells that are thought to encode Bayesian representations that generate top-down predictions. Having said this, models of perceptual hallucinosis and false inference may not be apt to explain the hallucinations seen in psychosis. This brings us back to the distinction between perceptual and active inference. Our explanation for hallucinations above rests explicitly on beliefs about agency and how “I cause sensory consequences”. In short, to develop a formal understanding of the sorts of hallucinations in conditions like schizophrenia, one might suppose that deep priors about how we actively generate our sensorium lie at the heart of auditory hallucinations, which are generally comorbid with delusions of control and other false concepts (see Box 1). This has the important implication that subsequent studies of hallucination may focus not so much on prior beliefs about the exteroceptive world, but on how we infer the consequences of our action across all (exteroceptive, interoceptive and proprioceptive) domains. This is particularly prescient given the engagement of the anterior insula and other frontal regions in hallucinations.

These computational models provide much needed support and constraint in what we mean when we evoke hierarchical layers of processing. It is clear that modulating the function of high and low levels of the hierarchy have different consequences for the presence and content of hallucinations in network models. We observe similar hierarchical organization of beliefs relevant to hallucinations in computational analyses of human behavior [34]. In neuroimaging data, high-level predictions about multi-sensory learned associations correlated with activity in supramodal regions such as orbitofrontal cortex and hippocampus [110], whilst dynamic low-level predictions were associated with activity in primary cortices [110]. Both levels were implicated in hallucinatory perception, but clinical AVH were associated with dysfunction in higher-level neural and behavioral responses [34]. The explanatory power of hierarchical differences – rather than being a salve for data that challenge the theory - appears to have empirical reality.

Concluding Remarks

Arguing whether hallucinations are driven by high- or low-precision prior beliefs may seem arcane. However, we believe it has profound clinical implications. If priors are dominant, then the neurobiological [34] and psychological [102] interventions indicated would be different from if they are imprecise. At the very least, the contemporary data suggest a precision-weighted trade-off for the contents of experience, in higher versus lower hierarchical layers and more or less agentic hierarchies, which mediate the different contents and complexities of hallucinations. More broadly, the observation from our laboratories, that people who do not hallucinate in their everyday lives can nevertheless evince hallucinations in the laboratory, favors the active inference model of perception and agency. We hope that this understanding of hallucinations as an exaggeration of normal non-hallucinatory

perception, to which we are all sometimes prone, may help further our understanding of perception and cognition (see Box 2) and go some way to de-stigmatize clinical hallucinations and encourage empathy for those who experience them.

Acknowledgements:

PRC was supported by NIMH R01MH112887, NIMH R21MH116258 as well as a Rising Star award from the International Mental Health Research Organization, the Clinical Neurosciences Division, U.S. Department of Veterans Affairs, and the National Center for Post-Traumatic Stress Disorders, VACHS, West Haven, CT, USA. BAD is supported by the Wellcome Trust (WT108720). KS is supported by a Research Fellowship of Leopoldina – German National Academy of Sciences (LPDS 2018–03). PCF is funded by the Wellcome Trust and the Bernard Wolfe health Neuroscience Fund. GH is supported by NIMH R01MH114965.

ARP is supported by a NARSAD Young Investigator Award from the Brain and Behavior Research Foundation, a K23 Career Development Award from the National Institute of Mental Health (K23 MH115252–01A1), a Career Award for Medical Scientists from the Burroughs-Wellcome Fund, and by the Yale University School of Medicine and Department of Psychiatry. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official view of NIH or the CMHC/DMHAS. The authors report no relevant biomedical conflicts of interest.

GLOSSARY

Agency
The inference that one has acted to effect some particular result

Allocentric
Having ones' interest and attention directed externally (often toward other persons)

Charles-Bonnet Syndrome
Named after a Swiss Philosopher, who wrote about his Grandfather's experiences losing his sight to cataracts. Common amongst those who have lost their sight, this condition causes visual hallucinations. The sight loss can also result from macular degeneration, glaucoma or diabetic eye disease

Corollary Discharge
The command sent to the muscles to produce a movement, directed to other regions of the brain to inform them of an impending movement

Deep-Boltzman Machine
A type of generative neural network (see below) that can learn internal representations

Delusions
Fixed, false beliefs that persist despite overwhelming contradictory evidence. Often held with considerable preoccupation and distress

Egocentric
Having one's interest and attention directed toward the self (and away from others)

Exteroceptive
Relating to stimuli that are external to the organism

Hallucinations

Percepts without corresponding external stimulus

Illusions
Something likely to be mis-perceived by the sensory apparatus. Typically, an erroneous conclusion about something actually present

Interoceptive
Relating to stimuli that are internal to the organism

Neural Network
A computing system vaguely inspired by nervous systems, based on a collection of connected units or nodes which loosely model neurons

Precision
A property of priors and prediction errors (see below). Formalized as the inverse variance or negative entropy. Precision underwrites notions like certainty, confidence, salience and reliability

Prediction Error
The mismatch of expectancy and experience that can, if precise (see above), lead to belief updating

Predictive Coding
A scheme used in signal processing that compresses signals using a predictive model. Predictive processing models of brain function suggest that similar processing efficiencies are employed by the brain. This entails a generative model of expected inputs and a means to broadcast (and learn from) failures in those predictions (prediction errors – see above)

Prior
Ones' beliefs about some quantity or dimension before new evidence is taken into account

Psychosis
Experiences that depart from consensual reality

Voices
Or Auditory Verbal Hallucinations, experiences that someone is talking when no-one is around

References

1. Taine H (1870) De l'intelligence, Librairie Hachette et Cie.
2. Tracy DK and Shergill SS (2013) Mechanisms Underlying Auditory Hallucinations-Understanding Perception without Stimulus. Brain Sci 3 (2), 642–69. [PubMed: 24961419]
3. McCarthy-Jones S and Longden E (2015) Auditory verbal hallucinations in schizophrenia and post-traumatic stress disorder: common phenomenology, common cause, common interventions? Front Psychol 6, 1071. [PubMed: 26283997]
4. Hauf M et al. (2013) Common mechanisms of auditory hallucinations-perfusion studies in epilepsy. Psychiatry Res 211 (3), 268–70. [PubMed: 23154091]

5. Sommer IE et al. (2014) Hearing loss; the neglected risk factor for psychosis. Schizophr Res 158 (1–3), 266–7. [PubMed: 25096542]
6. Salakhutdinov R and Hinton G (2012) An efficient learning procedure for deep Boltzmann machines. Neural Comput 24 (8), 1967–2006. [PubMed: 22509963]
7. Honig A et al. (1998) Auditory hallucinations: a comparison between patients and nonpatients. J Nerv Ment Dis 186 (10), 646–51. [PubMed: 9788642]
8. de Lange FP et al. (2018) How Do Expectations Shape Perception? Trends Cogn Sci 22 (9),764–779. [PubMed: 30122170]
9. Barlow H (1990) Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. Vision Res 30 (11), 1561–71. [PubMed: 2288075]
10. Grush R (2004) The emulation theory of representation: motor control, imagery, and perception. Behav Brain Sci 27 (3), 377–96; discussion 396–442. [PubMed: 15736871]
11. Powers AR, III et al. (2016) Hallucinations as top-down effects on perception. Biol Psychiatry Cogn Neurosci Neuroimaging 1 (5), 393–400. [PubMed: 28626813]
12. Powers AR, 3rd et al. (2017) Varieties of Voice-Hearing: Psychics and the Psychosis Continuum. Schizophr Bull 43 (1), 84–98. [PubMed: 28053132]
13. Shergill SS et al. (1998) Auditory hallucinations: a review of psychological treatments. Schizophr Res 32 (3), 137–50. [PubMed: 9720119]
14. Halligan PW and David AS (2001) Cognitive neuropsychiatry: towards a scientific psychopathology. Nat Rev Neurosci 2 (3), 209–15. [PubMed: 11256082]
15. Friston KJ (2005) Hallucinations and perceptual inference. Behavioral and Brain Science 28 (6), 764–766.
16. Helmholtz H.v. (1867) Handbuch der physiologischen Optik, Voss.
17. Feldman H and Friston KJ (2010) Attention, uncertainty, and free-energy. Front Hum Neurosci 4, 215. [PubMed: 21160551]
18. Friston K and Kiebel S (2009) Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci 364 (1521), 1211–21. [PubMed: 19528002]
19. Friston KJ and Stephan KE (2007) Free-energy and the brain. Synthese 159 (3), 417–458. [PubMed: 19325932]
20. Teufel C et al. (2013) The role of priors in Bayesian models of perception. Front Comput Neurosci 7, 25. [PubMed: 23565091]
21. Seashore CE (1895) MEASUREMENTS OF ILLUSIONS AND HALLUCINATIONS IN NORMAL LIFE. Studies from the Yale Psychological Laboratory 3.
22. Barber TX and Calverley DS (1964) An Experimental Study of "Hypnotic" (Auditory and Visual) Hallucinations. J Abnorm Psychol 68, 13–20. [PubMed: 14105174]
23. Mintz S and Alpert M (1972) Imagery vividness, reality testing, and schizophrenic hallucinations. J Abnorm Psychol 79 (3), 310–6. [PubMed: 5033372]
24. Pavlov IP (1928) Natural science and the brain In ectures on conditioned reflexes: Twenty-five years of objective study of the higher nervous activity (behaviour) of animals, Liverwright Publishing Corporation.
25. Ellson DG (1941) Hallucinations produced by sensory conditioning. Journal of Experimental Psychology.28 (1), pp.
26. Davies P et al. (1982) An effective paradigm for conditioning visual perception in human subjects. Perception 11 (6), 663–9. [PubMed: 7186617]
27. Kot T and Serper M (2002) Increased susceptibility to auditory conditioning in hallucinating schizophrenic patients: a preliminary investigation. J Nerv Ment Dis 190 (5), 282–8. [PubMed: 12011606]
28. Teufel C et al. (2015) Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc Natl Acad Sci U S A.
29. Alderson-Day B et al. (2017) Distinct processing of ambiguous speech in people with non-clinical auditory verbal hallucinations. Brain 140 (9), 2475–2489. [PubMed: 29050393]
30. Ellson DG, Hallucinations produced by sensory conditioning, Journal of Experimental Psychology, American Psychological Association, 1941, p. 1.

31. Warburton DM et al. (1985) Scopolamine and the sensory conditioning of hallucinations. Neuropsychobiology 14 (4), 198–202. [PubMed: 3835496]
32. Agathon M and Roussel A (1973) [Use of a sensory conditioning test in patients treated with psychotropic drugs]. Int Pharmacopsychiatry 8 (4), 221–33. [PubMed: 4773019]
33. Brogden WJ (1947) Sensory pre-conditioning of human subjects. J Exp Psychol 37 (6), 527–39. [PubMed: 18920626]
34. Powers AR et al. (2017) Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. Science 357 (6351), 596–600. [PubMed: 28798131]
35. Jardri R et al. (2011) Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. Am J Psychiatry 168 (1), 73–81. [PubMed: 20952459]
36. Jauhar S et al. (2017) A Test of the Transdiagnostic Dopamine Hypothesis of Psychosis Using Positron Emission Tomographic Imaging in Bipolar Affective Disorder and Schizophrenia. JAMA Psychiatry 74 (12), 1206–1213. [PubMed: 29049482]
37. Cassidy CM et al. (2018) A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. Curr Biol.
38. O'Callaghan C et al. (2017) Visual Hallucinations Are Characterized by Impaired Sensory Evidence Accumulation: Insights From Hierarchical Drift Diffusion Modeling in Parkinson's Disease. Biol Psychiatry Cogn Neurosci Neuroimaging 2 (8), 680–688. [PubMed: 29560902]
39. Factor SA et al. (2017) The role of neurotransmitters in the development of Parkinson's disease-related psychosis. Eur J Neurol 24 (10), 1244–1254. [PubMed: 28758318]
40. Crapse TB and Sommer MA (2008) Corollary discharge circuits in the primate brain. Curr Opin Neurobiol 18 (6), 552–7. [PubMed: 18848626]
41. Friston KJ et al. (2010) Action and behavior: a free-energy formulation. Biol Cybern 102 (3), 227–60. [PubMed: 20148260]
42. Geisler WS and Kersten D (2002) Illusions, perception and Bayes. Nat Neurosci 5 (6), 508–10. [PubMed: 12037517]
43. Weiss Y et al. (2002) Motion illusions as optimal percepts. Nat Neurosci 5 (6), 598–604. [PubMed: 12021763]
44. Dakin S et al. (2005) Weak suppression of visual context in chronic schizophrenia. Curr Biol 15 (20), R822–4. [PubMed: 16243017]
45. Koethe D et al. (2006) Disturbances of visual information processing in early states of psychosis and experimental delta-9-tetrahydrocannabinol altered states of consciousness. Schizophr Res 88 (1–3), 142–50. [PubMed: 17005373]
46. Norton D et al. (2008) Altered 'three-flash' illusion in response to two light pulses in schizophrenia. Schizophr Res 103 (1–3), 275–82. [PubMed: 18423984]
47. Adams RA et al. (2013) The computational anatomy of psychosis. Frontiers in psychiatry 4, 47. [PubMed: 23750138]
48. Heinks-Maldonado TH et al. (2007) Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. Arch Gen Psychiatry 64 (3), 286–96. [PubMed: 17339517]
49. Ford JM and Mathalon DH (2005) Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? Int J Psychophysiol 58 (2–3), 179–89. [PubMed: 16137779]
50. Ford JM et al. (2007) Synch before you speak: auditory hallucinations in schizophrenia. Am J Psychiatry 164 (3), 458–66. [PubMed: 17329471]
51. Frith C (2005) The neural basis of hallucinations and delusions. C R Biol 328 (2), 169–75. [PubMed: 15771003]
52. Rosler L et al. (2015) Failure to use corollary discharge to remap visual target locations is associated with psychotic symptom severity in schizophrenia. J Neurophysiol 114 (2), 1129–36. [PubMed: 26108951]
53. Thakkar KN et al. (2017) Oculomotor Prediction: A Window into the Psychotic Mind. Trends Cogn Sci 21 (5), 344–356. [PubMed: 28292639]
54. Thakkar KN et al. (2015) Disrupted Saccadic Corollary Discharge in Schizophrenia. J Neurosci 35 (27), 9935–45. [PubMed: 26156994]

55. Shergill SS et al. (2014) Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. JAMA Psychiatry 71 (1), 28–35. [PubMed: 24196370]
56. Canal-Bruland R et al. (2015) Size estimates of action-relevant space remain invariant in the face of systematic changes to postural stability and arousal. Conscious Cogn 34, 98–103. [PubMed: 25913547]
57. Badcock JC (2010) The cognitive neuropsychology of auditory hallucinations: a parallel auditory pathways framework. Schizophr Bull 36 (3), 576–84. [PubMed: 18835839]
58. Ylinen S et al. (2015) Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. Cereb Cortex 25 (6), 1576–86. [PubMed: 24414279]
59. Griffin JD and Fletcher PC (2017) Predictive Processing, Source Monitoring, and Psychosis. Annu Rev Clin Psychol 13, 265–289. [PubMed: 28375719]
60. Rohe T and Noppeney U (2018) Reliability-Weighted Integration of Audiovisual Signals Can Be Modulated by Top-down Attention. eNeuro 5 (1).
61. Rohe T and Noppeney U (2016) Distinct Computational Principles Govern Multisensory Integration in Primary Sensory and Association Cortices. Curr Biol 26 (4), 509–14. [PubMed: 26853368]
62. Rohe T and Noppeney U (2015) Sensory reliability shapes perceptual inference via two mechanisms. J Vis 15 (5), 22.
63. Rohe T and Noppeney U (2015) Cortical hierarchies perform Bayesian causal inference in multisensory perception. PLoS Biol 13 (2), e1002073. [PubMed: 25710328]
64. Daw ND et al. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8 (12), 1704–11. [PubMed: 16286932]
65. Botvinick M and Cohen J (1998) Rubber hands ‘feel’ touch that eyes see. Nature 391 (6669), 756. [PubMed: 9486643]
66. Thakkar KN et al. (2011) Disturbances in body ownership in schizophrenia: evidence from the rubber hand illusion and case study of a spontaneous out-of-body experience. PLoS One 6 (10), e27089. [PubMed: 22073126]
67. Samad M et al. (2015) Perception of body ownership is driven by Bayesian sensory in ference.PLoS One 10 (2), e0117178. [PubMed: 25658822]
68. Ehrsson HH et al. (2007) Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. Proc Natl Acad Sci U S A 104 (23), 9828–33. [PubMed: 17517605]
69. Stevenson RA et al. (2017) The associations between multisensory temporal processing and symptoms of schizophrenia. Schizophr Res 179, 97–103. [PubMed: 27746052]
70. Suzuki K et al. (2013) Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. Neuropsychologia 51 (13), 2909–17. [PubMed: 23993906]
71. Sterzer P et al. (2016) Thought Insertion as a Self-Disturbance: An Integration of Predictive Coding and Phenomenological Approaches. Front Hum Neurosci 10, 502. [PubMed: 27785123]
72. Frith C (1999) How hallucinations make themselves heard. Neuron 22 (3), 414–5. [PubMed: 10197519]
73. Moore JW et al. (2011) Sense of agency, associative learning, and schizotypy. Conscious Cogn 20 (3), 792–800. [PubMed: 21295497]
74. Moore JW et al. (2011) Ketamine administration in healthy volunteers reproduces aberrant agency experiences associated with schizophrenia. Cogn Neuropsychiatry 16 (4), 364–81. [PubMed: 21302161]
75. Moore JW and Fletcher PC (2012) Sense of agency in health and disease: a review of cue integration approaches. Conscious Cogn 21 (1), 59–68. [PubMed: 21920777]
76. Hauser M et al. (2011) Sense of agency is altered in patients with a putative psychotic prodrome. Schizophr Res 126 (1–3), 20–7. [PubMed: 21112189]
77. Voss M et al. (2010) Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. Brain 133 (10), 3104–12. [PubMed: 20685805]

78. Pezzulo G et al. (2015) Active Inference, homeostatic regulation and adaptive behavioural control. Prog Neurobiol 134, 17–35. [PubMed: 26365173]
79. Apps MA and Tsakiris M (2014) The free-energy self: a predictive coding account of self-recognition. Neurosci Biobehav Rev 41, 85–97. [PubMed: 23416066]
80. Zmigrod L et al. (2016) The neural mechanisms of hallucinations: A quantitative meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 69, 113–23. [PubMed: 27473935]
81. Moerel M et al. (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J Neurosci 32 (41), 14205–16. [PubMed: 23055490]
82. Wilkinson S and Bell V (2016) The Representation of Agents in Auditory Verbal Hallucinations. Mind Lang 31 (1), 104–126. [PubMed: 26900201]
83. Hirsh JB, Mar RA, Peterson JB (2013) Personal narratives as the highest level of cognitive integration. Behav Brain Sci 36 (3).
84. Woods A et al. (2015) Experiences of hearing voices: analysis of a novel phenomenological survey. Lancet Psychiatry 2 (4), 323–31. [PubMed: 26360085]
85. Fletcher PC and Frith CD (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 10 (1), 48–58. [PubMed: 19050712]
86. Sokolov A (1972) Inner Speech and Thought, Plenum Press.
87. Robbins TW (2017) Animal Models of Hallucinations Observed Through the Modern Lens. Schizophr Bull 43 (1), 24–26. [PubMed: 28053128]
88. Konorski J (1967) Integrative Activity of the Brain, University of Chicago Press.
89. Teufel C et al. (2015) Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc Natl Acad Sci U S A 112 (43), 13401–6. [PubMed: 26460044]
90. Holland PC (1981) Acquisition of representation-mediated conditioned food aversions. Learning and Motivation 12 (1), 1–18.
91. Holland PC (2005) Amount of training effects in representation-mediated food aversion learning: no evidence of a role for associability changes. Learn Behav 33 (4), 464–78. [PubMed: 16573217]
92. McDannald MA et al. (2011) Impaired reality testing in an animal model of schizophrenia. Biol Psychiatry 70 (12), 1122–6. [PubMed: 21798517]
93. Busquets-Garcia A et al. (2017) Representation-mediated Aversion as a Model to Study Psychotic-like States in Mice. Bio Protoc 7 (12).
94. Langdon AJ et al. (2018) Model-based predictions for dopamine. Curr Opin Neurobiol 49, 1–7. [PubMed: 29096115]
95. Takahashi YK et al. (2017) Dopamine Neurons Respond to Errors in the Prediction of Sensory Features of Expected Rewards. Neuron 95 (6), 1395–1405 e3. [PubMed: 28910622]
96. Sharpe MJ et al. (2017) Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nat Neurosci 20 (5), 735–742. [PubMed: 28368385]
97. Whitson JA and Galinsky AD (2008) Lacking control increases illusory pattern perception. Science 322 (5898), 115–7. [PubMed: 18832647]
98. Fiorillo CD et al. (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299 (5614), 1898–902. [PubMed: 12649484]
99. Mondragon E et al. (2017) Associative Learning Should Go Deep. Trends Cogn Sci 21 (11), 822–825. [PubMed: 28668210]
100. Versace E et al. (2018) Priors in Animal and Artificial Intelligence: Where Does Learning Begin? Trends Cogn Sci.
101. Karmiloff-Smith A (2006) The tortuous route from genes to behavior: A neuroconstructivist approach. Cognitive, affective & behavioral neuroscience 6 (1), 9–17.
102. Powers AR, Bien C, Corlett PR (2018) Hearing their voices: Aligning Computational Psychiatry with the Hearing Voices Movement. JAMA Psychiatry In Press.
103. Hoffman RE and Dobscha SK (1989) Cortical pruning and the development of schizophrenia: a computer model. Schizophr Bull 15 (3), 477–90. [PubMed: 2814376]
104. LeCun Y et al. (2015) Deep learning. Nature 521 (7553), 436–44. [PubMed: 26017442]

105. Yuille A and Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? Trends Cogn Sci 10 (7), 301–8. [PubMed: 16784882]
106. Reichert DP et al. (2013) Charles Bonnet syndrome: evidence for a generative model in the cortex? PLoS Comput Biol 9 (7), e1003134. [PubMed: 23874177]
107. Hinton GE and Dayan P (1996) Varieties of Helmholtz Machine. Neural Netw 9 (8), 1385–1403. [PubMed: 12662541]
108. Dayan P et al. (1995) The Helmholtz machine. Neural Comput 7 (5), 889–904. [PubMed: 7584891]
109. Hinton GE et al. (1995) The “wake-sleep” algorithm for unsupervised neural networks. Science 268 (5214), 1158–61. [PubMed: 7761831]
110. Weilnhammer V et al. (2018) The Neural Correlates of Hierarchical Predictions for Perceptual Decisions. J Neurosci.
111. Schmack K et al. (2013) Delusions and the role of beliefs in perceptual inference. J Neurosci 33 (34), 13701–12. [PubMed: 23966692]
112. Schmack K et al. (2015) Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. Schizophr Res Cogn 2 (2), 72–77. [PubMed: 29114455]
113. Stuke H et al. (2018) Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. Schizophr Bull.
114. Schmack K et al. (2017) Enhanced predictive signalling in schizophrenia. Hum Brain Mapp 38 (4), 1767–1779. [PubMed: 28097738]
115. Corlett PR et al. (2009) From drugs to deprivation: a Bayesian framework for understanding models of psychosis. Psychopharmacology (Berl) 206 (4), 515–30. [PubMed: 19475401]
116. Liddle PF (1992) Syndromes of schizophrenia on factor analysis. Br J Psychiatry 161, 861.
117. Minas IH et al. (1992) Positive and negative symptoms in the psychoses: multidimensional scaling of SAPS and SANS items. Schizophr Res 8 (2), 143–56. [PubMed: 1457393]
118. Davies DJ et al. (2017) Anomalous Perceptions and Beliefs Are Associated With Shifts Toward Different Types of Prior Knowledge in Perceptual Inference. Schizophr Bull.
119. Kwisthout J et al. (2017) To be precise, the details don’t matter: On predictive processing, precision, and level of detail of predictions. Brain Cogn 112, 84–91. [PubMed: 27114040]
120. Yang GJ et al. (2016) Functional hierarchy underlies preferential connectivity disturbances in schizophrenia. Proc Natl Acad Sci U S A 113 (2), E219–28. [PubMed: 26699491]
121. Jardri R and Deneve S (2013) Circular inferences in schizophrenia. Brain 136 (Pt 11), 3227–41. [PubMed: 24065721]
122. Murray JD et al. (2014) Linking microcircuit dysfunction to cognitive impairment: effects of disinhibition associated with schizophrenia in a cortical working memory model. Cereb Cortex 24 (4), 859–72. [PubMed: 23203979]
123. Anticevic A et al. (2012) NMDA receptor function in large-scale anticorrelated neural systems with implications for cognition and schizophrenia. Proc Natl Acad Sci U S A 109 (41), 16720–5. [PubMed: 23012427]
124. Corlett PR et al. (2016) Prediction error, ketamine and psychosis: An updated model. J Psychopharmacol 30 (11), 1145–1155. [PubMed: 27226342]
125. Powers AR, 3rd et al. (2015) Ketamine-Induced Hallucinations. Psychopathology 48 (6), 376–85. [PubMed: 26361209]
126. Firestone C and Scholl BJ (2015) Cognition does not affect perception: Evaluating the evidence for ‘top-down’ effects. Behav Brain Sci, 1–77.
127. Fodor JA (1983) The modularity of mind : an essay on faculty psychology, MIT Press.
128. Fotopoulou A (2014) Time to get rid of the ‘Modular’ in neuropsychology: a unified theory of anosognosia as aberrant predictive coding. J Neuropsychol 8 (1), 1–19. [PubMed: 23469983]
129. Quine WV and Quine WV (1951) Two Dogmas of Empiricism. Philosophical Review 60.
130. McKay RT and Dennett DC (2009) The evolution of misbelief. Behav Brain Sci 32 (6), 493–510; discussion 510–61. [PubMed: 20105353]

131. Johnson DD and Fowler JH (2011) The evolution of overconfidence. Nature 477 (7364), 317–20. [PubMed: 21921915]

132. Firestone C and Scholl BJ (2016) Cognition does not affect perception: Evaluating the evidence for “top-down” effects. Behav Brain Sci 39, e229. [PubMed: 26189677]

133. Pylyshyn Z (1999) Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. Behav Brain Sci 22 (3), 341–65; discussion 366–423. [PubMed: 11301517]

134. Macpherson F (2017) The relationship between cognitive penetration and predictive coding. Consciousness and Cognition: An International Journal, 6–16.

135. Suzuki K et al. (2017) A Deep-Dream Virtual Reality Platform for Studying Altered Perceptual Phenomenology. Sci Rep 7 (1), 15982. [PubMed: 29167538]

Box 1. Halucinations & Delusions – commonalities, differences

Delusions and hallucinations tend to co-occur though not always. In this computational psychiatry literature, they have been dissociated. Whereas hallucinations have been consistently related to more precise priors, delusions have been associated with less precise priors.

For instance, examining the relation between priors and delusional experiences with ambiguous visual motion stimuli reveals that delusion-proneness both in healthy individuals as well as in schizophrenia patients was related to weaker statistical learning of new priors [111–113]. However, delusion-proneness both in healthy individuals as well as in schizophrenia patients is correlated with high-precision priors induced by a high-level cognitive manipulation (a suggestion that 3D glasses will shift their perception of the ambiguous motion [111, 114]). These apparently contradictory observations might be reconciled by assuming a hierarchy in the brain's predictive machinery: Here, imprecise perceptual priors (or failure to attenuate sensory precision) may lead to the formation of delusional beliefs. At the same time, precise cognitive priors at a higher hierarchical level will sculpt perception, subtending hallucinations and the maintenance of delusions (by engendering delusion consistent percepts [115]).

This comorbidity of hallucinations and delusions is consistent with the factor structure of psychotic symptoms; a 3-factor solution: Negative symptoms, Disorganization and Positive Symptoms (comprising hallucinations *and* delusions) [116]. However, there may be structure within the positive symptom factor [117]. Hallucinations may co-occur more frequently with delusions of parasitosis (that there are insects on the skin) and less frequently with grandiose ideas, religious and referential delusions [117]. Might there be psychological and biological evidence in favour of such relationships between symptoms?

A recent study probed priors for the gist vs. the details of ambiguous images in a healthy population with varying degrees of hallucination- and delusion-proneness [118]. Hallucination-proneness

The precision-weighting of priors and prediction errors may be independent at different hierarchical levels and in different modalities [18]. Neurobiologically, this may be mediated by the higher density of recurrent connections in higher-level association cortices, compared with primary sensory regions, such that a psychotogenic perturbation that impacts excitatory/inhibitory (E/I) balance may have more profound effects higher rather than lower in the hierarchy [120]. See [121] for a detailed exposition of the role of E/I balance in learning, inference and psychosis). In brief, the E/I relationships may implement exactly the precision weighting mechanisms that underlie predictive coding. Blocking NMDA receptors (with ketamine for example) profoundly alters E/I balance [122, 123], thus altering the balance between priors and prediction errors [115], perhaps differently at different hierarchical levels [120]. NMDA blockade would decrease the precision of priors facilitating delusion-like inferences [124]. However, ketamine does not typically engender hallucinations (except perhaps in circumstances of high environmental uncertainty[125]). Hallucinations (and sustained delusions) would entail

neuroplastic learned changes in priors which may arise from dopamine driven learning mechanisms or serotonergic effects on deep pyramidal cells[115]. These effects are not characteristic of acute ketamine. But they may arise with chronic administration of ketamine or amphetamine or the administration of serotonergic hallucinogens[115]. Comparing and contrasting the different phenomenologies of these drugs[115] as well as psychotic symptoms across different illnesses will be a key test of our hypothesis. We suspect that in each case, where there are hallucinations, there will be high-precision priors.

Box 2. Cognitive penetration

Many cognitive scientists have argued that cognition does not influence perception [126], that perception and cognition are separate modules. We believe the hallucinations data outlined presently are inconsistent with such encapsulated modularity. This is because learned beliefs about the task — ones that were not present within the perceptual module prior to the experiment— can be conditioned and appear to change perception.

In *The Modularity of Mind* (1983), Jerry Fodor sketched a blueprint of mental architecture comprised of modules—systems that process a single specific kind of information [127], strictly segregated into discrete mental faculties that can be damaged in isolation [128]. An encapsulated perceptual system, kept separate from the influence of beliefs, could have the advantage of keeping our beliefs grounded in the truth offered by our senses [129]. However, a cognitively penetrable perceptual apparatus may be equally adaptive, despite misperceiving, as long as the resulting behavior is adaptive [130, 131].

Staunch modularists might allege that the top-down effects of cognition on perception are merely an effect of attention [132]. This seems hard to reconcile with the new percepts we appear to be able to condition. Furthermore, they may allow for the type of learning we describe by suggesting that new priors that are learned in our tasks are compiled into the perceptual module with experience [133]. This seems like an influence of cognition on perception to us. Predictive perception need not necessarily demand penetration [134], but penetration of perception by cognition, exemplified by hallucinations, is at least consistent with hierarchical versions of predictive processing.

Outstanding Questions

- Are corollary discharge and strong prior effects anti-correlated, as our account would predict?
- Is the sensitivity to conditioned hallucinations in people who hallucinate domain specific? That is, are people with AVH also more susceptible to visual conditioned hallucinations?
- What are the neurochemical mechanisms of strong priors? Can they be remediated with pharmacological manipulations?
- Are the contemporary neural data a cause, a consequence or a correlate of hallucinations? We will need causal manipulations, like transcranial magnetic stimulation and effective connectivity analyses to make causal and directional claims about the neurobiology of priors and hallucinations.
- Does this account apply to hallucinations trans-diagnostically (i.e. outside of the context of schizophreniform illness, to hallucinations in Alzheimer's, Parkinson's, Charles Bonnet Syndrome, head injury and stroke)?

Highlights

- Recent data establish a role for strong prior beliefs in the genesis of hallucinations. These data are difficult to reconcile with aberrant inner-speech theories, in which *weaker* predictions about the potential consequences of ones' own inner-speech drive an inference that speech is emanating from an agent external to oneself.
- In the predictive-coding view, this failure of self-prediction renders the consequences of one's inner speech surprising. The prediction errors induced are explained away by the strong higher-level priors identified in recent work. The presence and contents of hallucinations can be understood in terms of learning, inference and a reliability-based trade-off between internal and external information sources, biased toward high-level priors.
- If priors divorce perception from sensation somewhat, then the distinction between hallucination and perception becomes less clear. We hope this explanation renders hallucinations more understandable and less stigmatizing.

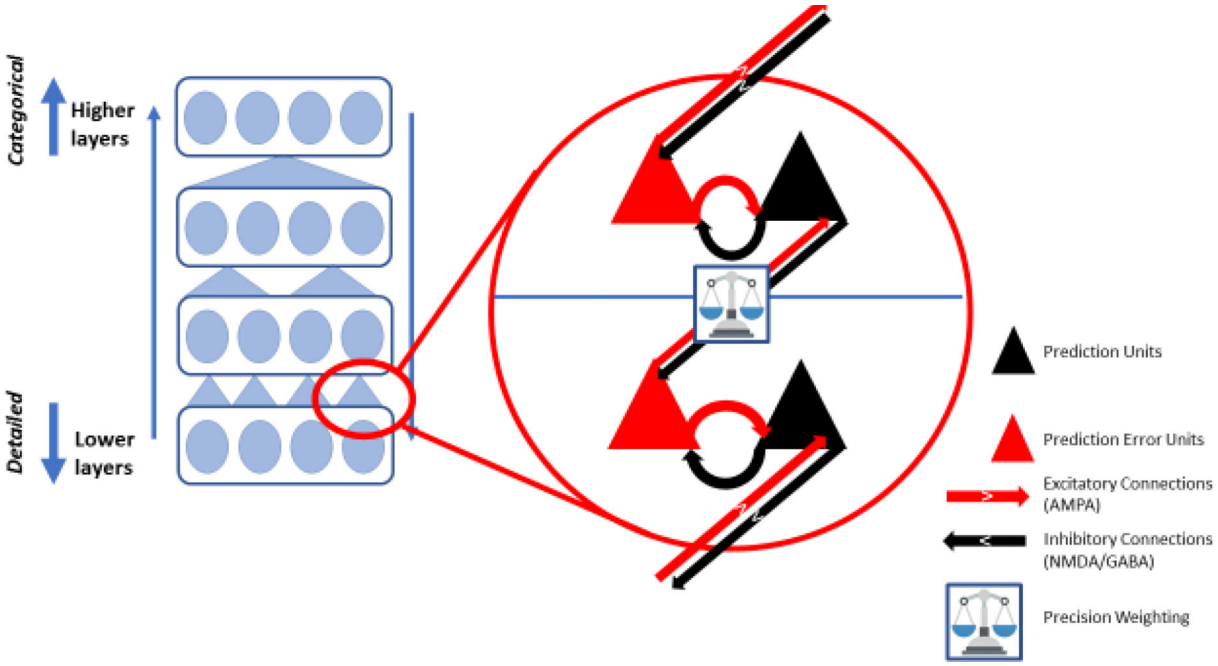

Figure 1. An Inferential Hierarchy.

Here we sketch the hierarchical message passing thought to underlie predictive coding and expand on the details of between- and within-layer computations. Sensory input is conveyed via ascending prediction errors in superficial pyramidal cells (red arrows, posited to be mediated by fast AMPA glutamate receptor signaling). Posterior expectations are encoded by the activity of deep pyramidal cells. These cells then provide descending priors (black arrows, mediated by slower and more diffuse NMDA and GABA signaling) that inform prediction errors at the lower level – instantiating the computations that remove the expected input, leaving prediction errors to be assimilated or accommodated, depending on their precision (depicted by the balance, hypothesized to be implemented via slower neuromodulators like dopamine, acetylcholine and serotonin, depending on the particular inferential hierarchy). Lateral interactions (horizontal arrows) mediated within-layer predictions about the precision of priors (black) and prediction errors (red).

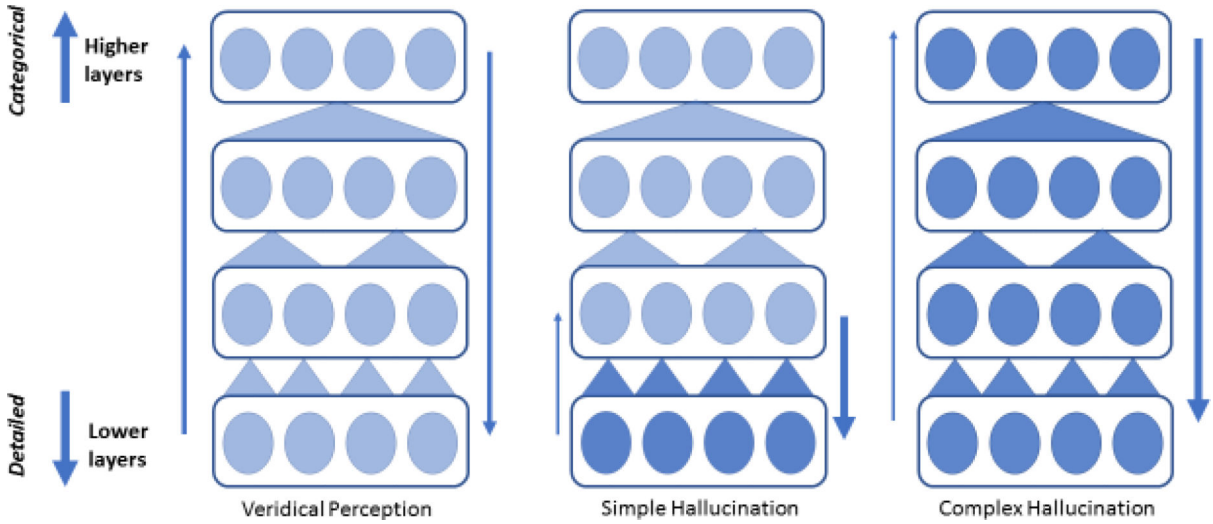

Figure 2. How Hierarchies Might Hallucinate

Problems lower in a hierarchy may generate simple hallucinations. Problems that pervade higher levels will engender more complex experiences. Consistent with the empirical data from humans viewing images generated from deep-networks perturbed at low and high levels [135]and DBM models of perception where in hallucinations only arise under the influence of the highest-level generative model [106]. Figure adapted from [135], licensed under a Creative Commons 4.0 International License (http://creativecommons.org/licenses/by/4.0/.)