Illusions, Delusions, and Your Backwards Bayesian Brain: A Biased Visual Perspective

Richard T. Born a Gianluca M. Bencomo b

a Department of Neurobiology, Harvard Medical School, Boston, MA, USA; b Department of Computer Science, Whittier College, Whittier, CA, USA

Abstract
The retinal image is insufficient for determining what is “out there,” because many different real-world geometries could produce any given retinal image. Thus, the visual system must infer which external cause is most likely, given both the sensory data and prior knowledge that is either innate or learned via interactions with the environment. We will describe a general framework of “hierarchical Bayesian inference” that we and others have used to explore the role of cortico-cortical feedback in the visual system, and we will further argue that this approach to “seeing” makes our visual systems prone to perceptual errors in a variety of different ways. In this deliberately provocative and biased perspective, we argue that the neuromodulator, dopamine, may be a crucial link between neural circuits performing Bayesian inference and the perceptual idiosyncrasies of people with schizophrenia. © 2021 S. Karger AG, Basel

Introduction

Vision as Perceptual Inference

For any given 2-dimensional (2D) projection on the retina (Fig. 1, gray parallelogram), there is an infinitude of possible 3D objects that could have produced that projection, only 2 of which are shown in the figure. Given this insufficiency of the visual stimulus alone, it should not be surprising that visual systems have evolved a number of mechanisms to resolve the ambiguity. For example, most animals, including humans, have 2 eyes that are laterally separated, and a comparison of the corresponding images from the 2 different viewpoints can be used to gain 3D information, a process known as “stereopsis” [Ponce and Born, 2008]. In addition, the comparison of successive “snap-shots” taken by the visual system as either the object or the observer moves through space can also help to disambiguate a scene. This is what happens in the award-winning visual illusion of “impossible motion” [Sugihara, 2010]: the explanation for why the balls seem to roll uphill is revealed as the camera angle is rotated. However, the initial misperception of the geometry of the apparatus and the resulting motion illusion reveals a third important mechanism: the visual system determines what is the most likely 3D cause of the 2D image, by combining
the ambiguous sensory data with prior knowledge. One of the more striking demonstrations of how powerful the influence of priors can be on perception is the so-called “hollow-face illusion” [also known as the “hollow-mask illusion”; Gregory, 1970]. In one version of the illusion, a rotating mask of Charlie Chaplin actually appears to reverse direction when viewed from the concave side [Ray-Omar, 2006], a misperception due to the fact that normal human observers see the concave inside surface of the mask as convex despite both conscious knowledge and low-level visual cues to the contrary. Interestingly, patients with schizophrenia are not as susceptible to this illusion [Dima et al., 2009], a point to which we will return below.

This process of combining sensory evidence with prior knowledge to infer what object or scene is the most likely external cause is what Hermann von Helmholtz [1925] referred to as “unconscious inference.” Precedence for the idea is most commonly attributed to Helmholtz; however, it can be found much earlier in the writings of the Arab polymath, Alhazen (Ibn al-Haytham), in the 10th century [Howard, 1996]. And despite doubts that Helmholtz was himself a Bayesian [Westheimer, 2008], in more recent times the idea has become tightly associated with so-called “Bayesian inference,” because it is through Bayes’ rule for inverting conditional probabilities that the inference is often performed. This topic has received much attention in the literature on vision [Weiss et al., 2002; Lee and Mumford, 2003; Yuille and Kersten, 2006; Fletcher and Frith, 2009; Moreno-Bote et al., 2011], including computational models for performing Bayesian inference with neurons and neural populations [Denève, 2005; Doya et al., 2007; Beck et al., 2008; Ganguli and Simoncelli, 2014; Hafner et al., 2016] and the underappreciated implication this viewpoint has for the important relationship between perception and learning [Barlow, 1990; Fiser et al., 2010].

Given the abundance of literature on the topic, including several excellent reviews [Yuille and Kersten, 2006; Fletcher and Frith, 2009; Fiser et al., 2010], we will not go
into detail here on Bayesian ideas, per se, but rather introduce a slightly different formulation put forward by Notredame et al. [2014] that incorporates the important concept of “predictive coding” into a hierarchical Bayesian scheme (Fig. 2). One key idea here is that prior knowledge is directly compared with the incoming sensory evidence to calculate a “prediction error” [Rao and Ballard, 1999], which is passed on to the next level of the hierarchy, where, after weighting by a factor that reflects the relative reliability of the prior and the likelihood [Ernst and Banks, 2002], it is combined with new prior information from above. The scheme presented in Figure 2 is greatly simplified, and it fails to adequately reflect the iterative and dynamic nature of the process after which one would expect all relevant variables (i.e., prior, prediction error, and posterior) to be represented at each level of the hierarchy with respect to that level’s degree of representational abstraction. We favor this scheme, because it calls for the explicit representation of “prediction error,” for which there is neurophysiological evidence (see below), and further suggests that the calculation involves cortico-cortical feedback. Finally, it allows for optimal combination of top-down and bottom-up information via the Kalman gain, which would naturally vary as a function of the reliability of the different sources of information. In the formula given in Figure 2c, the Kalman gain adjusts the strength of the prediction error, which can be thought of as the new (i.e., unpredicted) sensory evidence, so increasing the gain tips the balance away from prior beliefs and towards sensory evidence.

In the following section, we introduce a circuit motif that is widespread in the cerebral cortex and which could serve to implement Bayes-like calculations. We then describe a mechanism by which the nervous system might dynamically adjust the relative weights of evidence and priors (Kalman gain) using neuromodulators such as dopamine, and, finally, we briefly discuss some interesting sensory phenomena that occur in schizophrenia, a condition in which dysregulation of dopamine is a prominent feature.

**A Circuit for Integrating Top-Down with Bottom-Up Input**

The circuit depicted in Figure 3 emphasizes the role of pyramidal cells (PCs) in layer 5 of the cerebral cortex acting to compare feedforward (bottom-up) information through their basal dendrites with feedback (top-down) information via their apical dendrites, which ramify extensively in layer 1. Under certain circumstances, the layer 5 pyramidal neurons can act as coincidence detectors in which input to the apical dendrites, when paired within a narrow time window with input near the soma, can produce bursts of action potentials, as opposed to single spikes when no distal input is supplied (Fig. 3, Inset) [Cauller, 1995; Larkum et al., 1999, 2001]. It is unclear exactly what role the coincidence-detecting aspect of the

---

*Fig. 3. A circuit for Bayesian calculations. The main player is the layer 5 pyramidal cell (red) [Bekkers 2011] which receives bottom-up inputs via its basal dendrites and top-down inputs onto its apical dendrites, an example of which is shown for the V2-to-V1 projection [Rockland and Virga, 1989]. The question mark indicates the current lack of knowledge concerning how feedback axons might interact with SST neurons, particularly in layer 1 and the superficial part of layer 2. Inset The marked augmentation of PC neuron spiking when it receives near simultaneous inputs on its apical dendrite and cell body [Larkum et al. 1999].*
A basic circuit might account for a number of recent results in our laboratory when we reduced the top-down input to V1 in nonhuman primates (NHP) as well as some interesting similarities with the results of local circuit manipulations in rodent cortex. For example, reversibly inactivating V2/V3 (using “cryoloops” to cool regions of V2 and V3 within the lunate sulcus) of alert macaque monkeys produced reductions in the influence of visual context: surround suppression and end-stopping became significantly weaker [Nassi et al., 2013, 2014]. This result is consistent with the idea, put forward by computational neuroscientists within the framework of “predictive coding” in the visual system, that “prediction error” (in the form of contextual effects like end-stopping) is the result of interactions between top-down and bottom-up information [Rao and Ballard, 1999]. In this case, the top-down prediction from V2 is that most contours in natural images are extended and thus continue through most receptive fields (at least the very small ones present in V1); the spiking of an end-stopped cell in V1 signals the violation of this expectation and thus a “prediction error.” This effect of feedback inactivation on surround suppression was subsequently found in the V1 of other species, including marmosets [Nurminen et al., 2018] and mice [Vangeneugden et al., 2019], and, interestingly, was very similar to the reduction in visual surround suppression observed when SST neurons were optogenetically silenced locally in mouse V1 [Fig. 4b, c; Adesnik et al., 2012]. Further strengthening the possible association between cortico-cortical feedback and SST neurons, we discovered a dramatic reduction in visually evoked gamma oscillations in macaque V1 when inactivating V2/V3 [Hartmann et al., 2019], similar to that observed in mouse V1 with local SST inactivation [Fig. 4d, e; Veit et al., 2017]. A final bit of evidence connecting these results is that the strength of visually evoked gamma rhythms in monkey V1 appears to correlate with the engagement of surround suppression [Gieselmann and Thiele, 2008; Ray and Maunsell, 2011].

From this perspective, layer 1 is particularly interesting and important, as it serves as a nexus for the interactions of long-range sources of information with local cortical circuitry via the apical dendrites of pyramidal neurons and inhibitory interneurons. This layer remains a “crowning mystery” [Hubel, 1982], largely due to the fact that it contains very few neurons and most of its sources of input are remote. For example, approximately 90% of layer 1 of the primary visual cortex (V1) consists of “dark matter”: synapses for which the presynaptic source can not be accounted for by local neurons [Binzegger et al., 2004; Douglas and Martin, 2007]. This is because layer 1 of V1 receives long-range inputs from a multitude of sources, including cortico-cortical feedback from many higher visual areas [Rockland and Virga, 1989; Anderson and Martin, 2009], higher-order thalamic nuclei such as the pulvinar [Sherman and Guillery, 1996], and amygdala [Freese and Amaral, 2005], and neuromodulatory inputs such as those from the basal forebrain cholinergic [Mechawar et al., 2000] and midbrain dopamine [Seamans and Yang, 2004] neurons.

The layer 5 PCs are the principal output neurons of the cortex, sending the results of their calculations to subcortical targets that are often directly involved in producing movement. As an interesting aside, the axons of most, if not all, of these layer 5 PCs send a major collateral to higher-order thalamic nuclei, such as the pulvinar, which in turn send projections back up to the cortex in what may constitute a source of efference copy [Sherman and Guillery, 2011], a signal necessary for the corollary discharge system used to predict the sensory consequences of an animal’s actions (see below) [Crapse and Sommer, 2008; Sommer and Wurtz, 2008]. In the context of a Bayesian calculation, the bottom-up inputs convey the incoming sensory information (the so-called “likelihood”), the top-down inputs provide “prior” knowledge, and the output constitutes the “posterior,” which can be thought of as the probability of some state of the world, given the current sensory evidence. This circuit appears to be a very generic one that operates across different areas of the cortical mantle, allowing prior knowledge (and likely other signals as well) to influence everything from visual perception [Bullier, 2001] to memory retrieval [Tomita et al., 1999] and sensorimotor coordination [Xu et al., 2012].

From this perspective, layer 1 is particularly interesting and important, as it serves as a nexus for the interactions of long-range sources of information with local cortical circuitry via the apical dendrites of pyramidal neurons and inhibitory interneurons. This layer remains a “crowning mystery” [Hubel, 1982], largely due to the fact that it contains very few neurons and most of its sources of input are remote. For example, approximately 90% of layer 1 of the primary visual cortex (V1) consists of “dark matter”: synapses for which the presynaptic source cannot be accounted for by local neurons [Binzegger et al., 2004; Douglas and Martin, 2007]. This is because layer 1 of V1 receives long-range inputs from a multitude of sources, including cortico-cortical feedback from many higher visual areas [Rockland and Virga, 1989; Anderson and Martin, 2009], higher-order thalamic nuclei such as the pulvinar [Sherman and Guillery, 1996], and amygdala [Freese and Amaral, 2005], and neuromodulatory inputs such as those from the basal forebrain cholinergic [Mechawar et al., 2000] and midbrain dopamine [Seamans and Yang, 2004] neurons.
Fig. 4. Similarities between V1 effects seen after inactivation of top-down inputs in the monkey (b, d) and local inactivation of SST neurons in the mouse (c, e). For the experiments shown in b and c, V1 neurons were stimulated with grating patches of varying diameters (a, corresponding to the abscissas in b and c), displaying the well-known phenomenon of surround suppression, which is weaker when V2 is reversibly inactivated (b) or SST neurons are hyperpolarized in V1 (c). For the experiments shown in d and e, V1 neurons were stimulated with large, high-contrast gratings (a, right-most stimulus). The local field potential (LFP) was measured, revealing a pronounced “bump” in the gamma band (∼36 Hz in the monkey and ∼29 Hz in the mouse) under control conditions, which is diminished when V2 is cooled (d) or SST neurons are hyperpolarized (e). Modified from Nassi et al. [2013] (b), Adesnik et al. [2012] (c), Hartmann et al. [2018] (d), and Veit et al. [2017] (e).
Two other results from feedback inactivation experiments are potentially accounted for by the abovementioned fact that the circuit appears to amplify top-down and bottom-up inputs that arrive within a narrow time window (Fig. 3). First, we found that statistical measures of neuronal variability (both irregularity in the timing of individual spikes and trial-to-trial variability in spike counts) were modestly reduced in V1 when V2/V3 were inactivated [Gómez-Laberge et al., 2016]. Second, we observed that the delayed component (~100–200 ms after stimulus onset) of the visually evoked potential (VEP) in V1 was greatly reduced in amplitude during V2/V3 cooling [Hartmann et al., 2019]. We think that a switch from burst mode to single-spiking [Larkum et al., 1999], whether mediated by SST neurons or by a direct effect of feedback on the apical dendrites of layer 5 PCs, could account for the reduced spiking variability [Gómez-Laberge et al., 2016] and might, at least partially, explain feedback’s powerful influence over the volley of recurrent activity that follows the initial feedforward response in the VEP [Hartmann et al., 2019].

Taken together, these results support an important influence of feedback on local computations in the cortex and implicate SST neurons as a possible key circuit element for mediating some of its effects. Even though no evidence exists of a particular preference for feedback axons to make direct synapses onto SST neurons [Gonchar and Burkhalter, 2003], it remains possible that such interactions take place presynaptically in layer 1 or via a multisynaptic pathway.

Striking a Balance between Top-Down and Bottom-Up Information

The sensory world is constantly changing, and this results in variability in the quality of the evidence provided by our senses. To take an obvious but important example for vision, the sun rises and sets each day, creating a wide range of illumination conditions from the high signal-to-noise (SNR) regime of full sunlight at midday, to the low SNR periods during dawn and dusk and the extremely low-signal time of night. Moreover, prior knowledge has a strong learned component [Fiser et al., 2010], so the reliability of much top-down information is also changing over the course of an animal’s lifetime.

Statistical theory tells us that, to achieve “optimal” perception, i.e., to minimize the variability in our estimates of the state-of-the-world, animals should weight different sources of information in proportion to their reliability. And psychophysical experiments have shown that humans can and do integrate sensory information in this optimal way, at least under some circumstances [Ernst and Banks, 2002; Knill and Pouget, 2004; Körding and Wolpert, 2004; French and DeAngelis, 2020]. Thus, from a Bayesian perspective, our circuit (Fig. 3) would seem to require an additional level of flexibility to appropriately balance the influence of priors and sensory information as conditions change.

One candidate for mediating such changes in the balance of top-down and bottom-up inputs is the neuromodulator dopamine. We must emphasize up front that this hypothesis is highly speculative, since dopamine experts dozens of different effects on both pyramidal cells and interneurons acting through several different families of receptors, influencing everything from synaptic release to intrinsic electrical properties [Tritsch and Sabatini, 2012]. For simplicity, we focus here on effects of the intrinsic properties of the layer 5 PCs that are central to our circuit (Fig. 3) and the role of D1 receptors, which are the most highly expressed and widespread dopamine receptors in the rodent prefrontal cortex [Tritsch and Sabatini, 2012]. For our purposes, the most interesting finding is that dopamine appears to have different effects on the apical dendrites and cell body of these neurons (Fig. 5). At the cell body and basal dendrites, D1 receptor activation increases excitability by decreasing K+ currents [Yang and Seamans, 1996; Dong and White, 2003; Dong et al., 2004] as well as augmenting Na+ currents [Fig. 5b; Yang and Seamans, 1996; Gorelova and Yang, 2000]. Conversely, D1 agonists decrease the excitability of apical dendrites by suppressing Ca++ currents that give rise to regenerative high-threshold Ca++ spikes [Yang and Seamans, 1996; Young and Yang, 2004; Kisilevsky et al., 2008; Zhou and Antic, 2012], thus decreasing both the spatial range over which the apical tuft integrates top-down information and its net influence on the cell’s spike-generating mechanism (Fig. 5a). In the context of our Bayesian calculation, increasing levels of dopamine would then serve to progressively tip the balance away from priors and in favor of the sensory evidence. In the formulation put forward in Figure 2, increasing dopamine levels would be represented by an increase in the Kalman gain.

Back to Perception: Illusions and Delusions

It is traditional in psychiatry to distinguish between “false percepts” (hallucinations or illusions) and “false beliefs” (delusions); however, as previously pointed out in the review by Fletcher and Frith [2009], from a Bayesian perspective, the two are not so different (i.e., percepts are really beliefs about the external causes of the internal states of the nervous system) and, in fact, are likely to
Illusions, Delusions, and Your Backwards
Bayesian Brain

share common mechanisms. We will thus not treat them differently here, but rather focus on perceptual studies where the data are clearer and more plentiful.

If we are correct in our contention that what we see at any given moment is not just based on the bottom-up data coming in through our eyes but is also influenced by our top-down expectations, then we might predict that our visual systems would be prone to errors resulting from an imbalance in how these two sources of information are weighted. We have already encountered the powerful hollow-mask illusion, which we would interpret as the result of a very strong prior (“faces are convex”) overwhelming the visual evidence. In fact, a sizeable subset of visual illusions can be interpreted in this way [Weiss et al., 2002; Notredame et al., 2014]; in most instances, the priors are not so overwhelmingly strong (at least for high-contrast, low-noise stimuli), and the resulting illusions are correspondingly less spectacular. However, they are readily measurable and consistent across neurotypical observers.

One large class of visual illusions consists of those where the context in which an object appears causes systematic distortions in its appearance, a classic example of which is the Ebbinghaus illusion (Fig. 6). This particular example involves judgments of size, but the powerful influence of context has been demonstrated in virtually every visual domain, from color (simultaneous contrast; colored shadows) to orientation (tilt illusion), motion perception (Duncker illusion), and many more [Gregory, 1997; Mély et al., 2018]. Not surprisingly, many of these contextual illusions have been linked to extraclassical receptive field effects, such as surround suppression, found throughout the visual systems of NHP and other animals (Fig. 4) [Sériès et al., 2003; Mély et al., 2018].
Schizophrenia and Perception: Linking Dopamine to Bayesian Inference

Multiple lines of evidence, including pharmacology, genetics, and imaging, suggest that dopamine dysregulation is a major component of the pathophysiology of schizophrenia [Howes et al., 2017; Weinstein et al., 2017; Avramopoulos, 2018]. Drugs that block dopamine receptors have long served as primary treatments for this patient population [Stępnicki et al., 2018]. Given our above speculation that dopamine may be involved in adjusting the relative weighting of sensory data and prior information, perceptual studies on people suffering from schizophrenia provide intriguing data.

A number of previous groups have interpreted schizophrenia through a Bayesian lens, focusing particularly on the important role of predictive coding in perceptual inference [Fletcher and Frith, 2009; Brown et al., 2013; Sterzer et al., 2018, 2019; Heinz et al., 2019]. As described above (Fig. 2), the essence of predictive coding is a direct comparison of prior expectations with current sensory evidence, and any differences (the “prediction errors”) serve as spurs to modify internal models [i.e., update priors; Fiser et al., 2010] and can be used to calculate posteriors in hierarchical Bayesian inference schemes [Notredame et al., 2014]. Computational studies have demonstrated that predictive coding models naturally give rise to the phenomenon of “surround suppression” in sensory systems [Rao and Ballard, 1999; Schwartz et al., 2007; Spratling, 2010], and, as already described, animal studies have shown that surround suppression in the visual system is influenced by top-down inputs [Nassi et al., 2013, 2014; Nurminen et al., 2018; Vangeneugden et al., 2019]. It is thus particularly interesting that patients with schizophrenia appear to manifest perceptual signatures of weakened surround suppression (see below) as well as other indications that they may give decreased weight to perceptual priors.

Abnormalities in the neural representation of prior beliefs and sensory information have played an important role in theoretical treatments of the pathophysiology of schizophrenia in particular, and aberrant perceptual inference in general [Friston et al., 2016]. Nevertheless, the underlying mechanisms governing abnormal inference remain unclear. The Bayesian hierarchical predictive coding model described above (Fig. 2) provides a useful framework for understanding how prior beliefs are integrated with bottom-up information at different levels of the visual hierarchy [Felleman and Van Essen, 1991; Sterzer et al., 2018].

Perceptual studies on patients with schizophrenia generally support the idea that abnormally high dopamine levels produce an underweighting of priors. We will briefly describe two lines of evidence that support this notion: (i) the decreased susceptibility to certain visual illusions, and (ii) partial failure to incorporate sensory predictions of one’s own movements, known as “corollary discharge.”

As previously mentioned, a large body of literature supports the conclusion that patients with chronic schizophrenia have a lower sensitivity to expectation-driven visual illusions [Tschacher et al., 2006; Dima et al., 2009; Crawford et al., 2010; Horton and Silverstein, 2011; Keane et al., 2013; review, Notredame et al., 2014]. For example, several groups have found that the Ebbinghaus illusion (Fig. 6) is considerably weaker in people with schizophrenia [Uhlhaas et al., 2004, 2006; Bressan and Kramer, 2013]. And, as an interesting aside, it has also been reported that young children are less susceptible to the Ebbinghaus illusion [Doherty et al., 2010], a finding consistent with the above notion that some priors are learned. In fact, a host of studies has produced evidence that the influence of visual context in general, and of surround suppression (Fig. 4b, c) in particular, are reduced in people with schizophrenia [Dakin et al., 2005; Tadin et al., 2006; Yoon et al., 2009; Robol et al., 2013; Tibber et al., 2013; Yang et al., 2013]. These findings are consistent with both physiological evidence that surround suppression in V1 is influenced by top-down inputs [Nassi et al., 2013, 2014; Nurminen et al., 2018; Vangeneugden et al., 2019] and the biophysically based model suggesting that dopamine acts to diminish this influence [Fig. 5; Yang

Fig. 6. The Ebbinghaus illusion. Both pink disks are actually the same size, but the different contexts cause most neurotypical observers to perceive that on the left as smaller than that on the right.
and Seamans, 1996]. Moreover, this diminution of the influence of top-down information appears to extend to higher-order visual priors, such as that driving the hollow-mask illusion; several studies have reported that observers with schizophrenia are much more likely to report seeing the inside of masks as concave (i.e., veridically) than are neurotypical observers [Schneider et al., 2002; Dima et al., 2009, 2010; Koethe et al., 2009]. In Dima et al. [2009], virtually none of the observers misclassified normal (i.e., convex) faces; however, the neurotypical controls overwhelmingly misperceived concave faces as convex (71 of 72 images, on average) whereas the average patient with schizophrenia misclassified only 4 of the 72 images. This striking difference supports the possibility that dysregulation of dopamine can alter the balance between prior knowledge and sensory evidence, and this in a manner that is consistent with the proposed biophysical model (Fig. 4).

It is important to note that, for the class of visual illusions described above, patients actually perform “better” than neurotypical observers. From an experimental perspective, this is convenient because it helps rule out alternative explanations such as nonspecific impairments in attention, judgment, and cognition that plague many studies in populations with neuropsychopathology. However, we hasten to point out that these illusions are manifestations of mechanisms that are not just important, but actually optimal, for vision under normal conditions [Weiss et al., 2002]; it would therefore be a mistake to conclude that people with schizophrenia have “better vision.” In fact, the persistent inappropriate balance between priors and sensory evidence, when iterated up and down an inferential hierarchy [Jardri and Denève, 2013], can ultimately lead to aberrant perception and, paradoxically, unwarrantedly high levels of confidence [Hemsley and Garety, 1986; Garety et al., 1991; Warman, 2008], resulting in fixed, delusional beliefs [Woodward et al., 2008].

A second line of evidence supporting an underweighting of priors in patients with schizophrenia comes from studies of so-called “corollary discharge” (CD), i.e., a collection of circuits that attempts to account for the sensory consequences of an animal’s own motor actions [Crapse and Sommer, 2008; Sommer and Wurtz, 2008]. From the perspective of visual perception, it is obviously crucial to be able to distinguish sensory data that arises from external sources (“exafference”), such as the retinal motion produced by an approaching predator, from that produced by the movements of the observer or its sense organs (“reafference”), such as the retinal motion produced by movements of the eyes [Donaldson, 2000; Pack, 2014]. In the context of the hierarchical Bayesian inference model (Fig. 2), copies of the self-movement signals, processed through a “forward model” [Miall and Wolpert, 1996], yield a predicted sensory signal (a dynamically generated prior) that is subtracted from the actual sense data. In this case, the resulting prediction error is a sensory signal that is used to perform perceptual inference and, in addition, indicates the probability that the source of the information is external (as opposed to self-generated).

People with schizophrenia have shown deficits in this self-monitoring process across a variety of sensory modalities [Feinberg and Guazzelli, 1999; Pack, 2014]. For instance, when performing a force-matching task, patients experienced less attenuation due to self-generated forces, and out-performed controls in accurately reproducing an externally generated force [Shergill et al., 2005]. Similarly, patients revealed a diminished ability to cancel self-generated, tactile signals in a self-tickling paradigm [Blakemore et al., 2000a, 2000b]. In the visual domain, people with schizophrenia appear to lack the neural signals found in neurotypical observers, indicating discrepancies in visual feedback from performed motor sequences [Leube et al., 2010] and showing higher perceptual thresholds for detecting such incongruent visual feedback [Synofzik et al., 2010]. Further evidence of weakened CD in schizophrenic patients came from studies that found a decreased ability to integrate eye movement-related signals into the perception of visual motion [Taker et al., 1999] and location [Richard et al., 2014]. Interestingly, a failure of CD might also contribute to the auditory hallucinations frequently experienced by people suffering from schizophrenia [Ford and Mathalon, 2005, and references therein]. As noted above, the comparison between experienced sensory stimuli and those predicted from internal models constitutes an important mechanism by which humans and other animals determine which are “self” and which are “other.” This opens the possibility that deficits in the self-attribute of internal speech (i.e., many forms of “thinking”) may be misperceived as external voices.

In sum, a considerable body of evidence suggests that patients with schizophrenia have abnormalities in how they integrate priors with sensory information. We have emphasized the literature supporting “weak priors” in schizophrenia, but readers should be aware that other investigators have argued for an overweighting of priors in schizophrenia [Benrimoh et al., 2018; Cassidy et al., 2018; Corlett et al., 2019]. And still others have adduced evidence of stronger cognitive priors but weaker perceptual
ones [Schmack et al., 2013, 2015]. For lower-level influences, such as contextual visual illusions and corollary discharge, the preponderance of evidence clearly supports a decreased influence of priors in this population. However, one should keep in mind that the same neuromodulator can have different, even opposing, effects on different circuits, the “push/pull” role of dopamine on the striatal “direct/indirect” pathways via different dopamine receptor subtypes being a prime example [Albin et al., 1989; Kravitz et al., 2010]. Therefore, results that seem to be in conflict may be reconciled with a more detailed knowledge of these circuits and how they are modulated.

Closing Remarks and Future Directions
This has been an admittedly biased review of several different bodies of literature, perhaps illustrating the pitfalls inherent in overly strong priors. Our aim from the start was to be provocative and, whether the ideas presented here are absolutely correct in their detail is less important than the dialogue and future studies that we are hoping to inspire. Besides, it takes only one additional inhibitory interneuron intercalated into a circuit to completely invert the sign of a predicted effect or influence! And, as noted above, simply changing the subtype of a neuromodulator’s receptor can lead to very different effects on the same circuit.

In this spirit, we close with a few thoughts on several specific areas that we think merit deeper investigation. First, at the circuit level, the mechanisms by which top-down information interacts with local circuits remain largely unknown, exacerbated by the fact that many of these interactions take place in layer 1. While modern approaches using serial-section electron microscopy (EM) have begun to flesh out the details of local circuits [Morgan and Lichtman, 2013], layer 1 has not been amenable to traditional EM-based connectomics, because, as previously noted, the vast majority of the inputs are from distant sources. However, such distant sources might soon be identifiable in serial EM reconstructions by using recently developed methods that allow neural tracing with viral vectors carrying different genetically encoded labels that are distinguishable with EM [Cruz-Lopez et al., 2018; Zhang et al., 2019]. New, nondestructive imaging methods also promise to extend the distances over which circuits can be reconstructed at the ultrastructural level [Kuan et al., 2020].

Second, most studies on the influence of top-down information on perception and cognition have been done in humans and NHPs, where tools to study circuit mechanisms are lagging compared to those in rodent models. In the future, this border zone needs to be more thoroughly investigated, both by improving our toolkit for circuit-level manipulations in NHP [Dai et al., 2015; El-Shamayleh et al., 2016; Galvan et al., 2017] and seeking out meaningful touchpoints between studies on NHPs and rodents, in the spirit of Figure 4.

Third, the mechanisms by which neuromodulators influence specific cortical circuits are poorly understood; a myriad of cellular and synaptic effects have been described but understanding the overall effects will require sophisticated computational models [Seamans and Yang, 2004].

Fourth, how circuit-level influences of neuromodulators lead to changes in perception and behavior remains deeply mysterious. This is true, not only for dopamine, but other neuromodulators as well. Chief among those that seem ripe for investigation is serotonin (5-HT), given the powerful perceptual distortions that are produced by hallucinogenic drugs, most of which are believed to act through 5-HT_{2A} receptors [Nichols, 2004; González-Maeso et al., 2007; Halberstadt, 2015]. The historical events [Pollan, 2019] that led to these drugs being classified as “schedule 1” made them virtually inaccessible to the scientific community for many years. Thankfully, this historical influence appears to be on the wane, and we hope that perceptual scientists will make use of this powerful set of tools for future studies on perception.

Finally, the body of literature showing a reduced susceptibility to contextual visual illusions and abnormal corollary discharge in patients with schizophrenia, while suggestive, remains difficult to interpret for a variety of reasons including the fact that most of these patients are on a variety of psychoactive medications, are often condemned by their illness to extremely difficult socioeconomic situations, and frequently have other neuropsychiatric diagnoses. In this regard, several studies showing diminished top-down perceptual effects in the normal population that correlate with “cognitive-perceptual schizotypal traits” [Teufel et al., 2010; Bressan and Kramer, 2013] seem particularly promising, particularly given the possibility of conducting large-scale psychophysical studies online, using tools such as Amazon’s “Mechanical Turk” [Rajalingham et al., 2015; de Leeuw and Motz, 2016].

There is a tremendous gap between the conceptual simplicity of Bayesian inference and our understanding of the neural mechanisms that might implement it. Even such seemingly basic questions as how neural systems represent probability remain unsettled [Beck et al., 2008; Ma and Jazayeri, 2014; Haefner et al., 2016; Walker et al., 2020]. The situation might seem hopeless. Connectomics
has revealed seemingly Byzantine cortical circuitry [Bock et al., 2011] which can adopt a variety of different functional modes under the influence of multiple systems of neuromodulators [Bargmann and Marder, 2013], each having scores of effects at different levels of the circuit [Seamans and Yang, 2004; Tritsch and Sabatini, 2012]. While new experimental tools to probe circuit function are surely part of the solution, ultimately, what is most needed are synthetic computational models, i.e., models that themselves represent the consensus of an entire modeling community [Bower, 2015], which can integrate results across different levels of investigation into (hopefully) simpler explanations at the level of circuit motifs that perform canonical computations [Douglas and Martin, 2007; Kouh and Poggio, 2008; Carandini and Heeger, 2011; Miller, 2016] in the service of behavioral goals [Krakauer et al., 2017].

Acknowledgments

R.T.B. wishes to thank Dr. Michael Starnbach for shelter, support, and inspiration during the Covid-19 pandemic.

Author Contributions

Both authors contributed to the writing and editing of this review article.

References


Statement of Ethics

This is a review/opinion article and reports no new, original research. No experiments involving humans or other animals were performed. The authors declare that their contribution complies with internationally accepted standards for research practice and reporting.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

The writing of this article was supported by NIH R01 EY11379 and funds from the Edward R. and Anne G. Lefler Foundation and the Goldenson Family Research Fund.

Illusions, Delusions, and Your Backwards

Bayesian Brain

DOI: 10.1159/000514859

Brain Behav Evol
11
Illusions, Delusions, and Your Backwards Bayesian Brain

Brain Behav Evol
DOI: 10.1159/000514859


Symijnik M, Thier P, Leube DT, Schlotterbeck P, Lindner A. Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one’s actions. Brain. 2010 Jan;133( Pt 1):262–71.


