CellPress

Neuron
Perspective 30
THREE DECADES OF EXCITING NEUROSCIENCE SINCE 1988

Predictive Processing: A Canonical Cortical Computation

Georg B. Keller[1,2,*] and Thomas D. Mrsic-Flogel[3]
[1]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
[2]Faculty of Natural Sciences, University of Basel, Basel, Switzerland
[3]Sainsbury Wellcome Centre for Neural Circuits and Behaviour, University College London, London, UK
*Correspondence: georg.keller@fmi.ch
https://doi.org/10.1016/j.neuron.2018.10.003

This perspective describes predictive processing as a computational framework for understanding cortical function in the context of emerging evidence, with a focus on sensory processing. We discuss how the predictive processing framework may be implemented at the level of cortical circuits and how its implementation could be falsified experimentally. Lastly, we summarize the general implications of predictive processing on cortical function in healthy and diseased states.

Introduction

How does the brain distinguish between self-generated and externally generated sensory input? This was the basis of a disagreement between Hermann von Helmholtz and Charles Sherrington over a century ago. The echoes of this exchange enrich our pursuit of understanding the function of the neocortex to this day. Hermann von Helmholtz speculated that the absence of motion perception during eye movements is the result of an efference copy signal that cancels the visual feedback arising from self-generated eye movements (von Helmholtz, 1867). He argued that when pushing gently on one's eye, this cancellation does not occur, and we perceive a moving world. Less well known perhaps is the case of a patient with a unilateral traumatic lesion of the lateral rectus muscle that moves the eye temporally. When the patient would close the unaffected eye and attempt to initiate a movement of the affected eye temporally, he would report seeing the world rapidly moving in the direction of intended eye movement (von Helmholtz, 1867). Thus, the motor command to move the eye could drive perception in absence of any change in visual input. Based on these observations Helmholtz speculated that the brain must have an internal model of the sensory consequences of self-generated movements. He called this the "sense of innervation." Four decades later, Charles Sherrington revisited these ideas and argued that we have a sensory system in the musculature—the "muscular sense" (proprioception)—that provides direct sensory evidence of the position of our muscles. Based on this, he concluded that a sense of innervation would be an unnecessary assumption (Sherrington, 1900). Sherrington's reliance on bottom-up-driven sensory computations would extend to one of his most influential concepts—the receptive field—paving the way for a view of the brain that is driven to move by its sensorium. Sherrington's views of a nervous system built upon sensory-driven receptive fields would flourish over the next several decades. Describing the responses of ganglion cells in the frog's retina to small black spots, Horace Barlow argued that it is hard to avoid the conclusion that these neurons function as fly detectors (Barlow, 1953). Born was the concept of the feature detector, the postulate that the activity of neurons in sensory pathways is driven primarily by feed-forward sensory input and represents the presence of a feature or an object in the environment. The effects of this revolutionary idea are still apparent in most of our thinking of brain function. With the discovery of the simple cells in cat primary visual cortex (Hubel and Wiesel, 1959), the feature detector rapidly became the dominant narrative for our thinking about cortical function (Martin, 1994). This concept has been a guiding principle for scientific inquiry; it is apparent not only in the concept of receptive fields of neurons in visual cortex, but also in place cells (O'Keefe and Dostrovsky, 1971), grid cells (Hafting et al., 2005), face cells (Perrett et al., 1982), and concept cells (Quiroga et al., 2005). Once sensory systems of the brain have extracted an invariant representation from the sensory input, a separate part of the brain is then tasked with deciding and acting upon that representation. Following David Marr, we will call this the representational framework for describing the function of neocortex (Marr, 1982).

In parallel, the ideas of Helmholtz would resurface in the work of Erich von Holst, Horst Mittelstaedt (von Holst and Mittelstaedt, 1950), and Roger Sperry (Sperry, 1950). They were unsatisfied with an account of perception driven bottom-up by sensory input because it failed to explain how animals distinguish self-generated sensory feedback from externally generated input. One prominent example they used to illustrate that the brain must be able to make this distinction is the fact that the optokinetic reflex does not prevent self-motion of the eye. During passive viewing, full-field visual flow results in a movement of the eye that stabilizes the image on the retina; this is called the optokinetic reflex. If the animal could not distinguish between self-generated and externally generated visual input, then the optokinetic reflex would prevent any active movement of the eye. The argument is that the visual flow resulting from an eye movement would trigger the optokinetic reflex just as visual flow during passive viewing does and thus would result in a reflexive eye movement that counteracts the original eye movement. They concluded that one simple strategy to solve this problem of distinguishing self-generated sensory feedback from externally generated input in general would be to cancel the predictable consequences of

self-generated sensory feedback using an efference copy of a motor command. This requires that the brain has a mechanism to transform the efference copy of the motor command into the sensory coordinate system to cancel the reafferent sensory feedback. This transformed version of the efference copy is often referred to as a corollary discharge. Conceptually, such transformations, or internal models, are equivalent to a simulation of the external world and function to make predictions of sensory input. Kenneth Craik formulated this idea in the early 1940s as: "My hypothesis then is that thought models, or parallels, reality—that its essential feature is not 'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation" (Craik, 1943).

The mapping of the motor command onto the sensory consequences of the movement functions to simulate the environment and thus *is* the internal model of the world. The idea that the brain uses an internal model to predict sensory input based on movements and past sensory experience has been formalized in several different variants: predictive coding, hierarchical temporal memory, and Bayesian inference (Friston, 2005; Hawkins and Blakeslee, 2004; Körding and Wolpert, 2004; Rao and Ballard, 1999; Spratling, 2010). All of these are based around the idea of a generative model of the world used to predict sensory input. Following Andy Clark (Clark, 2016), we will refer to this family of theories as the predictive processing framework. Of note, we do not wish to diminish the importance of the discrepancies between the different theories we are grouping here (see, e.g., Spratling, 2017 for a review of different variants of predictive coding), but will focus on their common premise. Here, we will focus on aspects of predictive processing that are based on a comparison of sensory input with a generative model of the environment. Our aim is to discuss the physiological evidence that has convinced us that the predictive processing framework is more consistent with the data than the representational framework (see, e.g., Marr, 1982 and Martin, 1994 for discussions of the representational framework).

Predictive processing as a conceptual framework for understanding brain has a long tradition in the fields of computational and cognitive neuroscience and has been elegantly summarized elsewhere (Clark, 2013; Koster-Hale and Saxe, 2013). The principle of a comparison between predicted and actual feedback is also often used to model the function of the cerebellum (Wolpert et al., 1998) and the dopaminergic reward system (Schultz et al., 1997). Surprisingly, however, predictive processing in the neocortex has received little attention at the physiological level. This is in part due to the difficulty of designing experiments that can effectively disambiguate between the neuronal activity associated with bottom-up representation and that associated with predictive processing hypotheses, and it is in part because we have poor experimental access to internal models or control over the associated predictions. In this perspective, we argue that existing data about neural activity and neural circuit organization of the (sensory) cortex can be understood in the context of a predictive processing framework, and we highlight recent direct evidence in support. We then discuss how computations required for predictive processing might be implemented at the circuit level and propose experiments that would provide a mechanistic corroboration.

Theoretical Framework for Predictive Processing

Prediction-Error Neurons and Internal Representation Neurons

At the core of all predictive processing theories is the idea that the brain develops a generative model of the world that it uses to predict sensory input (Barlow, 1961; Craik, 1943; Gregory, 1980). The comparison of predicted and actual sensory input then updates an internal representation of the world. This process is often described as a processing hierarchy. A brain area at a higher level of the hierarchy sends a top-down signal to an area at lower level in the form of a prediction of the bottom-up input to that area. Predictions are compared to bottom-up input to compute the difference between the two (Figures 1A and 1B). This requires at least two functional classes of neurons: an internal representation neuron and a comparator or prediction-error neuron. Internal representation neurons project downward in the neural hierarchy and encode predictions about the bottom-up input. Prediction-error neurons project upward in the hierarchy and encode a difference between prediction and bottom-up input. Thus, in the lowest level of the hierarchy the bottom-up input is the sensory input, while in higher levels of the hierarchy it is the prediction errors from lower levels. When the bottom-up information matches the information carried by internal representation neurons, the responses in prediction-error neurons decrease. In sensory cortex, both internal representation neurons and prediction-error neurons are expected to be selective for specific stimulus features.

Prediction errors may come in two flavors. The bottom-up input can be stronger than predicted (for example, when an unpredicted stimulus appears) or it can be weaker than predicted (for example, when the expected stimulus does not appear or a stimulus disappears). In theory, a bidirectional change could be signaled by one neuron that has a sufficiently high basal firing rate. Increases in activity could signal more input than predicted, while a decrease could signal less input than predicted. Such a bidirectional modulation of a prediction-error signal has been observed in the dopaminergic system (Schultz et al., 1997). In the neocortex, and particularly in layer 2/3, the baseline firing rates of principal neurons are much lower (de Kock et al., 2007; Niell and Stryker, 2008; Sakata and Harris, 2009) and bidirectional modulation of activity is less plausible. In agreement with previous suggestions (Rao and Ballard, 1999), we think it is more likely that the error computation is carried out by two separate prediction-error circuits: one to signal more and one to signal less input than predicted (Figure 2). We will refer to these two types of prediction error as positive prediction error and negative prediction error.

In the predictive processing framework, predictions that arrive in a target area are based on an internal representation in the source area. To illustrate this, assume two hypothetical visual areas: one coding for geometric shapes and the other for edges. If the internal representation of a triangle is active in the geometric shape area, it will send a prediction of three edges to the edge area. Prediction-error neurons will be activated only if the bottom-up input does not match the top-down prediction. In absence of prediction errors, the internal representation for edges in the edge area and the internal representation for the triangle in the geometric shape area will remain active. These internal representations (of the triangle in the geometric shape area and edges in the edge area) are equivalent to those postulated by the

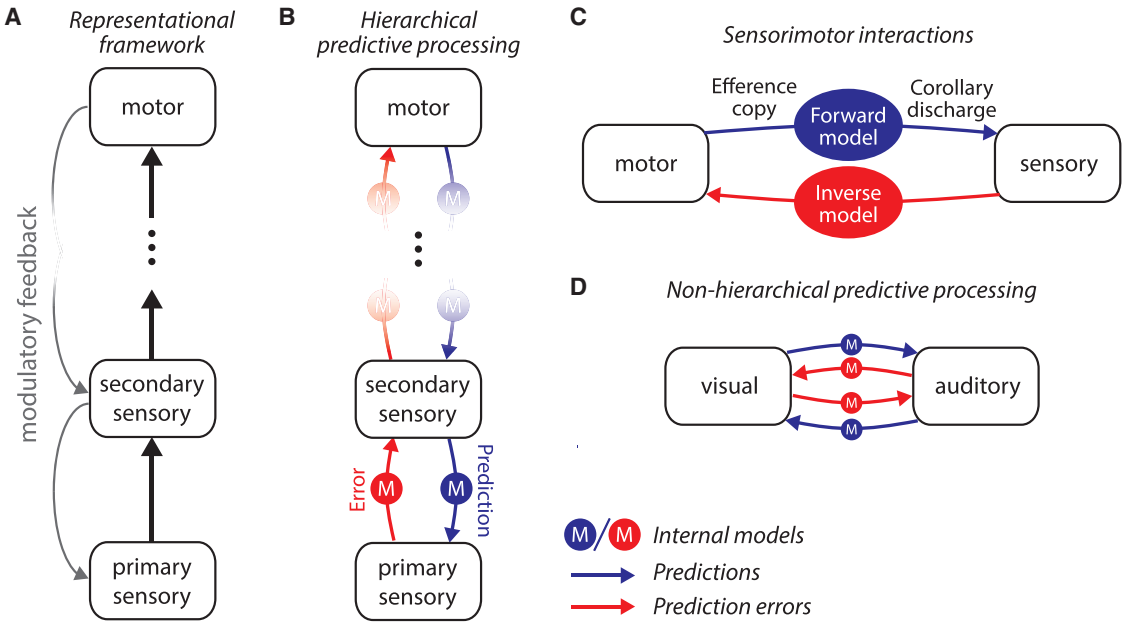

Figure 1. Inter-Areal Communication

(A) In the representational framework, internal representations are generated by bottom-up input, while top-down inputs act as modulatory signals.

(B) In a hierarchical predictive processing framework, internal representations are updated based on a comparison of a top-down prediction and bottom-up input. Prediction errors are sent forward in the hierarchy, while predictions are sent backward. The coordinate transformations between the different areas are the internal models (M).

(C) To predict the sensory consequences of self-generated movement, motor areas provide an efference copy of the motor command to sensory areas. The transformation from the motor coordinate system to the sensory coordinate system is referred to as a forward model (e.g., what do I hear when I speak). The transformed efference copy that can be directly compared to sensory signals is referred to as a corollary discharge. The transformation from the sensory coordinates to motor coordinates is referred to as an inverse model (e.g., what are the muscles I need to activate to reproduce a sound I just heard).

(D) Predictive processing does not need to follow a strict hierarchy. In the communication between two areas, both predictions and prediction errors can be sent in both directions.

representation framework. The key difference lies in how the internal representations are updated: in the representation framework through feature detectors and bottom-up drive and in predictive processing through a comparison between bottom-up input and top-down predictions based on an internal representation.

A common assumption is that predictive processing is advantageous because it is efficient; fewer spikes are necessary because only prediction errors are transmitted up the hierarchy. While prediction-error signals are sparser when input is predictable, for every bottom-up spike cancelled there needs to be a spike in a top-down prediction. In a first approximation, this means that the total number of spikes (bottom-up and top-down) remains unchanged. Hence, although there are circumstances under which predictive processing can be more efficient, in cortex this is likely not the case if efficiency is measured as the number of spikes per bit of information transmitted. We propose that the main advantage of predictive processing is that the internal representation is updated by a combination of bottom-up and top-down input and can thus be modified in absence of bottom-up input. This would provide a framework to simulate and predict the environment.

Coordinate Transformations across Cortical Areas (Internal Models)

Cerebral cortex is a network of interconnected areas that are distinguishable by their connections to the sensory input and motor output streams and by their connections to each other. We refer to the part of cortex that is the principal target of the afferents from primary sensory thalamus as primary sensory cortex. By virtue of its connectivity to the periphery, each cortical area has a unique basis for the representation of body and environment. We refer to this basis of representation as the area's coordinate system. The coordinate system of visual cortex, for example, appears to be built on Gabor filters of the visual input (such as the receptive field of simple cells), and that of auditory cortex is built on spectro-temporal filters of the auditory input. In motor cortex, the coordinate system is built on motor commands, and in inferotemporal cortex, possibly on objects or concepts (Quiroga et al., 2005). Each coordinate system only spans part of the total space of all sensory input and motor output. The transformation from one coordinate system to another is referred to as an internal model. For instance, given a current motor state and visual input, an efference copy of a motor command can be transformed to a prediction of the corresponding consequences in visual input. The motor command for an eye movement to the left can be transformed to the corresponding shift of the visual image to the right. The transformation from a motor coordinate system to a sensory coordinate system is referred to as a forward model, while a transformation from a sensory coordinate system to a motor coordinate system is referred to as an inverse model (Jordan and Rumelhart, 1992; Wolpert et al., 1995) (Figure 1C). More generally, any communication between two cortical areas will require a transformation that describes how activity in the source

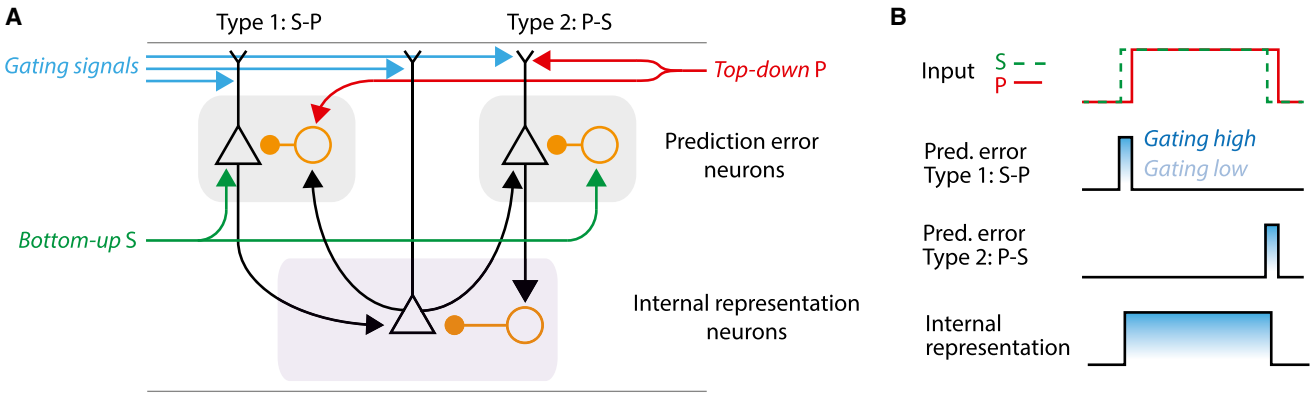

Figure 2. Schematic of the Canonical Microcircuit for Predictive Processing

(A) Positive prediction errors are computed in Type 1 neurons (Sensory Input - Prediction [S-P]), while negative prediction errors are computed by Type 2 neurons (Prediction – Sensory Input [P-S]). Triangles represent excitatory neurons, while circles represent inhibitory neurons. Note this schematic assumes hierarchical processing. In the case of a non-hierarchical communication between two areas, both areas will send and receive both top-down-like and bottom-up-like signals. (B) Schematic responses of positive and negative prediction-error neurons and internal representation neurons. Assuming the bottom-up input (S) to the circuit increases unexpectedly, positive prediction-error neurons will fire, activating both the internal representation neurons and the top-down prediction (P) from a higher area. This in turn will inhibit the positive prediction-error neuron. If the bottom-up input decreases again, negative prediction-error neurons will be activated and inhibit both internal representation neurons and top-down predictions. Responses of all three neuron types should be influenced by separate gating signals that modulate response amplitude.

area relates to activity in the target area. If such a transformation between two areas exists, activity in one area can serve as a prediction of bottom-up input in the other area. Although cortical processing can be hierarchical, especially in the vicinity of primary sensory areas, cortex as a whole is likely not arranged as a hierarchy (Gămănuț et al., 2018). Given that there are systematic correlations between auditory and visual inputs, for example, activity in an auditory area could serve as a prediction of bottom-up input in a visual area and vice versa. Thus, predictive processing does not have to follow a strict hierarchical arrangement of inter-areal connections (Figure 1D). Interestingly, a model that has been proposed recently as an alternative to hierarchical predictive processing is a variant of a predictive processing architecture in which the flow of signals is reversed, predictions are sent up the hierarchy, and errors are sent down the hierarchy (Heeger, 2017). In the absence of a strict hierarchy, the communication between areas would always entail the exchange of predictions and errors in both directions.

Experimental Considerations

When evaluating evidence that may distinguish the two alternative descriptions of cortical function, it is worth noting that the predictive processing framework is an extension of the representational framework. To illustrate this, we will make a few simplifying assumptions. In the representational framework, the response *R* of a neuron can be modeled as a function *V* of the bottom-up input.

$$R = V(\text{bottom-up}) \quad (1)$$

This function can be arbitrarily complex and, in the case of visual or auditory receptive fields, is in the form of a convolution with a receptive field. The predictive processing framework differs to this in that, in addition to internal representation neurons, it postulates the existence of prediction-error neurons. The response of prediction-error neurons is the difference between a function *V* that depends on the bottom-up input and a function *P* that depends on the top-down input—or, more specifically, the prediction of the bottom-up input *V*.

$$R = \pm\,(V(\text{bottom-up}) - P(\text{top-down})) \quad (2)$$

For simplicity, we have ignored multiplicative gains of response magnitude, which can be incorporated in both frameworks. The reason the two response types are hard to distinguish is that experimentalists have some control over the bottom-up input—at least in sensory areas of the brain—but have only poor control of the top-down input or predictions generated on a moment-by-moment basis. If experiments are performed by averaging data over many trials, for each of which the top-down input may vary, or experiments are performed under conditions in which top-down input is altered or gated off (e.g., by anesthesia), *P* reduces to a constant and (2) can be written in the form of (1). With the prediction error driven just by the stimulus, the internal representation will be updated by bottom-up input and will look like the one postulated by the representation framework (Figure 2). Under these conditions, both internal representation neurons and positive prediction-error neurons will have responses identical to the ones predicted by the representational framework. Thus, to design experiments that could distinguish between the two frameworks, experimentalists must be able to control or measure the prediction. Typically, this is not possible, and instead a proxy is used for the animal's predictions. In the context of sensory processing, self-generated motion is one possible proxy for a prediction of the resulting visual feedback (e.g., optic flow). This assumes that animals learn how sensory feedback couples to movement with experience. In a first approximation, the representational framework predicts that neuronal responses will not differ in conditions when the stimulus is externally generated versus when it is the consequence of self-motion. The predictive processing framework

instead postulates that responses in a subset of neurons, the prediction-error neurons, signal a deviation, or a mismatch, between predicted and actual sensory input. Based on this argument, much of the experimental focus in the effort to test the hypothesis of predictive processing in cortex was on prediction-error responses. In the next section, we will summarize the evidence for cortical responses that are consistent with predictive processing.

Evidence for Predictive Processing in Cortical Circuits

Behavioral Evidence

The idea that our perception of the world is an active and constructive process has an intuitive appeal to explain much of our everyday experience of the world. Our predictions frequently interfere with what we perceive. Our voice sounds eerily different when we hear it in a recording, and we perceive our own singing to be much closer to pitch than it actually is. In visual illusions, we see color where there is none, simply because we know objects rarely change color (Foster, 2011), or miss things that happen right in front of our eyes (Simons and Chabris, 1999). In these cases, what we expect to hear or see interferes with, and even supersedes, what we actually hear and see. We refer to the conditions in which we can prove that our predictions interfere with perception as sensory illusions. Given our frequent disagreements with others over the attributes of objects we see, or over what we hear, it is probably appropriate to describe perception as a controlled hallucination (Clark, 2016). This tainting of our access to reality must come at some advantage. One advantage of having an internal model of the world is to allow us to predict the future. We cannot only anticipate the sensory consequences of our own movements, but also physical attributes or dynamics of objects and other agents in the world. You can look at a photograph of a football player - one leg on the ground in front of the player, the other retracted behind the ball – and know instantly and without deliberation what will happen next. Often such predictions are not trivial and depend on detailed knowledge about the physical properties of the objects we are looking at, our model of the intentions or actions of the agents, and their context in the particular moment. These examples, and many others like it, give intuitive support for the idea that the brain is a predictive processing machine. In this section, we highlight the physiological evidence for predictive processing in neural circuits of the sensory neocortex.

Prediction-Error Signals

In neocortex, early evidence for the predictive processing framework did not arise with new data, but from the demonstration that classical visual phenomena, like end-stopping (Hubel and Wiesel, 1965), can be explained as a prediction error (Rao and Ballard, 1999). One central idea here was that the suppression of the response that appears when a stimulus extends into the surround of the classical receptive field is the consequence of top-down inhibition. In this way, the stimulus in a given location acts as a prediction of the stimulus in the neighboring region. This prediction, relayed via activation of a higher-level representation, inhibits responses of neurons with receptive fields in neighboring parts of the visual field to the same stimulus. In layer 2/3 of mouse visual cortex, somatostatin-positive interneurons, likely driven by lateral projections from neighboring cortical neurons, have been shown to have a causal role in surround suppression (Adesnik et al., 2012; Angelucci et al., 2017). The idea of a top-down prediction that acts to inhibit bottom-up input was later used to demonstrate that a large variety of classical visual receptive field properties can be explained in a predictive processing framework (Spratling, 2010). This type of comparison is consistent with a positive prediction error: the top-down prediction acts to inhibit the predictable bottom-up input. A top-down prediction that functions to inhibit bottom-up input should result in a response decrease when stimuli become predictable. This is indeed the case when stimuli become predictable, either as the result of a learned association with a preceding stimulus (Egner et al., 2010; Meyer and Olson, 2011) or after frequent presentation of the same stimulus, in which case the suppression is often described as sensory adaptation (Ulanovsky et al., 2003). Another simple form of increased predictability of a stimulus is prolonged presentation of the same stimulus, during which sensory responses typically decrease in magnitude. This form of sensory adaptation occurs at many levels in the sensory processing hierarchy, but certain forms, like contrast adaptation in visual cortex, are thought to be, at least in part, cortical in origin (Carandini, 2000; Keller et al., 2017; Maffei et al., 1973). Although adaptation would be consistent with top-down inhibition, early experiments studying mechanisms of contrast adaptation using intracellular recordings in visual cortex of anesthetized animals found no evidence of inhibition contributing to contrast adaptation (Carandini and Ferster, 1997). More recently, it was found that levels of inhibition increase with stimulus duration (Keller and Martin, 2015) and are selectively suppressed by anesthesia (Haider et al., 2013). Consistent with a strong top-down influence, contrast adaptation has been shown to depend on the behavioral relevance of a stimulus (Keller et al., 2017). Hence, it is possible that certain forms of sensory adaptation in the awake animal are driven by top-down inhibition.

Similarly, there may be top-down inhibition of the sensory consequences of self-generated movement. There is evidence for this in auditory cortex, where responses are generally supressed during self-generated locomotion via top-down projection that recruits local inhibition (Schneider et al., 2014). Consistent with the idea that the effect of these top-down predictions can be modulated in a context-dependent manner, certain forms of response adaptation in visual cortex have been shown to be dependent on the task relevance of the stimulus (Keller et al., 2017).

If increasing stimulus predictability results in a response reduction, a violation of a strong prediction should trigger a response increase. Evidence in support comes from the discovery of prediction-error signals in primary sensory areas of cortex, where responses were quantified to unexpected changes in the coupling between self-generated movements and sensory feedback. Using manipulations of visual feedback from hand movements, work in humans found a selective activation of primary visual cortex to incongruences between hand movements and visual feedback that could not be explained by the visual input alone (Stanley and Miall, 2007). Manipulating auditory feedback of self-generated vocalizations in marmosets revealed responses in primary auditory cortex that were selective to deviations between expected and actual auditory feedback (Eliades and Wang, 2008). Similar observations were made in primary auditory pallium of the songbird (Keller and Hahnloser, 2009).

These responses could not be explained by the change in sensory input, as they were only apparent during manipulations of self-generated feedback and not when the animal was passively observing or hearing the same stimulus. However, in all of these experiments, the responses were triggered by an unexpected change to sensory feedback in the form of an additional stimulus that differed from the one expected. The key signal that is more difficult to explain in a representation framework is a response to the absence of a predicted sensory input or a negative prediction error. Such signals have been found in layer 2/3 of primary visual cortex (V1) of the mouse, where a subset of neurons responds selectively to the absence of expected visual flow (Keller et al., 2012) or the absence of an expected visual stimulus (Fiser et al., 2016). We have referred to this type of negative prediction error as a mismatch response. Although mismatch responses also exist in layer 5 neurons, they are likely more prevalent in layer 2/3 (Saleem et al., 2013).

In the predictive processing framework, prediction-error signals in sensory cortices are expected to be feature-specific and not simply the result of a surprise response. That is, they should signal the type of deviation from prediction and not simply the fact that there was a deviation. Accordingly, responses of mismatch neurons in layer 2/3 of mouse V1 were found to signal deviations between predicted and actual visual flow in spatially confined areas of the visual field (Zmarz and Keller, 2016). These mismatch signals parallel visual signals in magnitude, spatial resolution and retinotopic organization, suggesting that mismatch signals are computed based on local visual cues and that visual and mismatch signals are separate aspects of the same computation.

Circuits for Predictive Processing

Observing prediction-error signals in the neocortex does not prove they are computed therein. However, if cortical circuits do implement predictive processing, this requires at least three components: a comparator circuit that computes the prediction error between bottom-up input and predictions, a circuit to maintain an internal representation that gives rise to predictions, and a modulating or gating signal that sets the precision or weight of the prediction error. The circuit elements required to generate prediction errors are present in each module of the neocortex. Cortical areas receive bottom-up input from the thalamus or other cortical areas as well as extensive top-down inputs from many nearby and distal cortical areas and higher-order thalamic nuclei (Felleman and Van Essen, 1991; Markov et al., 2014; Oh et al., 2014; Sherman, 2016; Zingg et al., 2014), consistent with predictions from multiple modalities. The top-down inputs can be very dense—as, for instance, the top-down input from anterior cingulate cortex to V1 (Zhang et al., 2014)—and target both excitatory and inhibitory neurons in layer 2/3 monosynaptically (Leinweber et al., 2017; Mao et al., 2011; Yang et al., 2013; Zhang et al., 2014). The comparator circuits that generate negative and positive prediction errors require differential wiring of bottom-up and top-down inputs onto subsets of excitatory and inhibitory neurons. Negative prediction-error neurons will respond when top-down excitation exceeds bottom-up inhibition (whereby increasing strength or saliency in predictions should result in increasing strength of mismatch). It follows that subsets of inhibitory neurons are mainly bottom-up driven, either directly or via local excitatory relays, and that these provide input preferentially to negative prediction-error neurons. In layer 2/3 of visual cortex, a subset of somatostatin-expressing interneurons are thought to provide visually driven inhibition to negative prediction-error neurons (Attinger et al., 2017). Conversely, positive prediction-error neurons will respond when bottom-up excitation exceeds top-down inhibition. Accordingly, a different set of interneurons is expected to be driven more strongly by top-down input and provide inhibition to positive prediction-error neurons. This form of top-down inhibition is a frequent circuit motif in cortex (Lee et al., 2013; Schneider et al., 2014; Zhang et al., 2014).

Finally, we suggest that negative and positive prediction-error neurons exert opposite effects on their targets. Negative prediction errors should act mainly by engaging bottom-up inhibition in their target areas, thus suppressing the current internal representation. Conversely, positive prediction-error neurons provide bottom-up excitation to target areas, thus activating a new cohort of neurons. The combined effect of positive and negative prediction-error neurons is to update the internal representation that best approximates, or predicts, the current environment.

Top-Down Signals Are Predictions

With the discovery of strong motor-related signals in primary visual cortex in the complete absence of visual input (Keck et al., 2013; Keller et al., 2012; Saleem et al., 2013) came further evidence that a representational framework could explain only a fraction of the responses, even in primary sensory areas. Modulation of visual responses by locomotion or arousal (Niell and Stryker, 2010; Reimer et al., 2016; Vinck et al., 2015) is thought to be the consequence of neuromodulatory inputs (Fu et al., 2014; Polack et al., 2013), which exert context-dependent influence on responses in visual cortex (Pakan et al., 2016). However, modulatory inputs alone cannot account for motor-related signals in visual cortex in the absence of visual input. A driving motor-related prediction of visual input, however, could account for these non-visual signals. We have recently argued that in visual cortex, one source of the prediction of visual input given movement is the anterior cingulate cortex (Leinweber et al., 2017). Activity in axons of anterior cingulate neurons in visual cortex conveys an experience-dependent prediction of visual flow (rather than copies of motor commands) as a function of the turning of the mouse in a virtual environment. Importantly, we found that this motor-related input is shaped by the coupling between movement and visual feedback the mouse has experienced previously.

Locomotion is just one possible proxy for a prediction of visual input, and other signals, like spatial location, could serve a similar function. Consistent with this, neurons in layer 2/3 of V1 respond robustly to the omission of a stimulus the mouse expects to see at a certain location in a virtual environment (Fiser et al., 2016). In principle, any signal that explains some of the variance in the visual input can serve as a prediction of visual feedback. Vestibular or eye movement signals could serve as predictions of full-field visual flow. In a learned coupling between two sensory stimuli—e.g., a sound and a visual input—one can serve as a prediction of the other. It is therefore plausible that long-range cortical communication conveys specific predictions of input to the target areas that are associated by experience with signals in the source area (Larkum, 2013; Roelfsema and Holtmaat, 2018). Consistent with this view, specific signals related to self-generated movement, head direction, animal's

spatial location, and stimulus timing have been observed across several sensory areas (Lütcke et al., 2010; Manita et al., 2015; Mao et al., 2011; Poort et al., 2015; Schneider et al., 2014; Vélez-Fort et al., 2018). These diverse sources of contextual input may thus provide predictions required for computation of prediction errors and for updating internal representations based on information from a given sensory modality.

Learning to Predict

A key assumption of the predictive processing framework is that internal models are learned and that experience shapes the circuits required for generating predictions and computing prediction errors. While evolution has generated a template of reproducible long-range projections linking cortical areas, often reciprocally, it is the interaction with the world that refines these connections to generate internal models. Sensory experience sculpts the connectivity between neurons in an activity-dependent manner, such that nearby cortical neurons with similar responses (i.e., those that fire together) can preferentially link up into synaptically connected subnetworks with strong recurrent excitation (Cossell et al., 2015; Ko et al., 2011, 2013). We suggest that a similar principle may apply to the establishment of long-range networks across cortical areas, whereby a history of correlated firing determines which neurons become associated. In the context of predictive processing, this would apply equally to sculpting the bottom-up and top-down connectivity between internal representation neurons encoding components of the same object as well as between prediction-error neurons and internal representation neurons within and across areas. In visual cortex of rodents, predictive responses emerge in an experience-dependent way (Fiser et al., 2016; Makino and Komiyama, 2015; Poort et al., 2015). Through passive sensory experience, visual cortex responses become predictive of upcoming visual stimuli (Gavornik and Bear, 2014; Xu et al., 2012). Through experience of visuomotor coupling, predictions of visual flow are learned (Attinger et al., 2017; Leinweber et al., 2017), and through experience in a spatial environment, responses emerge that are predictive of the visual input at a given spatial location (Fiser et al., 2016). These predictive responses is sensory areas may thus be driven by long-range inputs whose influence is shaped by experience.

We assume that perception is linked to the internal representation of the world and that we only perceive a stimulus if the internal representation for that stimulus is active. This internal representation is what predictions are based on. During a given percept, internal representation neurons, likely distributed across several associated areas, are active. If neurons maintaining the internal representation are the basis for predictions of bottom-up input impinging on the same or other cortical areas, they should exhibit a set of functional and connectional features. First, an internal representation requires a circuit mechanism that maintains the activity in a population of neurons for the time a stimulus is perceived. A number of plausible mechanisms have been proposed for the persistence of neural activity, including strong and selective recurrent excitation between coactive neuronal assemblies within the cortex (Cossell et al., 2015; Li et al., 2016; Perin et al., 2011; Song et al., 2005) or via thalamocortical loops (Guo et al., 2017; Reinhold et al., 2015; Schmitt et al., 2017). Moreover, internal representations may not require stable patterns of activity, but could be maintained using dynamic attractors. In either case, neurons representing the internal model are expected to exhibit more sustained and dense activity than neurons that function as comparators. Second, internal representation neurons should make connections within the area they reside as well as provide top-down input to lower areas within the same sensory modality and/or project to associated cortical areas dedicated to other modalities. Finally, as internal representations need to be updated by prediction errors, the neurons encoding the internal representation should be densely connected with the comparator circuit encoding the same feature. Interestingly, these functional and anatomical characteristics are hallmarks of a subset of cortical neurons prevalent in deeper layers (Harris and Mrsic-Flogel, 2013; Harris and Shepherd, 2015). However, how internal representations are maintained in the cortical circuit and how they may be used to generate top-down predictions is still unclear.

Precision Signals

In sensory cortex, responses can be modulated and given precedence depending on the context in which the stimulus is perceived. This implies that predictions and prediction errors may be modulated in a context-dependent manner. Conceptually, a dynamic modulation of the influence of top-down and bottom-up input is consistent with an attentional modulation of sensory input (Posner and Gilbert, 1999). Direct evidence for a modulation of prediction and prediction-error signals comes from a variety of experiments. Experience-dependent predictive responses in visual cortex, for example, are only apparent under quiet wakefulness, but not if the animal is active (Xu et al., 2012). Similarly, adaptation of sensory responses depends on context and the task-relevance of sensory input (Keller et al., 2017). In sensorimotor learning, prediction errors during movement are thought to correct the motor program. Here, the requirement for a context-dependent gating of prediction errors stems from the fact that prediction errors that occur during passive observation should not interfere with the motor program. This requires an error signal that can gate plasticity, whose magnitude can be adjusted in a context-dependent manner.

The source of such a modulating or gating signal is not always clear. Attentional gain modulation in visual cortex has been speculated to be driven by long-range cortical input (Zhang et al., 2014), input from higher-order thalamus (Purushothaman et al., 2012; Roth et al., 2016; Wimmer et al., 2015), or neuromodulatory inputs (Fu et al., 2014; Polack et al., 2013; Pinto et al., 2013; Thiele and Bellgrove, 2018). Neuromodulatory input can not only gate plasticity (Kilgard and Merzenich, 1998; Martins and Froemke, 2015; Weinberger, 2004), but also change the balance of top-down versus bottom-up influence (Yu and Dayan, 2005). Specifically, the neuromodulatory tone may shift the relative contribution of bottom-up and top-down signals such that the influence of prediction errors can be modulated according to the internal state of the animal, which would determine the extent to which bottom-up inputs are used to update the internal model. It remains to be seen how different modulatory signals are combined to alter the sensitivity by which cortical circuits prioritize and respond to sensory information or report prediction errors. A more complete understanding of these modulation mechanisms requires further exploration and may be key to understanding cortical dysfunction.

Implications for Cortical Function and Dysfunction

The immediate appeal of predictive processing is that it could be a basic computational primitive implemented in different variants throughout the brain. Evidence consistent with predictive processing has been found in a variety of different brain regions. The function of the dopaminergic system has been described in terms of reward prediction errors (Schultz et al., 1997). Many of the models of cerebellar function are based on the concepts of internal models and prediction errors (Wolpert et al., 1995, 1998). Cerebellum-dependent sensorimotor learning is thought to be driven by sensory prediction errors computed as a comparison between intended and actual sensory feedback (Brooks et al., 2015). Similarly, certain forms of cortex-dependent sensorimotor learning are thought to be driven by performance errors (Houde and Jordan, 1998; Konishi, 1965). In vocal learning, these performance errors have been suggested to be computed based on a comparison of intended and actual sensory feedback (Keller and Hahnloser, 2009).

So, what could the implications be of describing brain function in terms of an internal representation of the world that is updated through comparison with incoming sensory information? Of course, we do not have a definitive answer to this question, but what we will attempt to do in this section is to explain where we see promise of predictive processing. First, temporarily decoupling the internal representation from sensory input would allow one to run the model as a simulation. In this way, one could simulate the consequences of one's actions without having to perform them. This is likely what we refer to as thinking. Second, we would postulate that perception is based on a finely tuned process that continuously balances internal predictions against bottom-up signals to update an internal representation. If this process is imbalanced such that the internal representation is driven too strongly by top-down predictions, one might perceive things that are not there, or interpret intention into action of others where there is none. This would likely resemble positive symptoms of schizophrenia, as has been argued previously (Corlett et al., 2009; Fletcher and Frith, 2009; Frith et al., 2000). Conversely, if the effect of top-down predictions were too weak and the internal representation were dominated by bottom-up sensory input, one might be unable to adequately predict sensory input or understand intentions of others. Assuming the brain lacks the ability to generate an internal model with sufficient predictive capacity, a simple behavioral strategy would be to engage in stereotyped repetitive behavior that makes the input more predictable. A dysfunction in the brain's ability to make accurate predictions has been proposed as one of the attributes of autism (Lawson et al., 2014, 2017; Sinha et al., 2014). Based on this, one could speculate that schizophrenia and autism are opposite ends of the same circuit imbalance in which the internal representation of the world is either driven too strongly or too weakly by predictions. We speculate that alterations in predictive processing circuits may be common to both disorders. Anti-NMDA receptor encephalitis, for example, in which NMDA receptors are targeted by the immune system, results in symptoms that resemble those of schizophrenia when adults are affected, while it results in symptoms that resemble those of autism when children are affected (Creten et al., 2011; Dalmau et al., 2007; Titulaer et al., 2013). In addition, there is a common gene expression network that is dysregulated in the two conditions (Gandal et al., 2018), possibly in opposite directions (Crespi et al., 2010). Thus, the absence of key molecular regulators of synaptic plasticity (e.g., glutamate receptors) may lead to a failed experience-dependent adjustment of the connections in circuits that maintain an internal representation of the world through a comparison with incoming sensory input. In turn, this may cause aberrations in predictive processing and altered cortical function in these neurodevelopmental disorders.

The Experiments That Need to Be Done

In this section, we outline experiments that may test, refine, or reject the model of predictive processing in the cerebral cortex using currently available technologies.

(1) One of the core postulates is that there are neurons in each area of cortex that maintain an internal representation of the world in a local coordinate system. To the best of our knowledge, there has been no clear demonstration of the existence of such neurons. The problem with identifying such neurons is that they will exhibit responses that appear driven by a bottom-up input in many conditions. However, there are a few functional and anatomical characteristics that might aid in identifying them. First, internal-representation neurons should comprise a class of neurons separate from the prediction-error neurons. It is possible that internal-representation neurons are intermixed with prediction-error neurons in different cortical layers or that they are enriched in deep layers of cortex (Bastos et al., 2012), which are the main source of top-down signals (Felleman and Van Essen, 1991; Markov et al., 2014). Second, internal-representation neurons provide input to both local prediction-error neurons and, either directly or indirectly, give rise to projections, which convey predictions to other cortical areas. Third, within a cortical area, the current internal representation should function like a prediction (the current scene is a decent predictor of future scenes). Therefore, local internal-representation neurons should interact with prediction-error neurons in the same way top-down predictions do. Negative prediction-error neurons should be net excited by internal representation neurons, while positive prediction-error neurons should be net inhibited. Fourth, activity in prediction-error neurons should update the local internal representation. Positive prediction-error neurons, which report more bottom-up input than expected, should net activate the corresponding local internal-representation neurons. Conversely, negative prediction-error neurons, which report less input than expected, should net inhibit the corresponding internal-representation neurons. Given that we do not have a genetic handle on the different functional neuronal classes, experiments would need to rely on the possibility that there is a predominance of one or the other neuron type in different cortical layers (for example, a preponderance of prediction-error neurons in layer 2/3 and of internal representation neurons in layer 5). In this way, one could test the influence of activation of a subset of putative internal representation neurons, either in a given cortical layer or through targeted photostimulation (Packer et al., 2015), on functionally identified prediction-error neurons in layer 2/3.

(2) Assuming that cortex is built based on a canonical circuit motif (Douglas et al., 1989), we should find prediction-error neurons for every instance of a behaviorally meaningful correlation of activity across any two cortical areas. To illustrate this, let us take

the interaction between motor areas and visual areas. Every movement that is coupled to a predictable change in visual input (whole-body translation; eye, head, limb, or whisker movements; etc.) should have a corresponding set of prediction-error neurons in a sensory cortex. We think that a subset of layer 2/3 neurons in mouse V1 functions to compute prediction errors between whole-body translation and visual input (Attinger et al., 2017; Zmarz and Keller, 2016). Similar predictions of sensory input may be based on spatial location, head direction, or sensory input in other modalities. Consistent with this, a subset of neurons in layer 2/3 of mouse V1 signal a prediction error between a prediction of visual feedback based on spatial location and visual input (Fiser et al., 2016). Similar prediction-error neurons may exist for predictions based on auditory input, vestibular input, etc.

(3) Most work on cortical circuits for predictive processing has focused on primary sensory areas. The advantage of examining in a primary sensory area is that there is some experimental control over bottom-up inputs. Assuming predictive processing describes a canonical cortical computation, we should find similar prediction-error signals in other cortical areas. These prediction-error signals would be encoded in the same coordinate system as the bottom-up input to the area. By this we mean that there should be neurons in prefrontal cortex that signal deviations in a conceptual rule the animal has learned or a deviation in social patterns the animal expects to encounter in its conspecifics. There is some evidence that a subset of neurons in motor cortex signals a deviation between intended and actual motor state, given proprioceptive or other sensory feedback (Heindorf et al., 2018; Inoue et al., 2016).

(4) In neocortex, it is likely that the prediction-error circuits are shaped by experience, as they are in primary visual cortex for sensorimotor and spatial predictions (Attinger et al., 2017; Fiser et al., 2016). What is still unclear is which synapses in this circuit undergo experience-dependent plasticity. In the case of the negative prediction-error circuit in layer 2/3 of mouse V1, we can constrain the site of plasticity to some extent. The activity in the somatostatin-positive inhibitory interneurons, which mediate the bottom-up inhibition, is not dependent on sensorimotor experience (Attinger et al., 2017). Hence, experience-dependent plasticity must modify at least one of the other connections in the circuit. This could be the synapse from the inhibitory neuron onto the prediction-error neuron, or the one from the top-down predictive input onto the prediction-error neuron. Identifying the site of plasticity could be achieved by preventing experience-dependent plasticity in specific cell types during sensorimotor learning (Sawtell et al., 2003).

(5) To explain the phenomenon of attention and the fact that passive experience does not modify the motor program during sensorimotor learning, we predict the existence of a modulating signal that can attenuate or amplify prediction-error signals. In vocal learning, for example, listening to conspecific vocalizations should not generate prediction errors that update the motor program for vocalization. Hence, prediction-error signals have to be gated on only during times of self-vocalization. A similar modulation mechanism is necessary to explain attention-related phenomena. There may be at least three defining characteristics of such a signal. First, this input should selectively alter the coupling between prediction-error neurons and internal representation neurons. Second, in sensory and motor regions, the modulating signal should be correlated with movement. Third, manipulations of the modulating system should result in shifts in the balance between top-down and bottom-up inputs, and this may change the gain of responses in prediction-error neurons according to internal state. Possible sources of modulatory signals include classical neuromodulatory systems (e.g., acetylcholine, noradrenaline) or the thalamus, both of which have been shown to change the operating regime of cortical circuits (Fu et al., 2014; Polack et al., 2013; Purushothaman et al., 2012; Wimmer et al., 2015; Pinto et al., 2013).

(6) Assuming psychosis is a state of imbalance in processing in which the internal representation is not updated by sensory feedback and thus dominated by predictions, and prediction errors are either too strong or too weak, we would expect to find a common functional signature of drugs that reduce psychosis. It is conceivable that antipsychotic drugs function by changing the balance between positive and negative prediction errors or by changing the balance between top-down and bottom-up input. Testing this hypothesis requires systematic characterization of the effects of drugs that are anti- or pro-psychotic on prediction errors, predictions, and bottom-up signals.

Predictive processing in the form we are proposing here will very likely not provide a complete description of cortical function. Hence, our intention should be to identify the limits and shortcomings of the framework in order to formulate a more complete theory. One thing is certain: we need to move away from a purely representational understanding of the cortex if we aim to make conceptual progress in this endeavor.

Conclusion

Inquiries into the function of any biological structure are best conducted with an eye on David Marr's three levels of analysis (Marr, 1982). We have discussed a possible algorithm for the function of neocortex and how this algorithm could be implemented in cortical circuits. What remains to be addressed is what the goal of the computation is. In other words, what is the evolutionary advantage of having a cortex? Cortex emerged in evolution on top of a fully functional brain capable of sensory processing, movement control, and decision making. It had to integrate its input and output circuitry into this functioning brain. It is highly probable that the influence of cortex was initially sparse and modulatory. Reminiscent of this idea is the fact that cortical lesions have relatively subtle phenotypes in rodents (Kawai et al., 2015; Miri et al., 2017) compared to the dramatic effect of similar lesions in humans (Twitchell, 1951). The effect of motor cortex lesions is pronounced, however, during certain forms of motor learning (Kawai et al., 2015) or when animals need to initiate a behavioral response to unexpected sensory feedback perturbations (Lopes et al., 2016). Thus, the function of nascent cortex may have been to enable behavioral alternatives in order to evaluate and select novel responses to a given sensory input. One strategy to expand behavioral flexibility is to employ a simulation of the world that allows for rapid testing and continuous preparation of possible motor plans. With the increasing importance of social interaction and coordination, this same mechanism may have been adapted to model and simulate other agents in the world (Rizzolatti et al., 2001). Thus, human neocortex may be the product of an evolutionary arms race to build a machine that allows us to make predictions of

the ever increasingly complex behavior of our conspecifics—or in the words of the Scottish poet Robert Burns:

> "But, Mousie, thou art no thy-lane,
>
> In proving foresight may be vain;
>
> The best-laid schemes o' mice an' men
>
> Gang aft agley,
>
> An' lea'e us nought but grief an' pain,
>
> For promis'd joy!
>
> Still thou art blest, compar'd wi' me
>
> The present only toucheth thee:
>
> But, Och! I backward cast my e'e.
>
> On prospects drear!
>
> An' forward, tho' I canna see,
>
> I guess an' fear!"
>
> Robert Burns, from "To a Mouse."

ACKNOWLEDGMENTS

We thank Rainer Friedrich, Andreas Keller, Marcus Stephenson-Jones, and the entire Keller lab for comments on earlier versions of this manuscript. This work was funded by the Gatsby Charitable Foundation (GAT3212 / GAT3361) and Wellcome (090843/E/09/Z) (T.D.M.-F.), the Swiss National Science Foundation and the Novartis Research Foundation (G.B.K.).

REFERENCES

Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z.J., and Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. Nature *490*, 226–231.

Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., and Bressloff, P.C. (2017). Circuits and Mechanisms for Surround Modulation in Visual Cortex. Annu. Rev. Neurosci. *40*, 425–451.

Attinger, A., Wang, B., and Keller, G.B. (2017). Visuomotor Coupling Shapes the Functional Development of Mouse Visual Cortex. Cell *169*, 1291–1302.e14.

Barlow, H.B. (1953). Summation and inhibition in the frog's retina. J. Physiol. *119*, 69–88.

Barlow, H.B. (1961). Possible principles underlying the transformations of sensory messages. In Sensory Communication, W. Rosenblith, ed. (MIT Press), pp. 217–234.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical microcircuits for predictive coding. Neuron *76*, 695–711.

Brooks, J.X., Carriot, J., and Cullen, K.E. (2015). Learning to expect the unexpected: rapid updating in primate cerebellum during voluntary self-motion. Nat. Neurosci. *18*, 1310–1317.

Carandini, M. (2000). Visual cortex: Fatigue and adaptation. Curr. Biol. *10*, R605–R607.

Carandini, M., and Ferster, D. (1997). A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. Science *276*, 949–952.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. *36*, 181–204.

Clark, A. (2016). Surfing uncertainty : prediction, action, and the embodied mind (Oxford University Press).

Corlett, P.R., Frith, C.D., and Fletcher, P.C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. Psychopharmacology (Berl.) *206*, 515–530.

Cossell, L., Iacaruso, M.F., Muir, D.R., Houlton, R., Sader, E.N., Ko, H., Hofer, S.B., and Mrsic-Flogel, T.D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. Nature *518*, 399–403.

Craik, K.J. (1943). The Nature of Explanation (Cambridge University Press London).

Crespi, B., Stead, P., and Elliot, M. (2010). Evolution in health and medicine Sackler colloquium: Comparative genomics of autism and schizophrenia. Proc. Natl. Acad. Sci. USA *107* (*Suppl 1*), 1736–1741.

Creten, C., van der Zwaan, S., Blankespoor, R.J., Maatkamp, A., Nicolai, J., van Os, J., and Schieveld, J.N. (2011). Late onset autism and anti-NMDA-receptor encephalitis. Lancet *378*, 98.

Dalmau, J., Tüzün, E., Wu, H.Y., Masjuan, J., Rossi, J.E., Voloschin, A., Baehring, J.M., Shimazaki, H., Koide, R., King, D., et al. (2007). Paraneoplastic anti-*N*-methyl-D-aspartate receptor encephalitis associated with ovarian teratoma. Ann. Neurol. *61*, 25–36.

de Kock, C.P.J., Bruno, R.M., Spors, H., and Sakmann, B. (2007). Layer- and cell-type-specific suprathreshold stimulus representation in rat primary somatosensory cortex. J. Physiol. *581*, 139–154.

Douglas, R.J., Martin, K.C., and Whitteridge, D. (1989). A Canonical Microcircuit for Neocortex. Neural Comput. *1*, 480–488.

Egner, T., Monti, J.M., and Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. J. Neurosci. *30*, 16601–16608.

Eliades, S.J., and Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature *453*, 1102–1106.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex *1*, 1–47.

Fiser, A., Mahringer, D., Oyibo, H.K., Petersen, A.V., Leinweber, M., and Keller, G.B. (2016). Experience-dependent spatial expectations in mouse visual cortex. Nat. Neurosci. *19*, 1658–1664.

Fletcher, P.C., and Frith, C.D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat. Rev. Neurosci. *10*, 48–58.

Foster, D.H. (2011). Color constancy. Vision Res. *51*, 674–700.

Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. *360*, 815–836.

Frith, C.D., Blakemore, S., and Wolpert, D.M. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. Brain Res. Brain Res. Rev. *31*, 357–363.

Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., and Stryker, M.P. (2014). A cortical circuit for gain control by behavioral state. Cell *156*, 1139–1152.

Gămănuţ, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., Van Essen, D.C., Knoblauch, K., and Burkhalter, A. (2018). The Mouse Cortical Connectome, Characterized by an Ultra-Dense Cortical Graph, Maintains Specificity by Distinct Connectivity Profiles. Neuron *97*, 698–715.e10.

Gandal, M.J., Haney, J.R., Parikshak, N.N., Leppa, V., Ramaswami, G., Hartl, C., Schork, A.J., Appadurai, V., Buil, A., Werge, T.M., et al.; CommonMind Consortium; PsychENCODE Consortium; iPSYCH-BROAD Working Group (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science *359*, 693–697.

Gavornik, J.P., and Bear, M.F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. Nat. Neurosci. *17*, 732–737.

Gregory, R.L. (1980). Perceptions as hypotheses. Philos. Trans. R. Soc. Lond. B Biol. Sci. *290*, 181–197.

Guo, Z.V., Inagaki, H.K., Daie, K., Druckmann, S., Gerfen, C.R., and Svoboda, K. (2017). Maintenance of persistent activity in a frontal thalamocortical loop. Nature *545*, 181–186.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E.I. (2005). Microstructure of a spatial map in the entorhinal cortex. Nature *436*, 801–806.

Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. Nature *493*, 97–100. Published online November 21, 2012.

Harris, K.D., and Mrsic-Flogel, T.D. (2013). Cortical connectivity and sensory coding. Nature *503*, 51–58.

Harris, K.D., and Shepherd, G.M.G. (2015). The neocortical circuit: themes and variations. Nat. Neurosci. *18*, 170–181.

Hawkins, J., and Blakeslee, S. (2004). On intelligence (Times Books).

Heeger, D.J. (2017). Theory of cortical function. Proc. Natl. Acad. Sci. USA *114*, 1773–1782.

Heindorf, M., Arber, S., and Keller, G.B. (2018). Mouse Motor Cortex Coordinates the Behavioral Response to Unpredicted Sensory Feedback. Neuron *99*, 1040–1054.e5.

Houde, J.F., and Jordan, M.I. (1998). Sensorimotor adaptation in speech production. Science *279*, 1213–1216.

Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. J. Physiol. *148*, 574–591.

Hubel, D.H., and Wiesel, T.N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J. Neurophysiol. *28*, 229–289.

Inoue, M., Uchimura, M., and Kitazawa, S. (2016). Error signals in motor cortices drive adaptation in reaching. Neuron *90*, 1114–1116.

Jordan, M.I., and Rumelhart, D.E. (1992). Forward Models: Supervised Learning with a Distal Teacher. Cogn. Sci. *16*, 307–354.

Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, A.L., Dhawale, A.K., Kampff, A.R., and Ölveczky, B.P. (2015). Motor cortex is required for learning but not for executing a motor skill. Neuron *86*, 800–812.

Keck, T., Keller, G.B., Jacobsen, R.I., Eysel, U.T., Bonhoeffer, T., and Hübener, M. (2013). Synaptic scaling and homeostatic plasticity in the mouse visual cortex in vivo. Neuron *80*, 327–334.

Keller, G.B., and Hahnloser, R.H.R. (2009). Neural processing of auditory feedback during vocal practice in a songbird. Nature *457*, 187–190.

Keller, A.J., and Martin, K.A.C. (2015). Local Circuits for Contrast Normalization and Adaptation Investigated with Two-Photon Imaging in Cat Primary Visual Cortex. J. Neurosci. *35*, 10078–10087.

Keller, G.B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. Neuron *74*, 809–815.

Keller, A.J., Houlton, R., Kampa, B.M., Lesica, N.A., Mrsic-Flogel, T.D., Keller, G.B., and Helmchen, F. (2017). Stimulus relevance modulates contrast adaptation in visual cortex. eLife *6*, e21589.

Kilgard, M.P., and Merzenich, M.M. (1998). Cortical map reorganization enabled by nucleus basalis activity. Science *279*, 1714–1718.

Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjöström, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. Nature *473*, 87–91.

Ko, H., Cossell, L., Baragli, C., Antolik, J., Clopath, C., Hofer, S.B., and Mrsic-Flogel, T.D. (2013). The emergence of functional microcircuits in visual cortex. Nature *496*, 96–100.

Konishi, M. (1965). The role of auditory feedback in the control of vocalization in the white-crowned sparrow. Z. Tierpsychol. *22*, 770–783.

Körding, K.P., and Wolpert, D.M. (2004). Bayesian integration in sensorimotor learning. Nature *427*, 244–247.

Koster-Hale, J., and Saxe, R. (2013). Theory of mind: a neural prediction problem. Neuron *79*, 836–848.

Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. Trends Neurosci. *36*, 141–151.

Lawson, R.P., Rees, G., and Friston, K.J. (2014). An aberrant precision account of autism. Front. Hum. Neurosci. *8*, 302.

Lawson, R.P., Mathys, C., and Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. Nat. Neurosci. *20*, 1293–1299.

Lee, S., Kruglikov, I., Huang, Z.J., Fishell, G., and Rudy, B. (2013). A disinhibitory circuit mediates motor integration in the somatosensory cortex. Nat. Neurosci. *16*, 1662–1670.

Leinweber, M., Ward, D.R., Sobczak, J.M., Attinger, A., and Keller, G.B. (2017). A Sensorimotor Circuit in Mouse Cortex for Visual Flow Predictions. Neuron *95*, 1420–1432.e5.

Li, N., Daie, K., Svoboda, K., and Druckmann, S. (2016). Robust neuronal dynamics in premotor cortex during motor planning. Nature *532*, 459–464.

Lopes, G., Nogueira, J., Dimitriadis, G., Menendez, J.A., Paton, J.J., and Kampff, A.R. (2016). A robust role for motor cortex. bioRxiv https://doi.org/10.1101/058917.

Lütcke, H., Murayama, M., Hahn, T., Margolis, D.J., Astori, S., Zum Alten Borgloh, S.M., Göbel, W., Yang, Y., Tang, W., Kügler, S., et al. (2010). Optical recording of neuronal activity with a genetically-encoded calcium indicator in anesthetized and freely moving mice. Front. Neural Circuits *4*, 9.

Maffei, L., Fiorentini, A., and Bisti, S. (1973). Neural correlate of perceptual adaptation to gratings. Science *182*, 1036–1038.

Makino, H., and Komiyama, T. (2015). Learning enhances the relative impact of top-down processing in the visual cortex. Nat. Neurosci. *18*, 1116–1122.

Manita, S., Suzuki, T., Homma, C., Matsumoto, T., Odagawa, M., Yamada, K., Ota, K., Matsubara, C., Inutsuka, A., Sato, M., et al. (2015). A Top-Down Cortical Circuit for Accurate Sensory Perception. Neuron *86*, 1304–1316.

Mao, T., Kusefoglu, D., Hooks, B.M., Huber, D., Petreanu, L., and Svoboda, K. (2011). Long-range neuronal circuits underlying the interaction between sensory and motor cortex. Neuron *72*, 111–123.

Markov, N.T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., et al. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. J. Comp. Neurol. *522*, 225–259.

Marr, D. (1982). Vision (MIT Press).

Martin, K.A.C. (1994). A brief history of the "feature detector". Cereb. Cortex *4*, 1–7.

Martins, A.R.O., and Froemke, R.C. (2015). Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. Nat. Neurosci. *18*, 1483–1492.

Meyer, T., and Olson, C.R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. Proc. Natl. Acad. Sci. USA *108*, 19401–19406.

Miri, A., Warriner, C.L., Seely, J.S., Elsayed, G.F., Cunningham, J.P., Churchland, M.M., and Jessell, T.M. (2017). Behaviorally Selective Engagement of Short-Latency Effector Pathways by Motor Cortex. Neuron *95*, 683–696.e11.

Niell, C.M., and Stryker, M.P. (2008). Highly selective receptive fields in mouse visual cortex. J. Neurosci. *28*, 7520–7536.

Niell, C.M., and Stryker, M.P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. Neuron *65*, 472–479.

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. *34*, 171–175.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al. (2014). A mesoscale connectome of the mouse brain. Nature *508*, 207–214.

Packer, A.M., Russell, L.E., Dalgleish, H.W.P., and Häusser, M. (2015). Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. Nat. Methods *12*, 140–146. Published online December 22, 2014.

Pakan, J.M., Lowe, S.C., Dylda, E., Keemink, S.W., Currie, S.P., Coutts, C.A., and Rochefort, N.L. (2016). Behavioral-state modulation of inhibition is

context-dependent and cell type specific in mouse visual cortex. eLife *5*, e14985.

Perin, R., Berger, T.K., and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. Proc. Natl. Acad. Sci. USA *108*, 5419–5424.

Perrett, D.I., Rolls, E.T., and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Exp. Brain Res. *47*, 329–342.

Pinto, L., Goard, M.J., Estandian, D., Xu, M., Kwan, A.C., Lee, S.H., Harrison, T.C., Feng, G., and Dan, Y. (2013). Fast modulation of visual perception by basal forebrain cholinergic neurons. Nat. Neurosci. *16*, 1857–1863.

Polack, P.-O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. Nat. Neurosci. *16*, 1331–1339.

Poort, J., Khan, A.G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G.B., Mrsic-Flogel, T.D., and Hofer, S.B. (2015). Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. Neuron *86*, 1478–1490.

Posner, M.I., and Gilbert, C.D. (1999). Attention and primary visual cortex. Proc. Natl. Acad. Sci. USA *96*, 2585–2587.

Purushothaman, G., Marion, R., Li, K., and Casagrande, V.A. (2012). Gating and control of primary visual cortex by pulvinar. Nat. Neurosci. *15*, 905–912.

Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102–1107.

Rao, R.P.N., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. *2*, 79–87.

Reimer, J., McGinley, M.J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D.A., and Tolias, A.S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. Nat. Commun. *7*, 13289.

Reinhold, K., Lien, A.D., and Scanziani, M. (2015). Distinct recurrent versus afferent dynamics in cortical visual processing. Nat. Neurosci. *18*, 1789–1797.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. Nat. Rev. Neurosci. *2*, 661–670.

Roelfsema, P.R., and Holtmaat, A. (2018). Control of synaptic plasticity in deep cortical networks. Nat. Rev. Neurosci. *19*, 166–180.

Roth, M.M., Dahmen, J.C., Muir, D.R., Imhof, F., Martini, F.J., and Hofer, S.B. (2016). Thalamic nuclei convey diverse contextual information to layer 1 of visual cortex. Nat. Neurosci. *19*, 299–307. Published online December 21, 2015.

Sakata, S., and Harris, K.D. (2009). Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. Neuron *64*, 404–418.

Saleem, A.B., Ayaz, A., Jeffery, K.J., Harris, K.D., and Carandini, M. (2013). Integration of visual motion and locomotion in mouse visual cortex. Nat. Neurosci. *16*, 1864–1869.

Sawtell, N.B., Frenkel, M.Y., Philpot, B.D., Nakazawa, K., Tonegawa, S., and Bear, M.F. (2003). NMDA receptor-dependent ocular dominance plasticity in adult visual cortex. Neuron *38*, 977–985.

Schmitt, L.I., Wimmer, R.D., Nakajima, M., Happ, M., Mofakham, S., and Halassa, M.M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. Nature *545*, 219–223.

Schneider, D.M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature *513*, 189–194.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Sherman, S.M. (2016). Thalamus plays a central role in ongoing cortical functioning. Nat. Neurosci. *19*, 533–541.

Sherrington, C.S. (1900). The Muscular Sense. In Textbook of Physiology, E.A. Sharpey-Schäfer, ed. (London: Edinburgh), pp. 1002–1025.

Simons, D.J., and Chabris, C.F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception *28*, 1059–1074.

Sinha, P., Kjelgaard, M.M., Gandhi, T.K., Tsourides, K., Cardinaux, A.L., Pantazis, D., Diamond, S.P., and Held, R.M. (2014). Autism as a disorder of prediction. Proc. Natl. Acad. Sci. USA *111*, 15220–15225.

Song, S., Sjöström, P.J., Reigl, M., Nelson, S., and Chklovskii, D.B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol. *3*, e68.

Sperry, R.W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. J. Comp. Physiol. Psychol. *43*, 482–489.

Spratling, M.W. (2010). Predictive coding as a model of response properties in cortical area V1. J. Neurosci. *30*, 3531–3543.

Spratling, M.W. (2017). A review of predictive coding algorithms. Brain Cogn. *112*, 92–97.

Stanley, J., and Miall, R.C. (2007). Functional activation in parieto-premotor and visual areas dependent on congruency between hand movement and visual stimuli during motor-visual priming. *34*, 290–299.

Thiele, A., and Bellgrove, M.A. (2018). Neuromodulation of attention. Neuron *97*, 769–785.

Titulaer, M.J., McCracken, L., Gabilondo, I., Iizuka, T., Kawachi, I., Bataller, L., Torrents, A., Rosenfeld, M.R., Balice-Gordon, R., Graus, F., and Dalmau, J. (2013). Late-onset anti-NMDA receptor encephalitis. Neurology *81*, 1058–1063.

Twitchell, T.E. (1951). The restoration of motor function following hemiplegia in man. Brain *74*, 443–480.

Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. Nat. Neurosci. *6*, 391–398.

Vélez-Fort, M., Bracey, E.F., Keshavarzi, S., Rousseau, C.V., Cossell, L., Lenzi, S.C., Strom, M., and Margrie, T.W. (2018). A Circuit for Integration of Head- and Visual-Motion Signals in Layer 6 of Mouse Primary Visual Cortex. Neuron *98*, 179–191.e6.

Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J.A. (2015). Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron *86*, 740–754.

von Helmholtz, H. (1867). Handbuch der physiologischen. Optik (Stuttg.).

von Holst, E., and Mittelstaedt, H. (1950). Das Reafferenzprinzip. Naturwissenschaften *37*, 464–476.

Weinberger, N.M. (2004). Specific long-term memory traces in primary auditory cortex. Nat. Rev. Neurosci. *5*, 279–290.

Wimmer, R.D., Schmitt, L.I., Davidson, T.J., Nakajima, M., Deisseroth, K., and Halassa, M.M. (2015). Thalamic control of sensory selection in divided attention. Nature *526*, 705–709.

Wolpert, D., Ghahramani, Z., and Jordan, M. (1995). An internal model for sensorimotor integration. Science *269* (80-.), 1880–1882.

Wolpert, D.M., Miall, R.C., and Kawato, M. (1998). Internal models in the cerebellum. Trends Cogn. Sci. *2*, 338–347.

Xu, S., Jiang, W., Poo, M.-M., and Dan, Y. (2012). Activity recall in a visual cortical ensemble. Nat. Neurosci. *15*, 449–455, S1-2.

Yang, W., Carrasquillo, Y., Hooks, B.M., Nerbonne, J.M., and Burkhalter, A. (2013). Distinct balance of excitation and inhibition in an interareal feedforward and feedback circuit of mouse visual cortex. J. Neurosci. *33*, 17373–17384.

Yu, A.J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. Neuron *46*, 681–692.

Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. Science *345*, 660–665.

Zingg, B., Hintiryan, H., Gou, L., Song, M.Y., Bay, M., Bienkowski, M.S., Foster, N.N., Yamashita, S., Bowman, I., Toga, A.W., and Dong, H.W. (2014). Neural networks of the mouse neocortex. Cell *156*, 1096–1111.

Zmarz, P., and Keller, G.B. (2016). Mismatch Receptive Fields in Mouse Visual Cortex. Neuron *92*, 766–772.