The Power of Predictions: An Emerging Paradigm for Psychological Research

J. Benjamin Hutchinson[1] and Lisa Feldman Barrett[2,3,4] iD
[1]Department of Psychology, University of Oregon; [2]Department of Psychology, Northeastern University;
[3]Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School; and
[4]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown,
Massachusetts

Abstract
In the last two decades, neuroscience studies have suggested that various psychological phenomena are produced by predictive processes in the brain. When considered together, these studies form a coherent, neurobiologically inspired program for guiding psychological research about the mind and behavior. In this article, we consider the common assumptions and hypotheses that unify an emerging framework and discuss the ramifications of such a framework, both for improving the replicability and robustness of psychological research and for renewing psychological theory by suggesting an alternative ontology of the human mind.

How does a human mind emerge within a human brain as it navigates within an uncertain world while attempting to efficiently regulate its body within hard biological constraints? Recently, a hypothesis has emerged that offers a possible answer: Actions, and their accompanying mental events, begin as top-down representations in the brain, fashioned from past experiences that are tested against the state of the world (see Table 1). This is not a new idea (see the Appendix), but increasingly, research in a variety of different psychological domains is testing this core hypothesis, largely along parallel trajectories.

In this article, we suggest that their integration offers a counterintuitive but coherent, neurobiologically plausible scientific paradigm that has implications for guiding psychological research. This family of research approaches is deeply rooted in various forms of *predictive coding* (e.g., A. Clark, 2013; Rao & Ballard, 1999; Spratling, 2017), in which prediction signals, as representations constructed from past experiences, are compared with incoming sensory information to form prediction errors; prediction errors can be encoded and learned to update stored experience, which is then available for use in future predictions. Such approaches also include the *Bayesian-brain* approach

(e.g., Vilares & Kording, 2011), which assumes that the brain performs (approximate) Bayesian inferences when computing predictions and prediction errors; *belief propagation* (e.g., Lochmann & Deneve, 2011), which proposes that predictions are anticipatory-cause explanations for sensations that are mapped, inversely, to those sensations; and *active inference* (e.g., Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017), which hypothesizes that the brain's model of how sensations are caused is constrained by the need to minimize the cost of prediction error.

Here, we first consider key elements of this overarching "predictive-processing" research program, discussing several examples of its utility for psychological science. We then discuss its potential to improve the robustness and replicability of psychological research, as well as its potential to offer unintuitive but powerful hypotheses that explain cognitions, emotions, perceptions, and actions as emerging from a smaller set of common computational ingredients.

Corresponding Author:
Lisa Feldman Barrett, Northeastern University, Department of Psychology, 360 Huntington Ave., Boston, MA 02115
E-mail: l.barrett@northeastern.edu

Table 1. Examples of Theory Building and Research Guided by a Predictive-Processing Framework

Psychological domain	Example reference
Sensation and perception	
Vision	Rao & Ballard, 1999
Audition	Carbajal & Malmierca, 2018
Somatosensory	Adams, Shipp, & Friston, 2013
Olfaction	Zelano, Mohanty, & Gottfried, 2011
Taste	Gardner & Fontanini, 2014
Interoception	Barrett & Simmons, 2015
Memory	Hindy, Ng, & Turk-Browne, 2016
Language	Kuperberg & Jaeger, 2016
Attention	Feldman & Friston, 2010
Emotion	Barrett, 2017a
Mood	J. E. Clark, Watson, & Friston, 2018
Reward	Schultz, 2016
Social cognition	Tamir & Thornton, 2018
Motor action	Shadmehr, Smith, & Krakauer, 2010
Depression	Barrett, Quigley, & Hamilton, 2016
The self	Seth & Tsakiris, 2018
Social pressure	Theriault, Young, & Barrett, 2019
Words as context	Lupyan & Clark, 2015

Predictive Processing: Some Core Hypotheses

A variety of research findings that seem counterintuitive within psychology's traditional framework (Fig. 1a) are plausibly explained by predictive-processing accounts of the human mind and brain (Fig. 1b), as discussed in a growing literature (e.g., Barrett, 2009, 2017a, 2017b; A. Clark, 2013, 2016; Friston et al., 2017; Hohwy, 2013; Keller & Mrsic-Flogel, 2018; Koster-Hale & Saxe, 2013). This research program is united by a core computational hypothesis: A brain is continually running an internal model of an animal's world. The model is generative, meaning that past experiences can be recombined in novel ways as they are remembered. Unlike psychology's traditional framework, in which perception and action are separate processes with one causing the other, the predictive-processing approach suggests that perception and action are united by the brain's internal model in its effort to efficiently navigate its body in the world. Efficient navigation entails predictively controlling the autonomic nervous system, the endocrine system, and other internal systems to anticipate the needs of the body in the service of upcoming motor actions (Sterling, 2012; Sterling & Laughlin, 2015). Consequently, the model predicts, or infers, the sensory inputs that are expected to derive from those movements, from which perceptions emerge (e.g., A. Clark, 2013; Keller & Mrsic-Flogel, 2018).

According to this account, neural predictions are a continuously changing filter through which sensory inputs are processed, influencing the relevance of those inputs and effectively deciding which sensory features warrant further processing. This filtering is thought to be experienced as effortless, without a sense of personal agency or deliberation and without a specific mechanism for appraising self-relevance per se (as suggested by Ellsworth & Scherer, 2003). Furthermore, the brain's internal model is refined on the basis of comparisons with incoming sensory information from the body and the world; that is, its intrinsic activity can be either confirmed or modified by comparison with sensory input from the world and in the body (Raichle, 2015). In effect, by registering deviations from its internal model, the brain establishes whether and how to spend energy resources to act, and correspondingly, learn any unanticipated sensory inputs to improve the internal model.

This emerging predictive-processing research program offers innovative and important hypotheses for psychological science, many of which are scattered throughout diverse literatures and growing number of summaries and reviews. Here, to illustrate the program's scientific utility, we focus on two hypotheses: (a) Single mental events do not arise in a vacuum but are temporally dependent on prior events, and (b) energy regulation, plus its affective consequences, are core features of all psychological phenomena, not just those that are emotional or involve fight or flight.

Temporal dependence of mental events

Psychologists and neuroscientists have known for some time that the brain processes information in a temporally dependent fashion, such that responses to incoming sensory inputs are conditional on current activity. For example, numerous findings suggest that perceptual processing on a current experimental trial depends on an internal state from a previous trial or is otherwise informed by the past (e.g., Fischer & Whitney, 2014; Kok, Mostert, & de Lange, 2017; St. John-Saaltink, Kok, Lau, & de Lange, 2016; Van de Cruys, Vanmarcke, Van de Put, & Wagemans, 2017). This dependence on memory occurs even when experimental tasks do not require participants to explicitly "remember" past events, suggesting that such an internal model can influence perceptual processing implicitly.

Studies such as these suggest specific hypotheses for understanding how perceptions are formed. One such hypothesis is that, in certain circumstances, perceptual reports will resemble a combination of a perceiver's
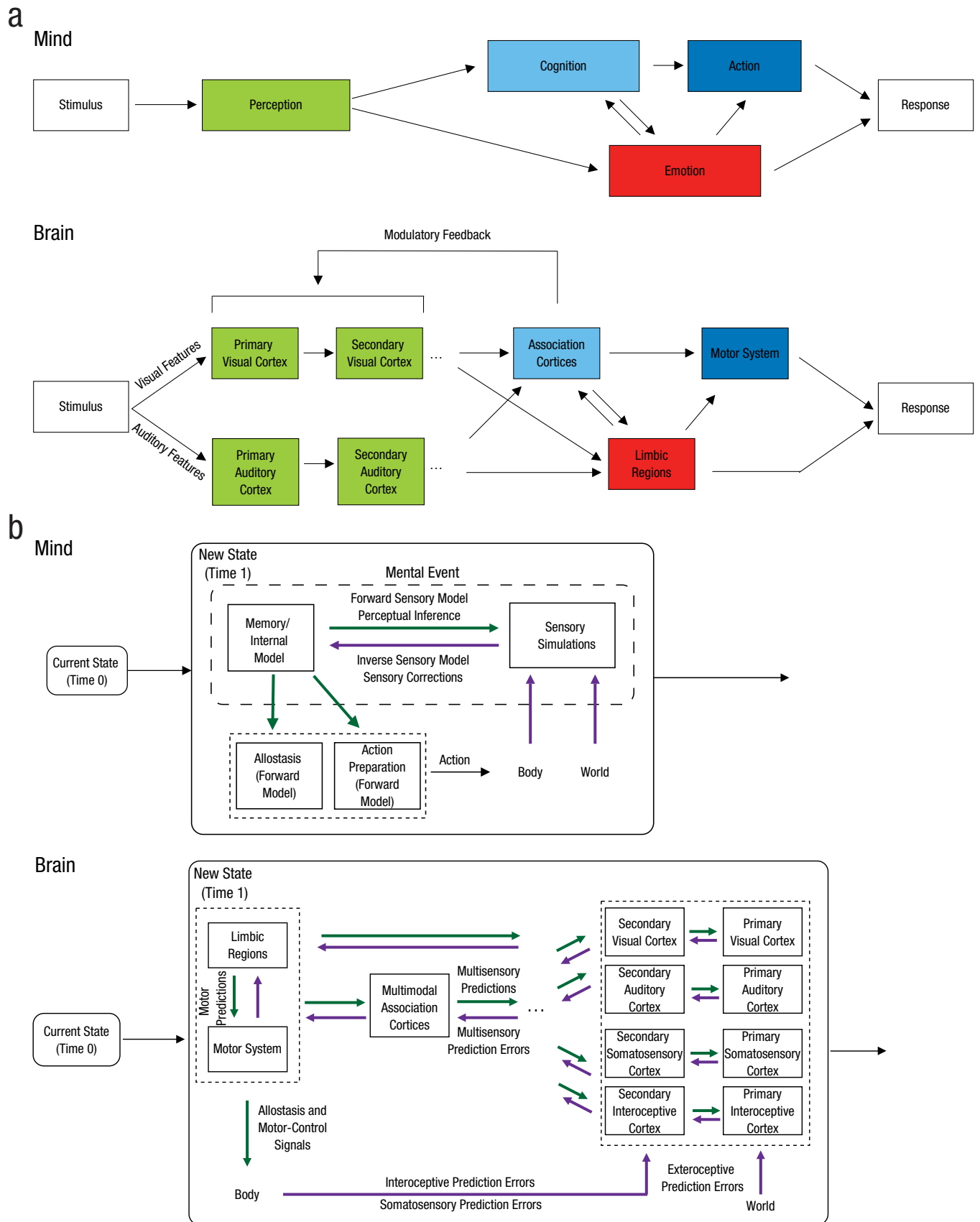
a

Mind

Brain

b

Mind

Brain

Fig. 1. *(continued on next page)*

Fig. 1. Psychology's representational framework compared with the emerging predictive processing framework. Since emerging in the mid-1800s, psychological science has been guided by what is called a representational framework (Keller & Mrsic-Flogel, 2018), in which sensory neurons are thought to create a representation of features in the world (i.e., the perception of a stimulus), which is then passed, like a baton in a relay race, to other parts of the brain that appraise it (i.e., cognition or emotion) and then respond. In the representational framework (a), internal representations and their outputs—the mental events you experience (thoughts, feelings) and the actions you take (behaviors)—are the result of sensory inputs (i.e., stimuli) from the world. Past experiences modulate these stimulus-response links. The colors of the boxes denote correspondence between proposed mental faculties and their mapping to the brain (green corresponds to sensation and perception, light blue to cognition, red to emotion, and dark blue to action). The emerging predictive-processing research paradigm (b) offers a counterintuitive correction, turning the implied causality on its head. Your experience of the world and your action in that world derive from an active, constructive process driven by your brain's internal model. Your brain starts with current conditions and "remembers" trajectories of prior experiences, projecting itself forward in time, assembling multiple competing prediction signals that prepare the body to move (green arrows; e.g., what muscles did I move the last time I was in a situation that is similar to this one?). Copies of these motor commands are thought to modulate the ongoing firing of sensory neurons, inferring the sensory consequences of those movements, thereby simulating some future state of the body and the world (green arrows; e.g., the last time I was in a similar situation and I prepared to move my body in a similar way, what did I see next?). If a brain's internal model are hypotheses about the future state of things, then incoming sensory inputs are the data used to test those hypotheses (purple arrows). The key structural hypothesis is that internal representations (prediction signals) and learning signals (prediction errors) are propagated by neurons arranged in a loose hierarchy (discussed in more detail in Fig. 2). It is hypothesized that when there is a mismatch between a prediction signal at a given level in the hierarchy and information being passed from a lower level, the neurons in question have the opportunity to change their pattern of firing to capture the unexpected input; this is how prediction error is thought to propagate up the processing hierarchy to modify the internal model in the moment and to optimize future predictions. When there is no mismatch, the prediction signal at a given level in the hierarchy is already firing in a way that represents incoming information. Once prediction errors are sufficiently minimized, these "inferences" become the brain's causal account of what caused the sensations in the first place, effectively categorizing the sensations so that they are meaningful.

internal model (i.e., the prediction) and a stimulus, rather than just the stimulus per se. For example, a number of studies show that people are experientially blind when presented with an apparently random selection of black and white blobs (*Mooney images*) that are actually visually degraded versions of regular images. These images are then perceived as coherent after exposure to the natural source images that were used to make them (e.g., Van de Cruys et al., 2017). Such findings suggest that perceptual experience is sculpted by an internal model after, but not before, exposure to the source images (also see *Kanizsa illusions*; Kok, Bains, van Mourik, Norris, & de Lange, 2016). In another example of combination, when participants are asked to adjust the color of a well-known object such as a banana to be gray, they tend to overadjust (i.e., they adjust it to be more bluish-gray, which combined with the predicted yellow produces a subjective judgment of gray; Hansen, Olkkonen, Walter, & Gegenfurtner, 2006). Consistent with these findings, research has shown that perception is also facilitated (e.g., breaking through visual suppression) when features of the internal model anticipate incoming stimulus features, creating perceptual fluency (Chanes, Wormwood, Betz, & Barrett, 2018).

The role of energy regulation in mental events

One important development within the predictive-processing framework comes from several recent studies in which computational hypotheses about predictive dynamics have been integrated with anatomical models of information flow in the cerebral cortex (e.g., Barrett,

2017a; Barrett & Simmons, 2015; Chanes & Barrett, 2016; Keller & Mrsic-Flogel, 2018), the hippocampus and medial temporal lobe (e.g., Gravina & Sederberg, 2017), the cerebellum (Schlerf, Ivry, & Diedrichsen, 2012), and the striatum and other subcortical regions comprising the dopaminergic reward system (Schultz, 2016). One attempt offers hypotheses about the flow of predictions and prediction-error signals within the brain using architectural models of cortical connections (Barrett & Simmons, 2015; Chanes & Barrett, 2016; Friston et al., 2017; see Fig. 2). A central implication arising from this approach is that the brain's internal model is centrally concerned with energy regulation, making energetics relevant for all mental events, not just those involving emotion or fight-or-flight responses. This hypothesis is consistent with the suggestion that predictive processing is ideal for minimizing free energy (Friston et al., 2017) or uncertainty, so that information in the brain is being represented as efficiently as possible.

A key structural hypothesis of predictive-processing accounts is that internal representations (prediction signals) and learning signals (prediction errors) are propagated by neurons arranged in a loose hierarchy (Barbas, 2015; Mesulam, 1998). For more than 30 years, research has used anatomical features describing the arrangement and connectivity of neurons to predict information flow across a loose cortical hierarchy (Barbas, 2015). This hierarchy suggests how a generative internal model built and constrained by sensory events in the periphery might be implemented in the brain. Specifically, neurons within cortical areas that are higher in this predictive hierarchy send prediction signals to neurons in regions that are lower in the hierarchy, with prediction errors flowing "upward" (see Fig. 2). An
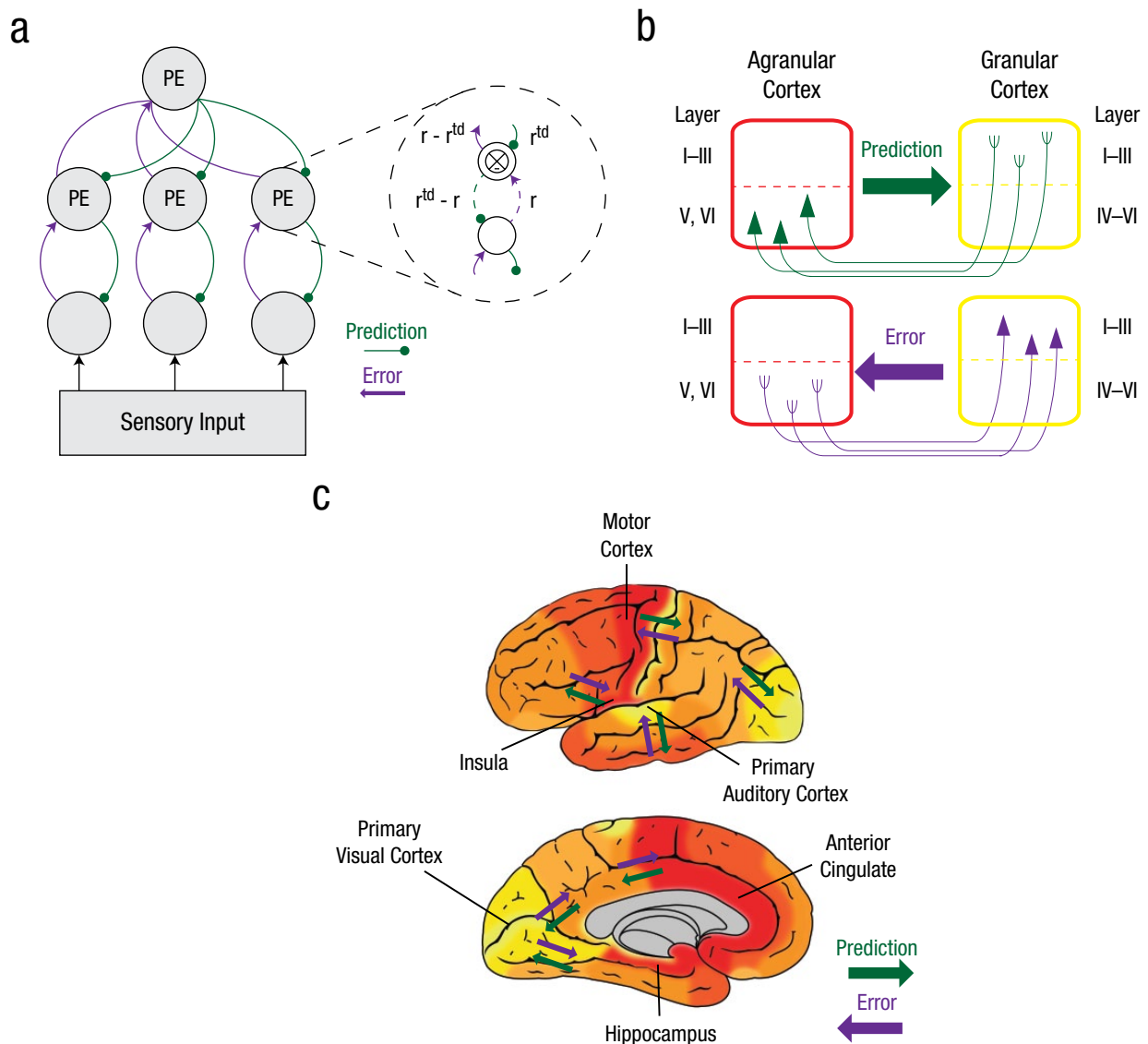
Fig. 2. Predictive processing from computation to whole-brain dynamics. In the original predictive coding model (a; Rao & Ballard, 1999), sensory input drives feed-forward error signals, which are processed by predictive estimators (PEs). PE modules consist of neurons whose activity (r) is compared with top-down activity from higher levels (i.e., prediction, r^{td}) and the difference of which is propagated forward to the next level as error ($r - r^{td}$). The flow of prediction and prediction-error signals between cortical columns based on cortical lamination gradients is shown in (b). Using anterograde and retrograde tracers, Barbas and colleagues (see Barbas, 2015, for a review) showed that the relative difference in laminar structure between two communicating cortical columns predicts whether the information flow is a feedback (prediction) or a feed-forward (prediction-error) signal. Prediction signals (green) originate in the deep layers (Layers V and VI) of less differentiated cortical areas (such as agranular cortex with undifferentiated Layers II and III and without a Layer IV, as depicted in the red column) and terminate in superficial layers of areas with a more developed laminar structure (such as dysgranular cortices with differentiated Layers II and III and a rudimentary Layer IV or granular cortices with differentiated Layers II and III and a well-defined Layer IV, depicted in the yellow column). Prediction-error signals (in purple) flow in the other direction, originating in the superficial layers (II and III) with more laminar differentiation and terminating in middle deep layers (V and VI) of areas with less differentiated laminar architecture. This structural model successfully predicts the flow of information in frontal, temporal, and parietal cortices in experiments with monkeys and cats (see Barbas, 2015, for a review). The estimates of flow of prediction and prediction-error signals (c) are based on whole-brain cortical granularity data (von Economo, 2009; von Economo & Koskinas, 2008). Predictions flow from cortical regions with less laminar differentiation to regions with increasing laminar differentiation (e.g., from limbic cortices to motor, interoceptive, and primary somatosensory, auditory, and visual cortices). Researchers are still investigating many of the anatomical and computational details, such as whether each individual neuron is capable of coding for internal representations (i.e., predictions) and comparing those predictions with incoming inputs from lower in the hierarchy (i.e., prediction errors) or whether predictions and prediction errors are coded by different neurons in the cortex.

important implication of this hypothesis is that limbic cortices, such as portions of the cingulate cortex, orbitofrontal cortex, entorhinal cortex, and anterior insula, along with the hippocampus, have anatomical features that place them at the top of this predictive architecture (Barbas, 2015) and consequently provide the initial representations of prediction signals that propagate throughout the cortex (Barrett & Simmons, 2015; Chanes & Barrett, 2016).

This limbic ensemble, via a series of connections to the hypothalamus and throughout the brain stem, is also responsible for regulating the body's global energy budget via control of the autonomic nervous system, the neuroendocrine and neuroimmune systems, and the other systems of the body's internal milieu (for a review of connectivity and evidence of intrinsic functional connectivity, see Kleckner et al., 2017). These limbic cortices are thought to regulate the body by anticipating its needs and attempting to meet those needs before they arise (Ganzel, Morris, & Wethington, 2010), a process called *allostasis* (Sterling, 2012), placing efficient energy regulation and metabolism at the core of the brain's internal model. Because learning new information has an energy cost, energetics may effectively determine the value of prediction errors that update the model (Theriault, Young, & Barret 2019). Limbic regions are still considered by many psychologists to be the most reactive parts of the brain, as home to emotions and therefore in need of control. But in this predictive-processing framework, limbic cortices, plus the hippocampus, are the source of prediction signals, driving action and perception in an inferential way concerned with energetics, not just during episodes of emotion, but during all mental events.

Several lines of research are broadly consistent with the hypothesis that metabolism and energetics are core concerns in the construction of action and mental events in the brain. The principles of neural design indicate that anticipatory regulation of the body is advantageous for energy efficiency (Sterling & Laughlin, 2015), consistent with evidence that this efficiency is a major constraint on brain evolution. As energetics determine an animal's internal state, they can influence contexts for learning (Yu & Dayan, 2005) and memory (Bouton, 2019). Energetics might play an underappreciated role in sensory processing, because there are long-range connections between limbic cortices (e.g., anterior cingulate cortex, or ACC) and primary sensory regions (e.g., primary visual cortex/V1; Zhang et al., 2014), as well as evidence that the ACC sends prediction signals to V1 (Leinweber, Ward, Sobczak, Attinger, & Keller, 2017). These prediction signals appear to be the source of neural firing in V1 after retinal lesions

and subsequent visual deprivation (Keck et al., 2013). Such evidence is consistent with other findings that a substantial fraction of activity in the visual cortex is not related to visual input per se (see references in Keck et al., 2013) but instead may be due to prediction signals (Liang, Mouraux, Hu, & Iannetti, 2013). Indeed, consistent with such predictive influence, research suggests that the majority of synapses in primary visual cortex originate from top-down sources (e.g., Sillito & Jones, 2002).

Furthermore, both structural- and functional-connectivity data provide additional evidence that limbic cortices are contained in two core intrinsic brain networks—the default mode and salience networks—that are implicated in various psychological phenomena, including memory, perception, attention, social affiliation, pain, empathy, reward, addiction, stress, emotion, and decision making (reviewed in Kleckner et al., 2017). These findings suggest that allostasis, in the service of efficient energy regulation, is a fundamental feature of the brain's internal model as supported by these two networks. Such findings suggest a provocative hypothesis for future research: While supporting a given psychological function in a given moment, these networks are simultaneously maintaining or attempting to restore allostasis.

Practical Implications: The Robustness and Replicability of Psychological Research

Both hypotheses we discussed—the temporal dependence of mental events and the importance of energetics in mental life—have important implications for the robustness and replicability of psychological studies. Some of these observations have been made before, but a predictive-processing framework helps to further unify and motivate them. The approach also suggests points for enhancing robustness and replicability when designing future experiments.

Temporal dependence of mental events

The predictive-processing framework suggests that the brain is constantly attempting to impose predictability over multiple time scales. Interestingly, this tendency is at odds with traditional laboratory experiments designed using unpredictable sequences of events, potentially limiting the generalizability of stimulus-response experiments, particularly when moving from the laboratory to the real world. That is, traditional laboratory experiments are typically constructed as random sequences of stimuli to minimize influence across trials. This allows adjacent trials to be treated as

independent, making them suitable for aggregation and traditional statistical analysis. In principle, many scientists would not defend the assumption that the mind works in similarly independent and discrete chunks in time. In practice, however, a participant's response on any given trial is some combination of the participant's internal model and a given stimulus, and ideally both should be modeled to maximize the robustness of scientific findings. Recent research has highlighted the utility of an experimental framework that moves beyond discrete individual events toward attempting to understand brain and behavior in terms of continuous, temporally dependent processes (Huk, Bonnen, & He, 2018).

The role of energy regulation in mental events

Psychological science routinely makes reference to "affective stimuli" and "rewards" or "threats" as if such qualities are embedded in objects and events in the world rather than features that arise from transactions between a person's current state and those objects and events. This is more than a trivial distinction: The features of an experimental stimulus that are deemed salient by an experimenter may not be experienced thus by a participant, and other features that are deemed psychologically impotent by an experimenter may be salient to participants (or to certain participants). If the brain runs an internal model that is concerned with efficient energy regulation, and if the processing of stimuli are conditioned on that model, then the state of a person's energy balance, and all the factors that can influence that balance (such as amount of sleep; ingestion of caffeine, sugar, or nicotine; and degree of hydration) can influence how the brain processes information and, correspondingly, task performance. For example, people perform better on a variety of cognitive tasks when they are tested at an optimal moment in their circadian cycle than they do at nonoptimal times (Yoon, May, & Hasher, 2012). If you schedule participants at a time of day that takes their circadian rhythm into account, then you will reduce what appears to be random variation which should increase replicability.

Even if any individual source of energetic influence is small, the overall impact on the robustness and replicability can be substantial, particularly with small sample sizes (Benjamin et al., 2018). These factors, when not measured, can add variance in an experiment that will appear as noise in the measurements, increasing the likelihood of Type I and Type II errors. That is, only the most potent effects will replicate, compounding the

impact of any Type II errors. Statistically controlling for such influences may remove meaningful variance that is better examined, again enhancing the likelihood of Type II errors. Additionally, when analyses do not control for such sources of variance, Type I errors may be inflated. When sample sizes are low, statistical power suffers, and any observed effect may be proportionally driven more by these factors (as power is reduced, so is the proportion of any observed effect that is due to reproducible variance). Thus, observed effects will be less likely to replicate as a consequence (Szucs & Ioannidis, 2017).

Conceptual Implications for Psychological Theory

The predictive-processing framework provides a principled approach for unifying psychological phenomena into a common explanatory framework with a shared vocabulary for theory building, offering several novel conceptual implications. One implication is that phenomena that we think of as separate processes arising from separate brain systems, such as automatic and controlled processing, could be conceptualized as different modes of whole-brain function (prediction based and prediction-error based).

According to this interpretation, the ubiquity of dual-process theories in psychology (e.g., Evans & Stanovich, 2013) may reflect a single underlying distinction between modes of prediction and prediction-error processing in the brain. For example, in the context of controlled versus automatic processing, on a given experimental trial, a participant's brain will launch a set of prediction signals, which scientists interpret as evidence for a rapid, automatic, and effortless process. If the stimulus is unexpected (such as when trials are randomized so stimuli remain unpredicted) or when a nonprepotent response is required, then participants will appear to correct the automatic process with a more deliberate, controlled, and effortful process. Learning and practice effects during the course of an experiment might be understood as the consequence of a brain successfully updating its internal model, just as inhibiting a prepotent response may reflect an updating of the model with prediction error. Participants thus appear to behave on the basis of rapid, automatic responses (predictions) followed by a more effortful choice that corrects them (updating with prediction error). What appears to be separate automatic and controlled processes in the mind may actually be different modes of operation for the brain: one emphasizing a participant's internal model (prediction) and the other emphasizing correction and learning (prediction error).

One consequence of this two-mode perspective is that traditional experiments, designed as independent sequences of stimulus and response, effectively sever the contingencies between one moment and the next. That is, the brain's predictions will almost certainly be wrong on many trials, forcing it into a mode that favors processing prediction error (e.g., driven by the stimuli), when in the real world the dynamics may frequently favor prediction (e.g., people with neurotypical brains are not continually in state of high prediction error). Thus, standard randomized designs encourage oversampling of what might be an unnatural state of error processing. Instead, psychological causation might be better measured and modeled as temporally extended, probabilistic sequences of brain states (Barrett, 2009). Indeed, a relatively unexplored, but potentially rich line of research would be to explore traditional psychological phenomena in predictable versus unpredictable experimental contexts (e.g., repeating vs. randomly generated sequences of trials).

In addition, a predictive-processing framework has the conceptual implication that certain psychological phenomena, such as memory, are actually ingredients in all psychological events, even those that do not appear to strictly involve memory. For example, attention has long been operationalized in terms of its influence on perceptual processing; however, a growing amount of evidence suggests that attention appears to be meaningfully guided by what one has encountered in the past (e.g., Hutchinson & Turk-Browne, 2012). Furthermore, neural signals in brain regions historically associated with memory, such as hippocampal activity, are also systematically observed during tasks of perception and attention (Aly, Ranganath, & Yonelinas, 2013; Aly & Turk-Browne, 2016). Predictions based on past experience influence relatively low-level perceptual processes and activity in primary visual cortex (den Ouden, Kok, & de Lange, 2012; O'Callaghan, Kveraga, Shine, Adams, & Bar, 2017; Rao & Ballard, 1999). Here, we suggest that such processes are pervasive and similar dynamics take place across a distributed cortical hierarchy where the interplay between an internal model and feedback from the external world, as well as the interoceptive state of the body, guide learning at multiple time scales (e.g., Honey, Newman, & Schapiro, 2017).

The prediction framework also leads us to speculate that affect is part of every psychological phenomenon, even those that are not explicitly emotional. If allostasis and energetics are key features of the brain's internal model, then so too are the predicted sensory consequences of those processes, referred to as *interoception* (Craig, 2014). Interoception is usually experienced in a low-dimensional form as the affective properties of valence and arousal (Barrett & Bliss-Moreau, 2009), suggesting that all psychological events exist in affective space. Valence and arousal might be better thought of as properties of consciousness, rather than properties of emotional episodes per se, as suggested by a number of philosophers. Consistent with this speculation, research shows that all words have affective connotations (Osgood, May, & Mirron, 1975) and even putatively "neutral" objects are experienced with subtle affective features (Lebrecht, Bar, Barrett, & Tarr, 2012). This insight, if correct, calls into question all hypotheses that cast cognition and emotion in a battle for the control of behavior, or as two separate processes that interact to produce behavior, and suggests that the concept of "rationality" must be redefined as something other than the absence of affect.

When taken together, these implications suggest several highly speculative ideas. First, many phenomena that go by different psychological names and up to now have been assumed to arise from distinct processes—such as memory, perception, and emotion—may actually be better understood as arising from a smaller set of common computational building blocks, with prediction-related processing at the core. Speculating even further, we surmise that predictive optimization may even be implemented at the level of species-wide neural development. Recently, it has been proposed that many brains within the animal kingdom are structured to function via prediction and correction (Sterling & Laughlin, 2015). Consistent with the notion of "experience-expectant" processes (i.e., information storage in anticipation of particular life periods of experience rather than in response to them; Greenough, Black, & Wallace, 1987), efficiency based on temporal regularities in the environment might even be built into the evolution of the brain itself.

Conclusions

The scientific story of predictive processing is still evolving, but many researchers believe this approach has ignited a paradigm shift in neuroscience. Here, we have proposed that this paradigm shift has important implications for psychological science, both in theory and in practice. The mind is a computational moment in a brain that creates a temporally continuous trajectory of neural activity, tasked with regulating a body in the world. Appreciating this perspective may improve the quality of our scientific findings and also offer opportunities for new discoveries about the nature of the human mind.

Appendix

Standing on the shoulders of giants: the rich history of internal models

Since psychology's emergence as an empirical science in the mid-1800s, research has largely relied on a stimulus (S) → organism (O) → response (R) model of the mind (Danziger, 1997). Yet this model has consistently been questioned through the ages (Barrett, 2009). Kant (1781/1929) proposed that we experience the world through a web of our own concepts, as did the seventh century Buddhist philosopher Dharmakīrti and the 11th century Islamic philosopher Ibn al-Haytham. At the dawn of psychological science, von Helmholtz's (1860/1924) idea of unconscious inference and Dewey's (1896) criticism of the reflex-arc concept cautioned against an SOR model of the mind. Decades later, Craik's (1943) and Johnson-Laird's (1983) internal models, and Tolman's cognitive maps (1948) all proposed, in different ways, that internal mental representations such as beliefs and knowledge influence subsequent perception and action at least as much, if not more than, the other way around. Early studies on attention posited that the degree to which a stimulus elicited an orienting response was related to how different it was from a "nervous model" of that stimulus based on past experience (Sokolov, 1963). The hypothesis that internal representations are hypotheses that play a key role in perception and action formed the basis of the cognitive revolution (e.g., Gregory, 1980; Neisser, 1967), and within social psychology, implicit attitudes, stereotyping, and prejudice are predicated on the idea that information inside the head shapes experience of and action in the world. A major benefit of predictive-processing-related accounts of psychological phenomena is that they are usually embedded within an anatomical or a computational framework, allowing, for the first time, a more direct assessment of their common (or distinct) implementations and consequences.

Recommended Reading

Barrett, L. F. (2017b). (See References). A broad integration of predictive processing, neuroanatomy, and psychological science that fashions a multilevel constructionist account of emotional phenomena and points toward a unified science of mind, action, and context.
Clark, A. (2013). (See References). An excellent summary and integration of hierarchical predictive-coding approaches that highlights key aspects of a predictive-processing approach, as well as its challenges.
den Ouden, H. E. M., Kok, P., & de Lange, F. (2012). (See References). An accessible overview of neural prediction errors across many domains of psychological research that outlines key empirical studies describing the origins, computational principles, and behavioral consequences of prediction errors throughout the brain.

Hohwy, J. (2013). (See References). An accessible introduction to a predictive-processing approach that outlines key concepts and speculates on how a predictive-processing approach can be used to understand a range of psychological phenomena.
Tamir, D. I., & Thornton, M. A. (2018). (See References). An examination of social inference through the lens of predictive processing.

Action Editor

Randall W. Engle served as action editor for this article.

ORCID iD

Lisa Feldman Barrett https://orcid.org/0000-0003-4478-2051

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure & Function*, *218*, 611–643.
Aly, M., Ranganath, C., & Yonelinas, A. P. (2013). Detecting changes in scenes: The hippocampus is critical for strength-based perception. *Neuron*, *78*, 1127–1137. doi:10.1016/j.neuron.2013.04.018
Aly, M., & Turk-Browne, N. B. (2016). Attention promotes episodic encoding by stabilizing hippocampal representations. *Proceedings of the National Academy of Sciences, USA*, *113*, E420–E429. doi:10.1073/pnas.1518931113
Bar, K.-J., de la Cruz, F., Schumann, A., Koehler, S., Sauer, H., Critchley, H., & Wagner, G. (2016). Functional connectivity and network analysis of midbrain and brainstem nuclei. *NeuroImage*, *134*, 53–63.
Barbas, H. (2015). General cortical and special prefrontal connections: Principles from structure to function. *Annual Review of Neuroscience*, *38*, 269–289.

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition & Emotion, 23,* 1284–1306. doi:10.1080/02699930902985894

Barrett, L. F. (2017a). *How emotions are made: The secret life the brain.* New York, NY: Houghton Mifflin Harcourt.

Barrett, L. F. (2017b). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12,* 1–23. doi:10.1093/scan/nsw154

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 41, pp. 167–218). doi:10.1016/S0065-2601(08)00404-8

Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1708), Article 20160011. doi:10.1098/rstb.2016.0011

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience, 16,* 419–429. doi:10.1038/nrn3950

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., . . . Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour, 2*(1), Article 6. doi:10.1038/s41562-017-0189-z

Bouton, M. E. (2019). Extinction of instrumental (operant) learning: Interference, varieties of context, and mechanisms of contextual control. *Psychopharmacology, 236,* 7–19.

Carbajal, G. V., & Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: From the subcortical roots to cortical deviance detection. *Trends in Hearing, 22.* doi:10.1177/2331216518784822

Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences, 20,* 96–106.

Chanes, L., Wormwood, J. B., Betz, N., & Barrett, L. F. (2018). Facial expression predictions as drivers of social perception. *Journal of Personality and Social Psychology, 114,* 380–396.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral & Brain Sciences, 36,* 281–253.

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

Clark, J. E., Watson, S., & Friston, K. J. (2018). What is mood? A computational perspective. *Psychological Medicine, 48,* 2277–2284. doi:10.1017/S0033291718000430

Craig, A. D. (2014). *How do you feel? An interoceptive moment with your neurobiological self.* Princeton, NJ: Princeton University Press.

Craik, K. (1943). *The nature of explanation.* Cambridge, England: Cambridge University Press.

Danziger, K. (1997). *Naming the mind: How psychology found its language.* London, England: Sage.

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motiva-

tion. *Frontiers in Psychology, 3,* Article 548. doi:10.3389/fpsyg.2012.00548

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review, 3,* 357–370. doi:10.1037/h0070405

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Sherer, & H. Hill Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). Oxford, England: Oxford University Press.

Evans, J. S., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8,* 223–241. doi:10.1177/1745691612460685

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4,* Article 215. doi:10.3389/fnhum.2010.00215

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience, 17,* 738–743. doi:10.1038/nn.3689

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation, 29,* 1–49. doi:10.1162/NECO_a_00912

Ganzel, B. L., Morris, P. A., & Wethington, E. (2010). Allostasis and the human brain: Integrating models of stress from the social and life sciences. *Psychological Review, 11,* 134–174. doi:10.1037/a0017773

Gardner, M. P., & Fontanini, A. (2014). Encoding and tracking of outcome-specific expectancy in the gustatory cortex of alert rats. *The Journal of Neuroscience, 34,* 13000–13017.

Gravina, M. T., & Sederberg, P. B. (2017). The neural architecture of prediction over a continuum of spatiotemporal scales. *Current Opinion in Behavioral Sciences, 17,* 194–202. doi:10.1016/j.cobeha.2017.09.001

Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). Experience and brain development. *Child Development, 58,* 539–559.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 290,* 181–197.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9,* 1367–1368. doi:10.1038/nn1794

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience, 19,* 665–667. doi:10.1038/nn.4284

Hohwy, J. (2013). *The predictive mind.* Oxford, England: Oxford University Press.

Huk, A., Bonnen, K., & He, B. J. (2018). Beyond trial-based paradigms: Continuous behavior, ongoing neural activity, and natural stimuli. *The Journal of Neuroscience, 38,* 7551–7558. doi:10.1523/JNEUROSCI.1920-17.2018

Hutchinson, J. B., & Turk-Browne, N. B. (2012). Memory-guided attention: Control from multiple memory systems. *Trends in Cognitive Sciences, 16,* 576–579. doi:10.1016/j.tics.2012.10.003

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Kant, I. (1929). *Critique of pure reason* (N. Kemp Smith, Trans). New York, NY: Palgrave Macmillan. (Original work published 1781)

Keck, T., Keller, G. B., Jacobsen, R. I., Eysel, U. T., Bonhoeffer, T., & Hübener, M. (2013). Synaptic scaling and homeostatic plasticity in the mouse visual cortex in vivo. *Neuron, 80*, 327–334. doi:10.1016/j.neuron.2013.08.018

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron, 100*, 424–435. doi:10.1016/j.neuron.2018.10.003

Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., . . . Barrett, L. F. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour, 1*, Article 0069. doi:10.1038/s41562-017-0069

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology, 26*, 371–376. doi:10.1016/j.cub.2015.12.038

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences, USA, 114*, 10473-10478. doi:10.1073/pnas.1705652114

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron, 79*, 836–848. doi:10.1016/j.neuron.2013.08.020

Kuperberg, G. R., & Jaeger, F. T. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*, 32–59.

Lebrecht, S., Bar, M., Barrett, L. F., & Tarr, M. J. (2012). Micro-valences: Affective valence in "neutral" everyday objects. *Frontiers in Perception Science, 3*, Article 107. doi:10.3389/fpsyg.2012.00107 PMC3328080

Leinweber, M., Ward, D. R., Sobczak, J. M., Attinger, A., & Keller, G. B. (2017). A sensorimotor circuit in mouse cortex for visual flow predictions. *Neuron, 95*, 1420–1432. doi:10.1016/j.neuron.2017.08.036

Liang, M., Mouraux, A., Hu, L., & Iannetti, G. D. (2013). Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. *Nature Communications, 4*, Article 1979. doi:10.1038/ncomms2979

Lochmann, T., & Deneve, S. (2011). Neural processing as causal inference. *Current Opinion in Neurobiology, 21*, 774–781. doi:10.1016/j.conb.2011.05.018

Lupyan, G., & Clark, A. (2015). Words and the world. *Current Directions in Psychological Science, 24*, 279–284. doi:10.1177/0963721415570732

Mesulam, M. M. (1998). From sensation to cognition. *Brain, 121*, 1013–1052.

Neisser, U. (1967). *Cognitive psychology.* New York, NY: Appleton-Century-Crofts.

O'Callaghan, C., Kveraga, K., Shine, J. M, Adams, R. B., Jr., & Bar, M. (2017). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition, 47*, 63–74. doi:10.1016/j.concog.2016.05.003

Osgood, C. E., May, W. H., & Mirron, M. S. (1975). *Cross-cultural universals of affective meanings.* Urbana: University of Illinois Press.

Raichle, M. E. (2015). The restless brain: How intrinsic activity organizes brain function. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1668), Article 20140172. doi:10.1098/rstb.2014.0172

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*, 79–87. doi:10.1038/4580

Schlerf, J., Ivry, R. B., & Diedrichsen, J. (2012). Encoding of sensory prediction errors in the human cerebellum. *The Journal of Neuroscience, 32*, 4913–4922. doi:10.1523/JNEUROSCI.4504-11.2012

Schultz, W. (2016). Dopamine reward prediction-error signalling: A two-component response. *Nature Reviews Neuroscience, 17*, 183–195. doi:10.1038/nrn.2015.26

Seth, A. K., & Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences, 22*, 969–981. doi:10.1016/j.tics.2018.08.008

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience, 33*, 89–108. doi:10.1146/annurev-neuro-060909-153135

Sillito, A. M., & Jones, H. E. (2002). Corticothalamic interactions in the transfer of visual information. *Philosophical Transactions of the Royal Society B: Biological Sciences, 357*(1428), 1739–1752. doi:10.1098/rstb.2002.1170

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual Review of Physiology, 25*, 545–580. doi:10.1146/annurev.ph.25.030163.002553

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition, 112*, 92–97. doi:10.1016/j.bandc.2015.11.003

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior, 106*, 5–15.

Sterling, P., & Laughlin, S. (2015). *Principles of neural design.* Cambridge, MA: MIT Press.

St. John-Saaltink, E. S., Kok, P., Lau, H. C., & de Lange, F. P. (2016). Serial dependence in perceptual decisions is reflected in activity patterns in primary visual cortex. *The Journal of Neuroscience, 36*, 6186–6192. doi:10.1523/JNEUROSCI.4390-15.2016

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology, 15*(3), Article e2000797. doi:10.1371/ journal.pbio.2000797

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences, 22*, 201–212. doi:10.1016/j.tics.2017.12.005

Theriault, J. E., Young, L., & Barrett, L. F. (2019). The sense of should: A biologically-based model of social pressure. *PsyArXiv.* doi:10.31234/osf.io/x5rbs

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*, 189–208.

Van de Cruys, S., Vanmarcke, S., Van de Put, I., & Wagemans, J. (2018). The use of prior knowledge for perceptual inference is preserved in ASD. *Clinical Psychological Science, 6*, 382–393. doi:10.1177/2167702617740955

Vilares, I., & Kording, K. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain.

Annals of the New York Academy of Sciences, 1224, 22–39. doi:10.1111/j.1749-6632.2011.05965.x

von Economo, C. (2009). *Cellular structure of the human cerebral cortex* (L. C. Triarhou Ed. & Trans.). Basel, Switzerland: Karger.

von Economo, C., & Koskinas, G. N. (2008). The 107 cortical cytoarchitectonic areas of Constantin von Economo and Georg N. Koskinas in the adult human brain. In L. C. Triarhou (Ed. & Trans.) *Atlas of cytoarchitectonics of the adult human cerebral cortex*. Basel, Switzerland: Karger.

von Helmholtz, H. (1924). *Hemholtz's treatise on physiological optics*. Menasha, WI: Optical Society of America and Banta Publishing. (Original work published 1860)

Yoon, C., May, C. P., & Hasher, L. (2012). Aging, circadian arousal patterns, and cognition. In D. Park & N. Schwarz (Eds.), *Cognitive aging: A primer* (2nd ed.). New York, NY: Psychology Press.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*, 681–692. doi:10.1016/j.neuron.2005.04.026

Zelano, C., Mohanty, A., & Gottfried, J. A. (2011). Olfactory predictive codes and stimulus templates in piriform cortex. *Neuron, 72*, 178–187. doi:10.1016/j.neuron.2011.08.010

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