# Cellular mechanisms of cooperative context-sensitive predictive inference

Tomáš Marvan [a],[*],[#], William A. Phillips [b],[#]

[a] *Institute of Philosophy, Czech Academy of Sciences (CAS), Czech Republic*
[b] *Faculty of Natural Sciences, University of Stirling, United Kingdom*

## ARTICLE INFO

## ABSTRACT

We argue that prediction success maximization is a basic objective of cognition and cortex, that it is compatible with but distinct from prediction error minimization, that neither objective requires subtractive coding, that there is clear neurobiological evidence for the amplification of predicted signals, and that we are unconvinced by evidence proposed in support of subtractive coding. We outline recent discoveries showing that pyramidal cells on which our cognitive capabilities depend usually transmit information about input to their basal dendrites and amplify that transmission when input to their distal apical dendrites provides a context that agrees with the feedforward basal input in that both are depolarizing, i.e., both are excitatory rather than inhibitory. Though these intracellular discoveries require a level of technical expertise that is beyond the current abilities of most neuroscience labs, they are not controversial and acclaimed as groundbreaking. We note that this cellular cooperative context-sensitivity greatly enhances the cognitive capabilities of the mammalian neocortex, and that much remains to be discovered concerning its evolution, development, and pathology.

## 1. Introduction

Mental life, and indeed our survival, depends upon maximizing valid perceptual inferences and minimizing invalid ones. In this sense, 'prediction error minimization' is a major goal for neural systems. However, it is often taken for granted that minimization of prediction errors subtracts predictions from ascending feedforward inputs such that the signals transmitted up a perceptual hierarchy code for the difference between descending predictions and ascending inputs. As we see it, that is neither necessarily so nor convincingly shown to occur in mammalian neocortex: the necessity of minimizing incorrect (and maximizing correct) perceptual inferences does not imply any specific form of coding.

In this paper, we show that certain classes of pyramidal cells with two points of integration can 'minimize errors of predictive inference' by amplifying relevant and suppressing irrelevant signals without using the error computations typical of predictive coding theories. Furthermore, we will argue that the hypothesis of predictive coding is neurobiologically implausible. Though the notion of subtractive prediction error has become prominent within theories of predictive coding, the evidence and arguments reviewed below imply that predictive inputs

regulate the salience of neuronal signals without changing that for which they code. The cellular mechanisms we briefly review below thus provide a sound neurophysiological basis for prediction error minimization, but not for the use of predictions to change that for which the ascending signals 'code', i.e., that about which they transmit information. Predictions are then seen as changing feedforward transmission by changing which signals are amplified and which are attenuated, not by dynamically changing that for which they code.[1]

As we see it, the conception of cellular cooperative context-sensitive predictive inference outlined in this paper preserves the strengths of previous research on 'predictive processing', even though it is skeptical of the notion of subtractive coding. The framework advocated here builds on discoveries in cellular neuroscience that are acknowledged to be groundbreaking milestones (e.g., by Ramaswamy and Markram, 2015), and which are poised to become mainstream in cognitive and computational neuroscience. These discoveries can be seen as rooting conceptions of inference and prediction in their cellular foundations, while simultaneously contributing to emerging medical and deep learning technologies.

In the following sections, we first state some of our working

---

[1] Throughout this paper, the target of our criticism are accounts of predictive coding which assume that the cells subtractively computing prediction errors switch from signaling the ascending information they are receiving to the *difference* between top-down generated predictions and the ascending information.

assumptions concerning the notions of predictions and predictive perceptual inference (Section 2). In Section 3, we spell out shortcomings of the idea of subtractive prediction error coding, which is often presented as being at the core of predictive coding theories. In Section 4, we briefly sketch the cellular processes that implement cooperative context-sensitive perceptual inferences. Section 5 explains how prediction error minimization can be achieved through such cellular processes. Section 6 indicates how dominant conceptions of predictive processing could be updated in light of the cellular discoveries of pyramidal neurons with two points of integration. Section 7 discusses implications of our perspective for a recent application of subtractive predictive coding to hippocampal-cortical interactions in episodic memory (Barron et al., 2020). Finally, Section 8 summarizes our conclusions and raises issues for future investigations.

## 2. Conceptual clarifications and working assumptions concerning predictive perceptual inference in mammalian neocortex

Despite the most natural reading of the word, 'prediction' does not necessarily concern the future. Although this default meaning of 'prediction' is often emphasized by philosophers of cognitive science (see, e. g., Clark, 2013a), the notion of prediction invoked in theories of perception is broader and includes predictive relations between concurrently observed datasets, e.g., to group data into coherent subsets and to disambiguate their interpretation. Many perceptual inferences do not aim to predict the future, but simply draw on stored inferential knowledge which is activated by incoming sensory data. Instead of 'predicting', it would thus be more appropriate to speak about our visual system *interpreting* the current sensory data.

In its contemporary forms, this inferential view of perception is infused with the Bayesian vision of evidential updating (Penny, 2012). But it is important to emphasize that, even in the Bayesian brain, predictions are not always anticipatory, as often implied by predictive processing theorists. Though perception frequently uses information from earlier events to guide interpretation of later events, and vice versa, ascending feedforward signals often convey consciously accessible information about events that could not have been either predicted or postdicted.

In this paper we use the notion of prediction in a broad sense, and we assume that many perceptual inferences are predictive in this sense. Taken thus broadly, predictions not only draw on feedback from higher to lower levels of perceptual processing hierarchies, but also, for example, relate concurrent events across modality specific hierarchies of sensory processing (Phillips, 2023, Chapter 5). They extend to action generating circuits because motor commands must be coordinated, so are to that extent mutually 'predictive'. Predictions in this broad sense can apply to events in the past. An exceptionally dramatic example of predicting the past occurred when observations of the crab nebula led to the prediction, subsequently confirmed, that a supernova had occurred in that part of the sky many centuries previously. More everyday examples include the disambiguation of earlier by later phonemes, words, or phrases within a sentence.

De-emphasizing the anticipatory sense of the notion of 'prediction' is

therefore one of the ways in which we distance our proposals about predictive inferences from those accounts of predictive processing that insist that generative models always issue anticipatory predictions before the corresponding sensory data arrive.[2] We also emphasize that higher levels of perceptual hierarchies do not necessarily predict the sensory *details* that are specified at lower levels of the hierarchy. Typically, they *generalize* over them: higher level percepts, or inferences, can be instantiated in more than one way, and usually in many different ways, at lower levels of the cortical hierarchies of abstraction.

A central aim of our paper is to provide grounds for supposing that perception does not depend upon the form of prediction error coding in which the output of individual pyramidal cells transmits information about the difference between the cell's ascending feedforward receptive field inputs and its contextual inputs, including those from feedback or other top-down sources. We will refer to that as 'subtractive coding'. In essence it is the form of coding used by communication engineers to greatly reduce the channel capacity required to transmit large amounts of slowly changing data. In contrast to that, we argue that in mammalian perceptual systems the predictive context is typically modulatory in that it either amplifies or attenuates transmission of information about its receptive field (RF) input. We, and many others, have referred to that as contextual modulation, but it could also be referred to as *modulatory Predictive Processing* (PP) to distinguish it from the subtractive coding that we dispute.[3] From that perspective, evidence for contextual modulation in cases where feedback provides the context can be seen as evidence for modulatory PP, including the many cases where it was not presented as being related to any form of predictive error coding (e.g. Lamme, 2004, 2020).

Given our focus on contextual modulation from the cellular to the cognitive systems and psychological levels, the key distinction between subtractive coding and modulatory PP is of obvious importance. Recent advances in the foundations of information theory (Williams and Beer, 2010; Wibral et al., 2015) have been used to define and quantify contextual modulation. Those advances in partial information decomposition quantify the information about which a local processor with two distinct sets of input transmits information uniquely about each of them (in addition to quantifying the transmitted information that depends on both subsets). They have recently been used to quantify the information transmitted specifically about contextual input to the apical dendrites of L5 cells in the cortex of mice (Kay et al., 2022). This directly confirms that under the conditions used apical input was modulatory rather than subtractive. In terms that are more explicitly related to research on predictive processing, the discoveries of Kay et al. (2022) directly and quantitatively demonstrate that in the conditions of that experiment modulatory PP occurred but subtractive coding did not.

It is sometimes taken for granted that evidence for the construction and use of internal models, in either perception or thought, implies subtractive coding. On that assumption our arguments against subtractive coding could be misinterpreted as implying that we doubt the use of internal models. We do not. We see the use of internal models as being crucial to perception and cognition in general. It is the assumption that evidence for the use of internal models is evidence for subtractive coding

---

[2] See, e.g., Clark (2019, p. 649): "Under that [Predictive Processing] schema, perception and action involve meeting the sensory flux with a stream of apt (generative-model based) top-down prediction." Note that this view is simply the logical consequence of the idea that "the forward flow of information is solely conveying error, and the backward flow is solely conveying predictions" (Clark, 2013a, p. 187f.). If *all* that gets passed 'upwards' through the hierarchies of neocortical perceptual abstraction are prediction errors, it follows that top-down predictions must be issued some time in advance before the sensory data arrive. If that were not the case, the predictions could not 'meet' the feedforward sweep in the early stages of the hierarchy.

[3] The notion of modulatory PP is meant to cover both the process of prediction error minimization and of prediction success maximization.

that we deny. Furthermore, if the free energy principle underlying that conception of PP does indeed define 'prediction error' such that it is both necessary to survival and unfalsifiable it could be neither confirmed nor refuted by specific evidence such as that reviewed here (see Fiorillo, 2008, 2010, 2012, and personal communication, 2023). Thus, as explained further below, we see our perspective as compatible with and complementary to basic insights expressed by the free energy principle.

Another potential misunderstanding of our arguments could also arise from assuming that we are concerned with brains in general. We are not. Our focus is specifically on cellular mechanisms for contextual modulation in *mammalian neocortex*. We make no claims about whether or how contextual modulation occurs in other species, though we do note that sensitivity to context is a key feature of mammalian cognition, and especially of human language and thought. We also note that cellular mechanisms for context-sensitivity that have been discovered in mammalian neocortex have not yet been seen in either subcortical regions of mammalian brains nor anywhere in other species. Perhaps such capabilities do not exist at the cellular level in those systems, and perhaps that is why the behavior of many other species is less flexibly sensitive to context than is that of mammals, and of humans in particular.

Information transmission as defined by Shannon implies that no information can be transmitted by neuronal signals about things that are known. Their informativeness increases as their probability decreases. Our main concern here is with the use of internal knowledge to choose between alternative interpretations of ascending feedforward data. In many cases the interpretation that is more likely given the current context seems the best bet and seems to be the one that is most often taken. Ascending data that contradicts confident expectations will be highly informative, however, and may thus be selected in those special circumstances. To show how our view offers a new perspective on the role of predictions in neocortical function, Section 7 of this paper will relate it to one that applies currently influential views of predictive processing to the role of hippocampal-neocortical interactions in episodic memory (Barron et al., 2020). Though highly influential and supported by much empirical evidence, the conceptual framework of which the work of Barron et al. (2020) is a good example remains inconclusive because it does not clearly and consistently distinguish between modulatory PP and subtractive coding. That is crucial because subtractive coding clashes with the evidence from cellular physiology that we emphasize throughout this paper. The aim of Section 7 will therefore be to show how these issues may be resolved by our emphasis upon context-sensitive neocortical pyramidal cells with two points of input integration.

The cellular mechanisms of modulatory PP, briefly sketched in Section 4, are able to build a system that minimizes errors in predictive inferences, a goal shared with predictive coding theories of brain function, but they do so without recourse to subtractive coding. A consequence of subtractive coding is that there would be no feedforward output from any level when top-down predictions are in full agreement with the forward flowing sensory data. In stark contrast to that, extensive evidence indicates that the feedforward outputs of many pyramidal cells in the neocortex are often *strengthened* when coincident feedforward and predictive contextual input agree in that both are depolarizing.[4] In this quickly emerging picture, functional segregation of prediction and error units postulated by most predictive coding theories

is replaced by a delicate interplay of amplification and attenuation that changes signal salience without changing what the cell's output transmits information about, i.e., its receptive field selectivity, other than by making it more precise in some cases.

To forestall further possible misunderstandings, we also note that some of our key working assumptions contradict other ideas that are widely taken for granted, often implicitly. These ideas, we argue, can be deeply misleading. Widely held assumptions in which readers must suspend belief if they are to understand what follows, are therefore explicitly listed here.

First, although the primary aim of this paper is to argue against subtractive coding, what we say still supports the hypothesis of prediction error minimization. We also argue for the objective of *prediction success maximization*. That objective is not equivalent to the objective of prediction error minimization, and is not open to the 'dark-room problem' (Sun and Firestone, 2020). There is evidence that neocortex uses predictions to amplify signals that contradict confident expectations, but that does not imply that what they code for is changed; indeed, that evidence is more plausibly interpreted as implying that what is coded for remains the same while becoming more salient when amplified. As we see it, the applied mathematicians and computer scientists who first proposed a theory of neocortex based on the notion of subtractive coding were misled by oversimplified views of the neocortex as having a purely hierarchical structure with feedback coming only from the next level up and without inputs from outside that hierarchy. They were also misled by assuming a major function of neocortex to be that of data reduction in the service of information transmission.

It is often taken for granted that information transmission is pivotal in the neocortex. We see that as seriously misleading. Transmitting information must be distinguished from the *interpretation and use* of information. Transmission is indeed the primary function of the optic tract, which transmits information about retinal input to the thalamus. Lossless data compression, such as that produced by Kalman filtering or Huffman coding, is indeed the primary function of the retinal microcircuitry that greatly reduces the channel capacity required of the optic tract. Information transmission is also the function of communication systems such as radios and TV sets, but, in contrast to neocortex, they do not interpret the information that they transmit, nor do they use it for any purpose. Neocortex both interprets and uses the information that is transmitted. Sensory information, and information from internal sources, is used to decide what to think and what to do but the purpose of those thoughts and actions is rarely, if ever, to simply transmit back to the world the information received from it. Overemphasis on information transmission is thus misleading when we are trying to understand the structure and function of neocortex.

When information theory is used in the sciences of brain and mind it is usually taken for granted that the more information transmitted the better. In stark contrast to that, we argue that selecting the currently relevant information for feedforward transmission is a crucial requirement of all neural systems, including mammalian neocortex.

Finally, the 20th century sciences of mind and brain took it for granted that neurons in general operate as integrate-and-fire point neurons. As argued in detail below, it is now known that many neocortical pyramidal cells have a second point of input integration near the top of the apical trunk. In contrast to proposals based on subtractive error coding, this second point of input integration is used as a context-processor which amplifies response to coincident basal and perisomatic activation.

In addition to rejecting these widely held assumptions, our perspective also advocates two other radical advances. Firstly, it is often implicitly taken for granted that, even if apical inputs to pyramidal cells can be distinguished on anatomical and functional grounds from that of basal/perisomatic inputs, they contribute to action potential generation in essentially the same way across time and hierarchical levels. In contrast to that, our perspective is open to the possibility that the exact form of the modulatory influences may vary across time and level in the

---

[4] When talking about 'coincidence' or 'agreement' between apical and basal/perisomatic input, we do not mean to imply that there is a match in representational 'content' at the level of single cell responses. Action potentials *per se* do not have content. They can be seen as simply saying, with various degrees of salience, 'There is evidence for the receptive field criteria to which I am selectively sensitive'. Truly representational properties only appear at the level of populations or networks of cells that jointly signal the presence of a perceptual feature.

hierarchy. Secondly, perception of the external world is sometimes thought of as being 'controlled hallucination' (e.g., in Seth, 2021). Controlled hallucination uses internal resources to interpret and systematize sensory data. If that process is restricted to higher perceptual levels, which perform what Ludwig Wittgenstein (1953) described as 'seeing as', then our perspective is also open to that possibility. Thus, a rich panorama of possibilities arises from the conception of neocortical pyramidal cells as having two functionally distinct points of input integration whose modes of operation may vary across time and level in the hierarchy.

## 3. Against subtractive prediction error coding

Theories of predictive coding ambitiously propose a general view of brain architecture and function: at each level of cortical signal processing, perceptual hypotheses are issued from level-specific generative models which represent how portions of sensory data are produced by hidden external causes of sensory input; these predictive hypotheses are sent as a feedback down the hierarchy to the preceding levels where they meet feedforward signals containing information about the stimulus. They typically propose that residual prediction errors are computed by subtracting descending predictions from feedforward sensory input. Signals coding for these prediction errors are then sent upwards to the next level in the perceptual hierarchy to correct activity at that level. In this way, prediction errors are continuously minimized, at every level of a predictive hierarchy (Mumford, 1992; Rao and Ballard, 1999; Huang and Rao, 2011; Bastos et al., 2012; Clark, 2013a; Hohwy, 2013).

The idea of prediction error coding was initially hypothesized by an eminent applied mathematician, Mumford (1992), and applied to the primary visual cortex by Lee and Mumford (2003) and by the computer scientists Rao and Ballard (1999). All of them were strongly influenced by anatomical and physiological evidence for hierarchical organization of neocortical regions. As we see it, they were misled by that evidence, however. Hierarchical aspects of neocortical structure and function are indeed important, as established most convincingly for primates by Markov and Kennedy (2013). Nevertheless, those hierarchical aspects are clearly embedded within a heterarchical system in which feedback from higher to lower hierarchical levels is inextricably intermingled with modulatory input from various other sources. Apical dendrites in layer 1 of neocortical pyramidal cells whose soma lie in deeper layers have a central role in mediating this diverse modulatory input. This contextual input is not restricted to feedback from higher levels (Schuman et al., 2021), let alone to feedback from the next level in a cortical hierarchy (cf. Litwin and Miłkowski, 2020), but includes contributions from subcortical centers such as the thalamus and amygdala. There is also clear evidence that this contextual modulation amplifies currently relevant neural signals and attenuates irrelevant signals.

We will look at some of this evidence more closely in the following sections. In this section we aim to spell out the main shortcomings of the subtractive coding hypothesis which are relevant for our argument. The simple idea underlying the notion of subtractive coding is that information transmission is more efficient if lower levels of abstraction do not tell higher levels what those higher levels 'know' already. So predictive coding accounts assume that feedback tells the lower levels what is already known at the higher levels, and that is subtracted from the feedforward input to the lower levels. As a result, all that needs to be transmitted forward to the higher levels is the 'error', i.e., the difference between feedforward input and feedback predictions. Prediction errors are then propagated through the perceptual hierarchy, until high-level perceptual hypotheses are accordingly adjusted.

A distinctive feature of theories based on subtractive coding is that subtractive coding does not just sometimes happen in the neocortex; it always happens. According to predictive coding views, all feedforward processing consists *solely* in the transmission of prediction errors defined as feedforward input minus feedback from a higher level in the hierarchy (Clark, 2013a, p. 187f.; Clark, 2019, p. 647; Sprevak, 2021).

However, the abstract idea of prediction error calculation by suppression of correctly anticipated information needs to be operationalized in neurobiological terms. The most natural way to operationalize the cancellation of successful predictions in neurobiologically realistic terms is to invoke neural inhibition, in which feedback from higher levels 'dampens' somatic action potentials at previous levels. In line with this, most predictive coding theories assume that feedback is in effect inhibitory. However, as noted by Spratling (2019) and by Aru et al. (2020a), that requirement puts those models of predictive coding at odds with single-cell neurophysiological data. That data shows that the effects of cortical feedback are often excitatory or disinhibitory as outlined further below (see also Johnson and Burkhalter, 1997). This is especially clear in the case of feedback and other types of input to layer 1 cortical terminals. Much of that input is conveyed by either direct excitatory synapses or by disinhibition.

One may attempt to reconcile such facts with subtractive coding by saying that excitatory feedback can be made negative through inhibitory interneurons (see Bastos et al., 2012; Kanai et al., 2015). The idea is that as feedback reaches the neurons carrying the feedforward information, activated inhibitory interneurons stop the production of action potentials in these neurons according to the computed (i.e., subtracted) prediction errors. But this attempt, too, is unconvincing. As noted, much of the input to layer 1 is either directly excitatory or disinhibitory. Although some of the input to layer 1 is indeed to inhibitory interneurons, some of these interneurons such as the vasoactive intestinal peptide-expressing (VIP) interneurons, activated by long-range inter-regional connections, *disinhibit* apical dendrites by inhibiting the local tonic inhibition that is produced by SOM/SST interneurons (predominantly activated by local intra-regional connections). VIP interneurons disinhibit apical dendrites by making small holes in the inhibitory blankets cast over them by SOM/SST cells (Karnani et al., 2014). So instead of suppressing action potentials, as predictive coding requires, activation of VIP interneurons *increases* the amplifying or driving effects of apical depolarization.

Our skepticism about the validity of neural architectures proposed by predictive coding theories is complemented by failures of studies designed to directly test predictions based on subtractive coding in the lab. Although some publications report positive evidence for subtractive coding (for overview, see Walsh et al., 2020; Shipp, 2024) other studies reached more pessimistic conclusions. For example, Alilović et al. (2019) report that the first transient wave of feedforward processing (<80 ms post-stimulus) is impenetrable by predictive effects. Solomon et al. (2021) purported to isolate the markers of subtractive predictive coding in the cortex, and found almost none (see also the negative results of Ouden et al. 2023). Given the lack of convincing empirical support, the popularity of predictive coding models may be surprising. Our simple argument here is that the hypothesis of subtractive coding may have gained much of its current popularity by being conflated with modulatory PP and by the necessity of prediction error minimization, neither of which necessarily imply subtractive coding.

## 4. A brief sketch of cooperative context-sensitive intracellular processes

The hypothesis of predictive coding was developed within a systems neuroscience based on the assumption that all neurons function as 'integrate-and-fire' processors in which all inputs are integrated to compute a single data point about which information is transmitted downstream for further processing. If cells are point processors, then they must transmit information about everything that affects them. For such processors, any differences in input will necessarily change what they transmit information about, including a possible change from representing a stimulus feature to conveying a prediction error.

In stark contrast to the integrate-and-fire point neuron assumption there is now convincing evidence for pyramidal cells that can act as processors with two distinct points of integration, somatic and apical.

The apical integration zone was first discovered by the technically demanding investigations of pioneers such as Larkum et al. (1999), Xu et al. (2012), Larkum (2013), Beaulieu-Laroche et al. (2018), Williams and Fletcher (2019), Harnett et al. (2015) and others. The potential functional utility of the apical integration zone was demonstrated by computational models such as those of Siegel et al. (2000) and Körding and König (2000). The functional specialization of the apical integration zone has been directly observed by multisite patch-clamping studies in rodents (Schulz et al., 2021). Further analysis of those observations confirms that they reveal modulatory effects as defined by Kay and Phillips (2020) using state-of-the-art three-way mutual information decomposition (Kay et al., 2022).[5]

Here we concentrate on the mode of apical function referred to as 'apical amplification' (AA), which is central to the perspective advocated by Larkum and Phillips (2016, 2017), Phillips et al. (2018), Marvan et al. (2021), and by the Dendritic Integration Theory (Aru et al., 2020a; Bachmann et al., 2020). In a nutshell, AA is that mode of operation in which pyramidal cells amplify transmission of information about their basal and perisomatic input when it is useful to do so in the context of input from diverse cortical and sub-cortical sources as signaled by input to their apical dendrites in layer 1. Thus, we and others interpret the discovery of AA as implying that the strength of the cell's output is increased when the basal and the apical inputs are both depolarizing (Phillips et al., 2015; Larkum and Phillips, 2016; Phillips, 2017; Aru et al., 2020a). The direct long-range inter-regional excitatory input to layer 1 (containing apical dendrites of pyramidal cells in layers 5, 3 and 2), whatever its origin, typically amplifies the cell's response to coincident feedforward basal and perisomatic excitation. The signature of a successful match between apical and basal inputs is the generation of a brief burst containing 2 to 4 spikes within about 20 msec separated by silence for approximately 100 msec and usually synchronized across a sparse subset of cells (e.g., as described by Naud et al., 2023).

Fig. 1 sketches some of the things that are now known about intracellular processes within context-sensitive neocortical pyramidal cells and about their (dis)inhibitory regulation. For reviews of evidence on which this sketch is based and of its implications for cognition see Larkum (2013, 2022), Phillips (2017, 2023), and Tantirigama et al. (2020). For a review of evidence showing that dysfunctions of the mechanisms sketched in Fig. 1 are central to the way in which cognition is impaired by the various genetic and environmental origins of Fragile X, Down Syndrome, and Fetal Alcohol Spectrum Disorders see Granato et al. (submitted). Though Fig. 1 is typical of cells whose soma lie in layer 5, and is thus applicable to all neocortical regions, it may, in some cases, also apply to cells in more superficial layers, as discussed in Section 8.

Apical amplification provides a mechanism by which information about basal and perisomatic inputs can be amplified when appropriate without being contaminated or corrupted by inputs coming from the modulatory context. Thus, apical input can function in a way that keeps its effects separate from those of feedforward inputs. Two-point processors can use their contextual input, including feedback, to *modulate* the strength, or salience, with which feedforward information is transmitted. Importantly, two-point processors do not trade in prediction errors. What is transmitted through perceptual hierarchies is the information about current sensory input to basal and perisomatic synapses. Apical depolarization amplifies the transmission of relevant signals without changing what those signals code for, i.e., what they transmit information about. Thus, in contrast to the hypothesized subtractive effects of predictive coding, the evidence for cellular cooperative context-sensitivity implies that the cell's output is strengthened when dendrites receiving predictive contextual input are activated at about the same time as dendrites receiving the feedforward input. AA thus

supports the common view that feedback signals are functionally distinct from feedforward signals, but the picture of brain function it offers is not that proposed by subtractive coding.

It is also possible that, in addition to the mode of apical amplification, there is also a mode of operation in which fictive percepts are generated by apical input instead of by RF input. One form of 'apical drive' could provide a cellular mechanism for dreaming (Hobson et al., 2014; Aru et al., 2020b). Other forms may provide mechanisms for imagination and thought when awake. If so, then refinements of those modes of apical operation might be involved in the creative and linguistic capabilities that are distinctively human (Phillips, 2023, Chapter 6).

When operating in the amplifying mode the apical site of these pyramidal cells regulates the salience of the somatic output without directly generating its own output. Modulatory effects of apical input are not limited to being either fully amplifying or fully driving, however. The outputs of context-sensitive pyramidal cells can be amplified to various extents depending upon their current relevance and informativeness. Signals known to be highly probable in a specified context carry little information in that context; so processing resources would be saved if amplification of those signals by apical dendrites were attenuated in that context. Indeed, there is direct evidence that such signals are attenuated by increased activation of the SOM interneurons, as first reported by Larkum et al. (1999). The signals transmitted in those conditions will therefore be attenuated relative to conditions in which amplification is not reduced. Signals that contradict context-dependent predictions will be highly informative, and there is direct evidence that in such cases the signals are amplified by disinhibiting the apical dendrites (Harris and Shepherd, 2015; Wang and Yang, 2018). Modes of operation intermediate between amplification and drive are also possible, so in that case the distinction between reality and imagination would become problematic.

Again, note that making a message less salient because it is not highly informative, or amplifying it because it is, implies that the message being conveyed *remains the same*. Thus, 'context-sensitivity' refers to modulatory effects that do not fundamentally change the information transmitted by the cell other than by making it more precise. 'Cooperativity' implies that if apical and basal/perisomatic inputs are both excitatory, i.e., depolarizing, then the cellular output generated is amplified, which implies a supra-linear interaction between the two sites. This cooperation implies an asymmetric interaction between the two input sites because the apical contribution depends on the presence of basal/perisomatic activation, but not vice versa (Kay et al., 2022). It may also be that if net apical input is inhibitory, cellular output generated by excitatory input to basal and perisomatic synapses is attenuated, but the physiological evidence on that is less clear.

Although not explicitly related to the distinct roles of basal and apical dendrites, the theory of Coherent Infomax was based on local processors with two functionally distinct points of input integration (Kay et al., 1998; Kay and Phillips, 2011). That theory defines RFs and contextual fields (CFs) mathematically and shows that together they provide computational capabilities of great generality, including figure ground organization and contextual disambiguation. It must not be confused with the theory of Infomax (Linsker, 1992), which simply maximizes the transmission of all information, whether currently relevant or not. In contrast to that, Coherent Infomax shows how it is possible to amplify only the transmission of information that is coherently related to that being transmitted by other cells, thus maximizing agreements between them and minimizing mismatches between them.

## 5. Predictive inferences do not imply subtractive prediction error coding

This section argues in more detail that increase of prediction success and reduction of prediction error do not imply subtractive coding. The cellular mechanisms of cooperative context-sensitivity sketched in the

---

[5] For overviews of the functional capabilities of pyramidal neurons with two sites of integration see Phillips (2017, 2023); Phillips et al. (2018); Aru et al. (2020a); Marvan et al. (2021); Aru et al. (2023).
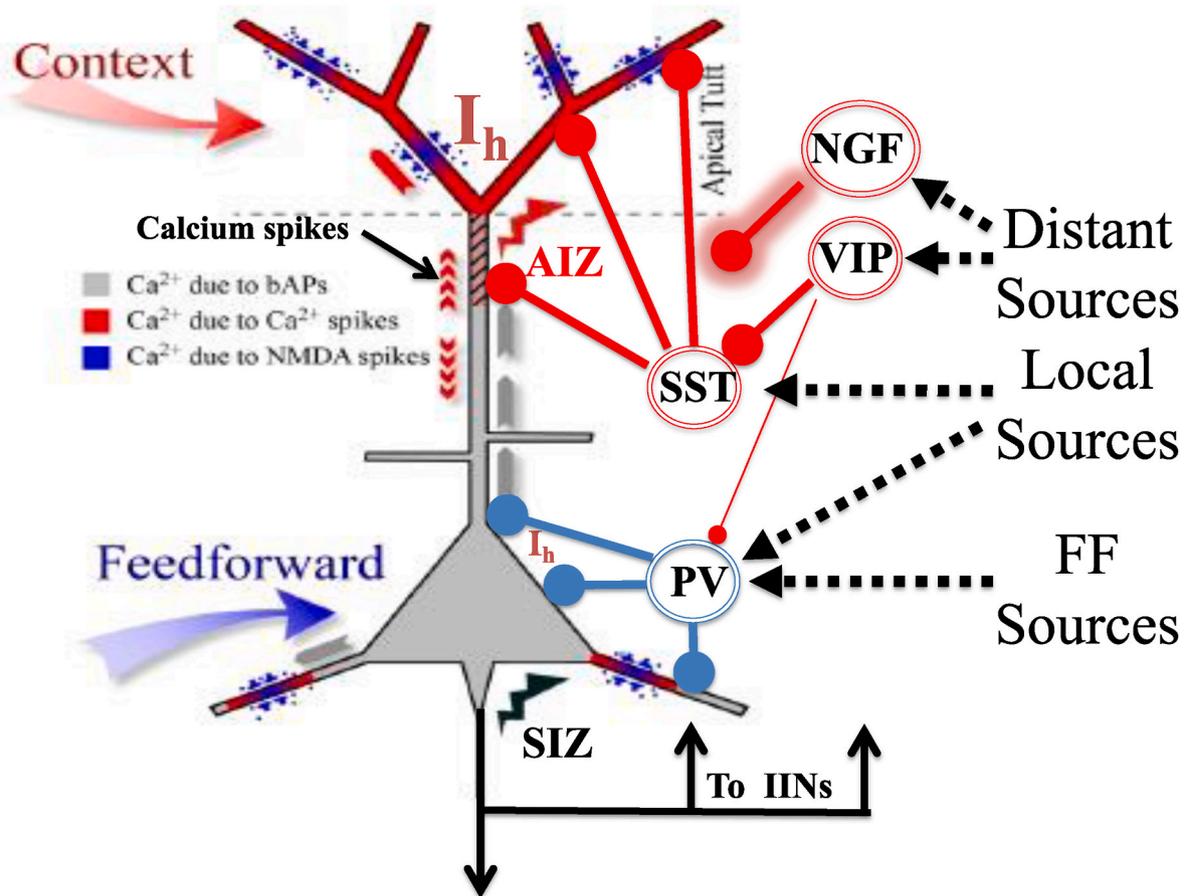
**Fig. 1.** A sketch of a context-sensitive pyramidal cell and its (dis)inhibitory regulation. The idealized diagram of a pyramidal cell shown on the left is adapted from Larkum and Phillips (2016). In the modulatory mode of operation, contextual input, which comes from diverse sources, amplifies transmission of information from the cell's RF, i.e., the specific set of feedforward sources to which the cell is selectively sensitive. Ih: hyperpolarization activated current flow through HCN ion channels. AIZ: apical integration zone. FF: Feedforward. SIZ: somatic integration zone. IINs: four classes of inhibitory interneuron, i.e., PV, SST, VIP, and NGF. The output of NGF cells is shown in a highlighted form and without a well-specified target because they inhibit nearby pyramidal cells and IINs via volume release of GABA mainly in the upper layers, but also to a lesser extent in deeper layers.

last section can achieve these objectives by amplifying coherent subsets of signals in the short-term, and by learning what predicts what by adapting the strengths of the apical and basal synapses in the long-term (Kay and Phillips, 2011).[6]

Shannon's information theory quantifies information as the reduction of uncertainty. Known signals carry no new information, and the salience with which features are signaled tends to increase as their improbability increases.[7] When feedforward input carries no new information, levels higher in the hierarchy can be seen as in effect saying to lower levels: 'Message received and understood, so no need to keep making that information salient'. This explains why onsets and offsets of new information usually produce far stronger internal responses than do sustained patterns of stimulation following their onset.

There is much evidence that unpredicted sensory information evokes greater neuronal response in relevant parts of the cortex, especially in omission responses, expectation suppression, and repetition suppression

(Walsh et al. 2020). *Prima facie* that may sound like subtractive coding, according to which unpredicted stimuli trigger stronger response in hypothetical 'error units'. However, there is an alternative explanation of these response profiles. The features about which the cells in V1 and other sensory areas transmit information, i.e., their receptive field selective sensitivity, *remain the same*, whether they were predicted or not. That is also the case in the mathematical model being developed by Mikulasch et al. (2023), so relations between our perspective and theirs merits close examination. In both their perspective and ours, cells in sensory areas tend to amplify their outputs when they are currently deemed to be relevant, including amplification that occurs because they contradict confident predictions, and are thus highly informative. Amplification of unexpected signals is by no means an unbroken principle of cortical function, however, because there is plenty of evidence for what is known as 'confirmation bias' in which interpretations of new data that confirm current beliefs are prioritized over interpretations that do not. How that evidence can be reconciled with evidence for the amplification of signals that contradict confident predictions remains to be clarified.

Hence, findings that have been interpreted as supporting subtractive coding can now be seen as providing further information on the context-sensitivity of pyramidal neurons. Instead of emphasizing prediction errors, AA distinguishes between contextual amplification of important sensory signals and contextual attenuation of irrelevant or incoherent signals by dedicated perceptual mechanisms. Note that just as amplification could be misinterpreted as the transmission of prediction errors,

---

[6] Though we have here considered only a generic conception of context-sensitive pyramidal cells, it is becoming clear that there are anatomical, physiological and functional differences between those in different layers, regions and species. Developing conceptions of such cells in ways that help us understand those differences is a major task for the future.

[7] Although we focus here on amplification, we also assume that correctly predicted sensory information can be made more precise; see, e.g., Teufel et al. (2018).

so attenuation could be misinterpreted as error subtraction in cases of successfully predicted sensory data. Yet, in contrast to subtractive coding, it is possible to interpret these cases of reductions of signal strength as being due to a form of attenuation that does not change the information transmitted by the signal that is attenuated.

Note that amplification and attenuation can occur concurrently in different cells. To take one example, the pulvinar conveys diverse contextual information to V1 that amplifies transmission of information about changes in the visual scene that are not predicted by the animal's own actions (Roth et al., 2016). This increases the response of V1 to non-self-generated events within a background of responses to self-generated events that are not amplified, and are usually to some extent suppressed. Similar principles lie behind a common neocortical process of 'normalization': amplification of selected signals automatically attenuates the transmission of their competitors. In fact, given that sensory signals in the neocortex typically have several competitors, if follows that amplification of selected signals in the neocortex entails attenuation of more signals than are amplified.[8]

As already noted, amplification of selected signals does not require that their message be anticipated by higher-level cell assemblies. For example, processes of spatial attention can increase the salience of selected signals, but that does not imply that attention anticipated the signals; amplification of the signals can be conditional upon their occurrence. By the same token, there is no need to assume that visual or auditory search is predictive in the anticipatory sense. Search operates as though it were simply saying: "If there is evidence of the thing searched for, then send it downstream loud and clear!" So, spatial attention could operate by amplifying the transmission of whatever (feedforward) information is present at the target location.

The mathematically explicit model of hierarchical inference in neocortex being developed by Mikulasch et al. (2023) shows that it can be efficiently implemented using spiking neurons with feedforward input to basal and top-down input to apical dendrites. Though they refer to 'coding' throughout, Mikulasch et al. (2023) do not imply that the signals transmitted are a code for errors. Indeed, they state clearly that, as in Spratling's models, it is inferences that are transmitted, not errors. The model of Mikulasch et al. (2023) has so much in common with the perspective presented here that the two approaches are strongly mutually supportive. The various differences between them may be worth exploring when their shared assumption of a distinct and crucial functional role for apical input receives the attention that it deserves.

## 6. Upgrading popular conceptions of predictive processing by reference to mechanisms of cellular cooperative context-sensitivity

The hypothesis of subtractive coding within conceptions of predictive processing has been used to explain many psychological and electrophysiological findings. Nevertheless, despite its widespread popularity, subtractive coding faces several difficulties which, we argue, can be overcome by combining the concepts of inference and prediction with the evidence for cooperative context-sensitive pyramidal cells. The empirical discoveries and conceptual advances outlined here and elsewhere support aspects of the predictive processing perspective, but they require it to be reinterpreted as being more concerned with enhancing the salience of coherent or relevant signals, rather than with changing the information for which the signals code, as the hypothesis of subtractive coding requires.

The extensive evidence for diverse sources of modulatory contextual input to the apical dendrites in layer 1 is a problem for the predictive coding hypothesis because feedback from hierarchical level n+1 to level n is inextricably intermingled with inputs from a diverse range of other sources. That becomes a strength from the viewpoint of cooperative context-sensitivity because it puts local processing in a broader context, which is not limited to activity at level n+1 in the same hierarchy. One direct upgrade of PP and the free energy reduction (FER) theories would then be that context-sensitive cellular and synaptic mechanisms could provide mechanisms for the *precision weighting* postulated by those theories (see, e.g., Clark, 2013b; Rigoli et al., 2019; cf. Shipp, 2016, sect. 11). Precision weighting is more concerned with the salience of sensory signals than with changes in coding. As has been noted several years ago, "the predicted precision has an excitatory modulatory effect, consistent with mediation through voltage-dependent NMDA receptors in pyramidal cells of the superficial layers" (Kanai et al., 2015). This conception of precision weighting is thus consistent with the predominantly excitatory consequences of input to layer 1 as noted above.

Conceptually, linking precision weighting to contextual amplification makes good sense because inputs that effect precision within PP and FER frameworks can be seen as being part of the modulatory contextual input. However, the discoveries showing that apical activation amplifies the cell's responses to feedforward basal activation contradicts the hypothesis of subtractive coding, as Aru et al. (2020a), amongst others, acknowledge. Our emphasis on maximizing agreement, rather than minimizing disagreement, between cellular activities is directly inherited from the theoretical perspective of Coherent Infomax, which focuses on interactions between local processors that, in addition to their RF inputs, receive CF inputs from diverse sources (including but not limited to feedback), with the RFs and CFs being integrated separately.

The perspective of Coherent Infomax is far more closely rooted in cellular neurophysiology than are current theories of predictive processing and free energy reduction. It largely ignores the hidden external states inferred by generative models, and focuses on the transactions between and within local neural processors. Nevertheless, despite these differences of emphasis, and despite the two perspectives having been developed independently, the agreements between them are more important than their disagreements. In essence they may therefore be more complementary than being opposed. Most importantly, both perspectives assume that a fundamental objective of neocortical dynamics is to reduce differences between predictions and observations, where 'predictions' are to be interpreted as specified by the working assumptions stated above (in Section 2). But the agreement between the two perspectives probably runs even deeper because unpublished work in the 1990s by John Hertz, a statistical physicist with advanced knowledge of neural computation (see Hertz et al., 1991), derived a proof, now sadly lost, that the theory of Coherent Infomax implies the reduction of free energy. Furthermore, Karl Friston has more recently shown that free energy reduction can be seen as a special case of the general objective on which Coherent Infomax is based. Unfortunately, that proof also remains unpublished.

What we in effect propose here is that modulatory contextual mechanisms can in principle do all the work entrusted to subtractive prediction error computations in previous conceptions of predictive processing and of the free energy principle. Relinquishing the emphasis upon subtractive coding would, of course, be a major concession on the part of many predictive processing theorists. However, the change of perspective that we advocate implies that one can preserve much of the core of those perspectives if they are presented in a way that does not depend on subtractive coding. This is clearly seen in those versions of 'active inference' theory that do not depend upon subtractive error

---

[8] For rigorous and extensive accounts of how perception, attention and working memory depend on competitive normalization combined with cooperative recurrent amplification, see Heeger and Zemlianova (2020) and Heeger and Mackey (2019).

coding (e.g., Linson et al., 2020).[9] We see the current advances in cellular physiology and multivariate information decomposition as offering adherents of subtractive error coding the opportunity to free themselves from the shackles of the point neuron and subtractive assumptions. Given the widespread influence of those two misleading assumptions, however, we do not expect that opportunity to be immediately popular.

## 7. Our view of predictive processing as applied to hippocampal-cortical interactions in episodic memory

Assuming that predictive processing depends on subtractive coding, Barron et al. (2020) seek to reconcile that assumption with ample evidence that input to the neocortex from the hippocampus generates neocortical activity during episodic memory recall. In essence, they conclude that during perception descending projections, including those from the hippocampus, directly inhibit feedforward output from neocortical pyramidal cells, whereas during recall hippocampal input facilitates or generates output. They propose that which of those two different processing modes operates at any moment depends on regulation by the classical neuromodulators, with NMDA receptors for glutamate also having a major role.

We strongly agree with many aspects of the case that Barron et al. present. First, we agree that precision has a central role. We see it as being much the same as what we call salience except that it could also include the width of the cell's RF tuning. Second, we agree that precision is increased by disinhibition. Third, we agree that attention involves disinhibition. Fourth, we agree that it is important to distinguish apical from basal inputs, and that the hippocampal input is predominantly to the apical tufts in layer 1. Fifth, we agree that there are different modes of neuronal operation, which are dynamically regulated by the classical neuromodulators and in ways that involve NMDA glutamate receptors. Finally, we agree that during episodic recall fictive percepts are generated in the higher perceptual regions that were activated during the perceptual experiences being recalled.

Some aspects of the hypothesis presented by Barron et al. (2020), however, are contradicted by direct evidence from anatomy and cellular physiology. Most, if not all, of them arise from their commitment to subtractive coding. Given that commitment, they bravely state that "descending projections that convey predictions must therefore either be inhibitory (e.g., long-range GABAergic projections) or target local inhibitory interneurons (e.g., in superficial cortical layers)" (Barron et al., 2020, p. 5). Despite the data that they interpret as supporting that inference, there is ample evidence that descending input is often either excitatory via direct glutamatergic synapses to pyramidal cells and/or disinhibitory via VIP mediated inhibition of somatostatin positive interneurons (e.g., Karnani et al., 2014; Wang and Yang, 2018; Khan and Hofer, 2018). Anatomical and physiological findings on inhibitory microcircuitry and its role in the plasticity of visual cortex is convincingly reviewed in quantitative detail by van Versendaal and Levelt (2016). In contrast to the purely inhibitory effects of feedback hypothesized by Barron et al. (2020), Figure 2 in van Versendaal and Levelt (2016, p. 3681) shows quantitatively the consensus view that VIP positive interneurons receive long-range input including feedback from higher levels and predominantly inhibit somatostatin positive interneurons that receive their input from local rather than from distant sources. Thus, there is much evidence that descending and other long-range inputs often amplify pyramidal cell outputs either directly or

by attenuating the local suppression that normalizes pyramidal cell outputs (Heeger and Zemlianova, 2020), which is consistent with the findings of Zhang et al. (2014).[10]

The extensive overlap between our view of predictive processing and more conventional perspectives is clear in a recent multi-scale multi-species overview of the microcircuitry underlying prediction and context-sensitivity (Muckli et al., 2023). That overview by 26 leading experts focuses on high-resolution fMRI recordings of neocortical activity in humans under conditions in which context-dependent activity is separated from feedforward activity and relates the data in detail to studies using monkeys and rodents in analogous conditions. Muckli et al. (2023) conclude that context-sensitivity observed using multiple methodologies in monkeys and rodents is conserved in humans, and that it has many properties compatible with conventional views of predictive processing. The overlap between our view of modulatory PP and those views is also clear in that our view is in essence much the same as that of Mikulasch et al. (2023) who present modulatory PP as being functionally equivalent to classical predictive coding, even though feedforward signals in their model transmit inferences, not errors. Such revised view of error computation and signaling is now explicitly being seen as compatible with the general predictive processing framework (Brouillet and Friston, 2023).

As Barron et al. (2020) claim that predictive coding is known as Kalman filtering in engineering, we note that we too are committed to the use of insights offered by principles or algorithms that have proven useful for data compression in real-world applications. Engineering techniques for data compression use various mathematical functions for transmitting information through the local processing nodes, however. Typically, those functions use subtraction combined with other simple arithmetic operators. The combination of subtraction with other arithmetic operators is crucial because the key properties of transfer functions specified by combinations of simple arithmetic operators are fundamentally different from those of the simple operators alone (Kay and Phillips, 2020). In contrast to what some readers may assume, though, modulatory PP cannot be interpreted as multiplicative predictive processing. Multiplication is intrinsically symmetric whereas modulation is intrinsically asymmetric (Kay and Phillips, 2020). Context-sensitive pyramidal cells transmit much unique information about basal input but little or none about apical input – even in cases where apical input has a large effect on the cell's output via its contribution to synergy, a component of output that depends on both apical and basal input (Kay et al., 2022).

Barron et al. (2020) focus largely on macroscopic imaging evidence and say explicitly that they are not concerned with how errors are coded neurally. Therefore, we note that macroscopic neuroimaging typically confounds axonal with dendritic currents, so it provides ambiguous evidence on events at the cellular level. It is of course possible that both subtractive coding and modulatory PP occur in neocortex. It is impossible to prove that subtractive coding does not. We do not contest the evidence that Barron et al. (2020) cite in support of their hypothesis. Interpretations of it as evidence for subtractive coding will not be convincing, however, unless they explicitly refute the more

---

[9] Leading proponents of the free energy reduction principle, such as Karl Friston, have from the outset noted that error coding is *a* way to implement that principle, not the only way. Furthermore, the work of Mike Spratling (such as Spratling, 2008, 2017) is often cited in support of subtractive coding even though its key contribution is to show that feedforward transmission of errors is not crucial to predictive coding.

[10] Christiaan Levelt (personal communication) notes that although neurogliaform (NGF) cells also receive descending input and inhibit cells in their vicinity, they do not prevent cells from responding to their ascending input as implied by predictive coding. He also tells us that inhibitory interneurons that do so have not yet been identified, though he is still looking for them. Even if cells that can do so with the required local specificity are discovered, however, that would not resolve the issue with which we are here concerned. Showing that feedback can attenuate, or even fully, prevent, feedforward transmission would not show that when there is transmission it codes for the feedforward signal minus the feedback signal. Nor would such a discovery weaken the direct evidence that descending and other contextual input can also be amplifying, and that under some conditions can even generate output by itself.

straightforward interpretation of it as evidence for modulatory interactions.

## 8. Concluding remarks

The hypothesis of predictive coding posits two sets of pyramidal cells that are distinguished by what they represent, i.e., what they transmit information about: those representing predictions, and those representing subtractive prediction errors. Our claim in this paper is that this functional bifurcation is not necessary to implement the general goal of maximizing the success of predictive perceptual inferences, and of minimizing the overall prediction error. We do not assume, though, that all inputs to the apical dendrites in layer 1 are usefully interpreted as being predictive, even when prediction is conceived broadly. In many cases it is more appropriate to think of input to apical dendrites in layer 1 not as a prediction but as saying: "*If* your RF is being activated then transmit that information loud and clear". For example, that is a more natural way of thinking of selective attention and emotional prioritization. Sensory inputs to a selected location in space can be attended to without predicting what, if anything, occupies that location. Signals with highly emotional connotations can be amplified without being predicted (although some disorders, such as PTSD for example, may arise because that prioritization becomes a self-fulfilling prophecy).

Conceptions of predictive error coding are so many and so various that we cannot discuss all of them. Our perspective shares so much with the version proposed by Shipp (2024), however, that it is worth seeking to reconcile the differences between his views and ours in a way that preserves the best of both. Our claims agree in the following ways (Shipp, personal communication). We both claim that receptive field selectivity is not changed by feedback. We both claim that FB can increase the strength, or salience of feedforward transmission to higher levels. We both claim that input to L1 cannot operate as a route for 'subtractive' prediction error coding. As the way in which those agreements are expressed implies, reconciliation of the differences between Shipp's views and ours will require some approximate terminological translations. From our perspective 'error cells' could be much the same as 'cells whose outputs are strengthened or weakened by context'. 'Feedback' could be much the same as 'context' with the proviso that context includes but extends beyond feedback and 'top-down'. 'Precision' could be much the same as salience, strength, or exactness of feedforward receptive field selectivity. Shipp's model proposes that backward precision signals are applied to apical dendrites (plus interneurons) in layer 1, specifically to those of error neurons; whereas the signal that is applied to basal dendrites of deep-layer cells is the information-predictive component of backward messaging that is then routed toward superficial error cells for the purpose of subtractive coding (Shipp, 2016, 2024; Shipp and Friston, 2023). The subtlety and complexity of that model is now so great, however, that understanding, testing, and developing it is far from easy. As we see it, however, the role of basal/perisomatic dendrites in predictive processing could be clarified by using partial information decomposition to quantify the distinction between receptive field selectivity and modulation as done by Kay and Phillips (2020). On the basis of this quantification it has already been demonstrated, both physiologically (Kay et al., 2022) and by high resolution modeling (Graham et al., 2024), that apical inputs can be modulatory. It might be possible to do something analogous to that for basal/perisomatic synapses by dividing them into two subsets, such as the proximal and distal parts of each dendrite (see Jadi et al., 2014), for example. If that were done, and if it turned out that a subset of basal/perisomatic synapses are modulatory, as so quantified, then that would show that apical dendrites in L1 are not the only route by which context can modulate feedforward transmission in a way that does not corrupt feedforward selectivity.

We fully acknowledge that subtractive prediction error coding can, in principle, be implemented in models of neocortical microcircuitry, such as that of Bastos et al. (2012). Indeed, it may be that some modified

version of them does apply to circuits outside of the neocortex, or even to the neocortex itself in early stages of development. As argued in detail above, however, we doubt its applicability to mature neocortex. The use of codes that change from moment to moment and are interpretable only if a prediction is known would raise serious difficulties in systems as heterarchical and dynamically adaptable as neocortex. Given the strength of the evidence for a functionally distinct apical site of integration, that evidence should either be disputed or explicitly incorporated into conceptions of neocortical microcircuitry. Capone et al. (2023) show in detail how it can be incorporated in ways that provide cellular mechanisms for target-based learning that propagates targets or inferences rather than errors. They show explicitly how that enables effective performance in real time of context-dependent spatiotemporal tasks, imitation learning, and the decomposition of long-horizon decisions into simpler sub-tasks. As shown by the direct physiological studies of Kay et al. (2022), a key property of the outputs of context-sensitive cells, including those whose capabilities are demonstrated by Capone et al. (2023), is that it consists of brief bursts of 2–4 spikes within about 20 ms separated by silent intervals of about 100 ms. That is compatible with evidence interpreted as implying that layer 2/3 cells have more power in high gamma band whereas cells in the deeper layers have more power in the alpha-beta band, because those relations depend on the number of spikes in the burst and the duration of the intervals between them. Further study of how the traditional decomposition of output power across spike frequencies relates to analyses of bursting in microcircuit models is therefore a major task for the future.

As our arguments focus on context-sensitive neurons with two functionally distinct points of input integration, their prevalence is a crucial issue. Empirical study of that by direct multisite patch-clamping is no easy matter because relatively few labs have the technical skills that it requires. Nevertheless, wherever it has been studied in layer 5 cells it has usually, though not always be found to be present. A quantitative study of that issue in mouse V1 found that pyramidal cells had a distinct apical point of integration if their apical trunks were more than 0.460 mm long (Williams and Fletcher, 2019). A useful rule of thumb for estimating their likely prevalence is thus that the two-point mode of input integration is likely to apply if apical trunk-length exceeds about half a millimeter, which is the case for the majority of layer 5 cells, and also for many superficial cells in human neocortex (Kalmbach et al., 2018). Our assumption of a high prevalence for context-sensitive cells is greatly strengthened by high resolution multicompartmental modeling (Shai et al., 2015). That kind of modeling has now been used to show explicitly that the distinct capabilities that we assume for context-sensitive cells is a direct consequence of their biophysics (Graham et al., 2024). Though interaction between the apical and somatic input integration sites has till now been described as active 'coupling' the calcium spikes by which it is achieved are excitatory. Modeling and physiological studies of that coupling typically consider excitatory apical inputs only because there are no known mechanisms whereby inhibitory apical inputs are actively transmitted to the soma. That is unlikely to occurs because of the distance involved and because of ion flow through HCN ion channels which opposes such transmission. Nevertheless, explicit physiological and modeling studies of the conditions under which inhibitory apical inputs are conveyed to the soma are now needed to further clarify this issue.

Another major issue that now requires in-depth investigation concerns the role of amplification in the processing of rewards that can be either positive or negative. One possibility is that basal and apical synapses are strengthened when feedforward transmission leads to positive reward and weakened when they lead to negative reward. Feedforward transmission might then become less context-sensitive if they do not depend on context, and more if they do. It is also possible that progress can be made on this issue by reconsidering the definition of 'prediction error' to align it more closely with information theory and with broader notions of probabilistic inference (Christopher Fiorillo, personal communication). Whether these speculations are useful is not yet clear,

however, so, as we see it, this is a major issue for the future.

The discovery of neurons with two functionally distinct sets of input is uncontroversial and at last being given the widespread attention that it deserves. This is witnessed by recent empirical findings supporting the two-point architecture (e.g., Takahashi et al., 2016; Dowdle et al., 2023; Pujol et al., 2023) and also by practical applications of this new view of cortical function. Perhaps most importantly, the discovery of two-point neurons is beginning to provide the basis for a far clearer understanding of various pathologies (Phillips, 2023, Chapter 7). Unbroken paths can now be mapped all the way from genetic mutations and environmental insults that influence apical function to their various cognitive consequences in various neurodevelopmental disabilities (Granato et al., submitted).

The two-point neuron view is beginning to inspire major advances in computer vision (Schmid et al., 2023) and in neuromorphic machine learning algorithms that are far more effective and energy-efficient than standard deep learning algorithms (see, e.g., Adeel et al., 2022, 2023; Capone et al., 2023) – although whether the net effect of such advances in neuromorphic computing will be for good or ill remains to be seen. Further concrete computational studies of real-world big-data processing are now required. The difficulties to be overcome in doing that are far from trivial, however, because as things are at present they require high performance computational capabilities. Current chip technology has been leveraged to provide a highly flexible model of context-sensitive cells with distinct modes of operation (Pastorelli et al., 2023). Further advances in our understanding of their capabilities are likely to arise from such models even though they require high performance computing power. In the medium-term, if not the short-term, even faster advance may arise from the development of new chip designs that are purpose-built for use as adaptive cooperative context-sensitive neural information processors. Advances such as those in optomemristor technology (Sarwat et al., 2022) may make that possible in the not too distant future.

Thus, the revolutionary discovery of context-sensitive two-point neurons briefly summarized here matters much more than just within the narrow context of competing theories of brain function and predictive inference. In the long run, it may well be that descriptions of context-sensitive capabilities enabled by distinctively human pyramidal cell morphologies (Beaulieu-Laroche et al., 2018) and by inhibitory interneurons in layer 1 (Boldog et al., 2018) will come to be seen as enhancing our understanding of neural mechanisms and cognitive capabilities that are distinctively human. These capabilities include language production and interpretation, which depend on context to an exceptionally high extent, and which may arise, not simply from bigger brains, but also from advanced forms of contextual modulation at the cellular level.[11] However, we acknowledge that if future research shows modulatory PP to be far less common in mammalian brains than is subtractive coding, then that will provide strong grounds for supposing that the modulatory capabilities that have been observed at a cellular level in the mammalian neocortex are far less important than we claim them to be.

The discovery of cellular mechanisms for cooperative context-sensitive computation raises a host of unresolved genetic, cellular, network, and psychological issues that are far too extensive to be listed here but many of them are clearly stated and motivated in Phillips (2023). Empirical and conceptual advances outlined in this paper, and the emerging field of cellular psychology based on them, may even be taken as supporting Kuhn's (1962, 1970) notion of scientific revolutions. Kuhn presents a vision of science as consisting of periods of exploration and careful hypothesis testing that are occasionally punctuated by radical replacement of an underlying conceptual framework, or

paradigm, by a more adequate framework. Observations of ocean-floor spreading in the 1960s initiated a revolution in geology that is now taken for granted. Presenting those observations as hypothetical constructs would have been misleading. Similarly, presenting the functionally distinct site of apical integration that has been observed by cellular physiologists as a hypothetical construct, rather than as a discovery, would be misleading. Given the long and lasting belief in the integrate-and-fire point neuron assumption in systems and cognitive neuroscience, it was correct to infer that the many descending inputs received by neocortical pyramidal cells must be an integral part of that about which those cells transmit information – as in subtractive coding. Since the discovery of the functionally distinct site of apical integration, however, the point neuron assumption can no longer be taken for granted. Furthermore, that assumption is directly contradicted by demonstrations that neocortical pyramidal cells can operate in a way such that much unique information is transmitted about basal input but little or none about apical input, even when it has large effects on output (Schulz et al., 2021; Kay et al., 2022). The theoretical implications of the abundant input to apical dendrites from descending and other diverse sources must therefore be re-assessed, as advocated here.

## CRediT authorship contribution statement

**Tomáš Marvan:** Conceptualization, Writing. **William A. Phillips:** Conceptualization, Writing.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tomas Marvan reports financial support was provided by Czech Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## Acknowledgement

## References

Adeel, A., Adetomi, A., Ahmed, K., et al., 2023. Unlocking the potential of two-point cells for energy-efficient and resilient training of deep nets. IEEE Transac. Emerg. Topics in Comput. Intelli. 1–11.

Adeel, A., Franco, M., Raza, M., et al., 2022. Context-sensitive neocortical neurons transform the effectiveness and efficiency of neural information processing. https://doi.org/10.48550/arXiv.2207.07338.

Alilović, J., Timmermans, B., Reteig, L.C., et al., 2019. No evidence that predictions and attention modulate the first feedforward sweep of cortical information processing. Cerebr. Cortex 29, 2261–2278.

Almeida, V.N., 2022. The neural hierarchy of consciousness: a theoretical model and review on neurophysiology and NCCs. Neuropsychologia 169, 108202.

Anderson, K.M., Collins, M.A., Chin, R., Ge, T., Rosenberg, M.D., Holmes, A.J., 2020. Transcriptional and imaging-genetic association of cortical interneurons, brain function, and schizophrenia risk. Nat. Commun. 11 (1), 1–15.

Aru, J., Suzuki, M., Larkum, M.E., 2020a. Cellular mechanisms of conscious processing. Trends Cognit. Sci. 24, 814–825.

Aru, J., Siclari, F., Phillips, W.A., Storm, J.F., 2020b. Apical drive—a cellular mechanism of dreaming? Neurosci. Biobehav. Rev. 119, 440–455.

---

[11] It has recently been found that inhibitory regulation of contextual input to pyramidal cells becomes more differentiated at higher levels of abstraction in neocortex (Anderson et al., 2020; Almeida, 2022).

Aru, J., Bachmann, T., Suzuki, M., et al., 2023. Primer on the dendritic integration theory of consciousness. PsyArXic. https://doi.org/10.31234/osf.io/vkdt2.

Bachmann, T., Suzuki, M., Aru, J., 2020. Dendritic integration theory: a thalamo-cortical theory of state and content of consciousness. Philosophy and the Mind Sci. 1 https://doi.org/10.33735/phimisci.2020.II.52.

Barron, H.C., Auksztulewicz, R., Friston, K., 2020. Prediction and memory: a predictive coding account. Prog. Neurobiol. 192, 101821.

Bastos, A.M., Usrey, W.M., Adams, R.A., et al., 2012. Canonical microcircuits for predictive coding. Neuron 76, 695–711.

Beaulieu-Laroche, L., Toloza, E.H.S., van der Goes, M.-S., et al., 2018. Enhanced dendritic compartmentalization in human cortical neurons. Cell 175, 643–651.e14.

Boldog, E., Bakken, T.E., Hodge, R.D., et al., 2018. Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. Nat. Neurosci. 21, 1185–1195.

Brouillet, D., Friston, K., 2023. Relative fluency (unfelt vs felt) in active inference. Conscious. Cognit. 115, 103579 https://doi.org/10.1016/j.concog.2023.103579.

Capone, C., Lupo, C., Muratore, P., Paolucci, P.S., 2023. Beyond spiking networks: the computational advantages of dendritic amplification and input segregation. Proc. Natl. Acad. Sci. U.S.A. 120 (49), e2220743120 https://doi.org/10.1073/pnas.22207431202023.

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

Clark, A., 2013b. The many faces of precision. (Replies to commentaries on "Whatever next? Neural prediction, situated agents, and the future of cognitive science"). Front. Psychol. 4, 270.

Clark, A., 2019. Consciousness as generative entanglement. J. Philos. 116, 645–662.

Dowdle, L., Ghose, G., Moeller, S., et al., 2023. Characterizing top-down microcircuitry of complex human behavior across different levels of the visual hierarchy. bioRxiv. https://doi.org/10.1101/2022.12.03.518973.

Fiorillo, C.D., 2008. Towards a general theory of neural computation based on prediction by single neurons. PLoS One 3 (10), e3298. https://doi.org/10.1371/journal.pone.0003298.

Fiorillo, C.D., 2010. A neurocentric approach to Bayesian inference. Nat. Rev. Neurosci. 11 (8), 605.

Fiorillo, C.D., 2012. Beyond Bayes: on the need for a unified and Jaynesian definition of probability and information within neuroscience. Information 3 (2), 175–203.

Graham, B.P., Kay, J.W., Phillips, W.A., 2024. Transfer functions for burst firing probability in a model neocortical pyramidal cell. bioRxiv, 575982. https://doi.org/10.1101/2024.01.16.575982, 01.16.

Harnett, M.T., Magee, J.C., Williams, S.R., 2015. Distribution and function of HCN channels in the apical dendritic tuft of neocortical pyramidal neurons. J. Neurosci. 35, 1024–1037.

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

Heeger, D.J., Mackey, W.E., 2019. Oscillatory recurrent gated neural integrator circuits (ORGaNICs), a unifying t heoretical framework for neural dynamics. Proc. Natl. Acad. Sci. USA 116 (45), 22783–22794.

Heeger, D.J., Zemlianova, K.O., 2020. A recurrent circuit implements normalization, simulating the dynamics of V1 activity. Proc. Natl. Acad. Sci. USA 117 (36), 22494–22505.

Hertz, J., Krogh, A., Palmer, R.G., 1991. Introduction to the Theory of Neural Computation. Addison-Wesley, Redwood City, CA.

Hobson, J.A., Hong, C.C.-H., Friston, K., 2014. Virtual reality and consciousness inference in dreaming. Front. Psychol. 5, 1133.

Hohwy, J., 2013. The Predictive Mind. Oxford University Press, Oxford, New York.

Huang, Y., Rao, R.P.N., 2011. Predictive coding. Wiley Interdiscipl. Rev. Cogn. Sci. 2, 580–593.

Jadi, J.P., Behabadi, B.F., Poleg-Polsky, A., Schiller, J., Mel, B.W., 2014. An augmented two-layer model captures nonlinear analog spatial integration effects in pyramidal neuron dendrites. Proc. IEEE 102, 782–798.

Johnson, R.R., Burkhalter, A., 1997. A polysynaptic feedback circuit in rat visual cortex. J. Neurosci. 17, 7129–7140.

Kalmbach, B.E., Buchin, A., Long, B., Close, J., Nandi, A., Miller, J.A., Bakken, T.E., Hodge, R.D., Chong, P., de Frates, R., Dai, K., 2018. h-Channels contribute to divergent intrinsic membrane properties of supragranular pyramidal neurons in human versus mouse cerebral cortex. Neuron 100 (5), 1194–1208. https://doi.org/10.1016/j.neuron.2018.10.012.

Kanai, R., Komura, Y., Shipp, S., et al., 2015. Cerebral hierarchies: predictive processing, precision and the pulvinar. Phil. Trans. Biol. Sci. 370, 20140169.

Karnani, M.M., Agetsuma, M., Yuste, R., 2014. A blanket of inhibition: functional inferences from dense inhibitory connectivity. Curr. Opin. Neurobiol. 26, 96–102.

Kay, J.W., Phillips, W.A., 2011. Coherent Infomax as a computational goal for neural systems. Bull. Math. Biol. 73, 344–372.

Kay, J.W., Phillips, W.A., 2020. Contextual modulation in mammalian neocortex is asymmetric. Symmetry 12, 815.

Kay, J.W., Schulz, J.M., Phillips, W.A., 2022. A comparison of partial information decompositions using data from real and simulated layer 5b pyramidal cells. Entropy 24, 1021.

Khan, A.G., Hofer, S.B., 2018. Contextual signals in visual cortex. Curr. Opin. Neurobiol. 52, 131–138.

Körding, K.P., König, P., 2000. Learning with two sites of synaptic integration. Netw. Comput. Neural Syst. 11, 25–39.

Kuhn, T.S., 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago.

Kuhn, T.S., 1970. Reflections on my critics. In: Lakatos, I., Musgrave, A. (Eds.), Criticism and the Growth of Knowledge. Cambridge University Press, Cambridge, pp. 231–278.

Lamme, V.A.F., 2004. Beyond the classical receptive field: contextual modulation of V1 responses. In: Werner, J.S., Chalupa, L.M. (Eds.), The Visual Neurosciences. MIT Press, Cambridge, MA, pp. 720–732.

Lamme, V.A.F., 2020. Visual functions generating conscious seeing. Front. Psychol. 11, 83.

Larkum, M.E., 2013. A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. Trends Neurosci. 36 https://doi.org/10.1016/j.tins.2012.11.006.

Larkum, M.E., Phillips, W.A., 2016. Does arousal enhance apical amplification and disamplification? Behav. Brain Sci. 39, e215.

Larkum, M.E., Zhu, J.J., Sakmann, B., 1999. A new cellular mechanism for coupling inputs arriving at different cortical layers. Nature 398, 338–341.

Lee, T.S., Mumford, D., 2003. Hierarchical Bayesian inference in the visual cortex. J. Opt. Soc. Am. Opt Image Sci. Vis. 20, 1434–1448.

Linsker, R., 1992. Local synaptic learning rules suffice to maximize mutual information in a linear network. Neural Comput. 4, 691–702.

Linson, A., Parr, T., Friston, K.J., 2020. Active inference, stressors, and psychological trauma: a neuroethological model of (mal)adaptive explore-exploit dynamics in ecological context. Behav. Brain Res. 380 https://doi.org/10.1016/j.bbr.2019.112421.

Litwin, P., Miłkowski, M., 2020. Unification by fiat: arrested development of predictive processing. Cognit. Sci. 44, e12867.

Markov, N.T., Kennedy, H., 2013. The importance of being hierarchical. Curr. Opin. Neurobiol. 23, 187–194.

Marvan, T., Polák, M., Bachmann, T., et al., 2021. Apical amplification – a cellular mechanism of conscious perception? Neurosci. Consciousness 2021, niab036.

Mikulasch, F., Rudelt, L., Wibral, M., Priesemann, V., 2023. Where is the error? Hierarchical predictive coding through dendritic error computation. Trends Neurosci. 46 (1), 45–59.

Muckli, L., Petro, L., Abbatecola, C., Adeel, A., Bergmann, J., Deperrois, N., Destexhe, A., Kriegeskorte, N., Levelt, C.N., Maass, W., Morgan, A.T., Papale, P., Pennartz, C.M.A., Peters, B., Petrovici, M., Phillips, W.A., Roelfsema, P.R., Sachdev, R.N.S., Seignette, K., Self, M.W., Smith, F.W., Storm, J.F., Svanera, M., Vanduffel, W., Senn, W., Larkum, M.E., 2023. The Cortical Microcircuitry of Predictions and Context – a Multi-Scale Perspective. In: Zenodo, 1. https://doi.org/10.5281/zenodo.8380094.

Mumford, D., 1992. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol. Cybern. 66, 241–251.

Naud, R., Friedenberger, Z., Toth, K., 2023. Silences, spikes and bursts: three-part knot of the neural code. bioRxiv. https://doi.org/10.48550/arXiv.2302.07206 arXiv: 2302.07206v1 [q-bio.NC].

Ouden, C den, Zhou, A., Mepani, V., et al., 2023. Stimulus expectations do not modulate visual event-related potentials in probabilistic cueing designs. bioRxiv. https://doi.org/10.1101/2023.04.05.535778.

Pastorelli, E., Yegenoglu, A., Kolodziej, N., Wybo, W., Simula, F., Diaz, S., Storm, J.F., Paolucci, P.S., 2023. Two-compartment neuronal spiking model expressing brain-state specific apical-amplification, -isolation and -drive regimes. bioRxiv. https://doi.org/10.48550/arXiv.2311.06074 arXiv:2311.06074 [q-bio.NC].

Penny, W., 2012. Bayesian models of brain and behaviour. Int. Sch. Res. Notices 2012, e785791.

Phillips, W.A., 2017. Cognitive functions of intracellular mechanisms for contextual amplification. Brain Cognit. 112, 39–53.

Phillips, W.A., Bachmann, T., Storm, J.F., 2018. Apical function in neocortical pyramidal cells: a common pathway by which general anesthetics can affect mental state. Front. Neural Circ. 12 https://doi.org/10.3389/fncir.2018.00050.

Phillips, W.A., Clark, A., Silverstein, S.M., 2015. On the functions, mechanisms, and malfunctions of intracortical contextual modulation. Neurosci. Biobehav. Rev. 52, 1–20.

Phillips, W.A., 2023. The Cooperative Neuron: Cellular Foundations of Mental Life. Oxford University Press, Oxford, New York.

Pujol, C.F., Blundon, E.G., Dykstra, A.R., 2023. Laminar specificity of the auditory perceptual awareness negativity: a biophysical modeling study. PLoS Comput. Biol. 19, e1011003.

Ramaswamy, S., Markram, H., 2015. Anatomy and physiology of the thick-tufted layer 5 pyramidal neuron. Front. Cell. Neurosci. 9 https://doi.org/10.3389/fncel.2015.00233.

Rao, R.P.N., 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.

Rigoli, F., Michely, J., Friston, K.J., et al., 2019. The role of the hippocampus in weighting expectations during inference under uncertainty. Neural Comput. 115, 1–14.

Roth, M.M., Dahmen, J.C., Muir, D.R., et al., 2016. Thalamic nuclei convey diverse contextual information to layer 1 of visual cortex. Nat. Neurosci. 19, 299–307.

Schmid, D., Jarvers, C., Neumann, H., 2023. Canonical circuit computations for computer vision. Biol. Cybern. https://doi.org/10.1007/s00422-023-00966-9.

Sarwat, S.G., Moraitis, T., Wright, C.D., Bhaskaran, H., 2022. Chalcogenide optomemristors for multi-factor neuromorphic computation. Nat. Commun. 13, 2247. https://doi.org/10.1038/s41467-022-29870-9.

Schulz, J.M., Kay, J.W., Bischofberger, J., et al., 2021. GABAB receptor-mediated regulation of dendro-somatic synergy in layer 5 pyramidal cells. Front. Cell. Neurosci. 15.

Schuman, B., Dellal, S., Prönneke, A., et al., 2021. Neocortical layer 1: an elegant solution to top-down and bottom-up integration. Annu. Rev. Neurosci. 44, 221–252.

Seth, A., 2021. Being You. Faber, London.

Shai, A.S., Anastassiou, C.A., Larkum, M.E., Koch, C., 2015. Physiology of layer 5 pyramidal neurons in mouse primary visual cortex: coincidence detection through bursting. PLoS Comput. Biol. 11 (3), e1004090 https://doi.org/10.1371/journal. pcbi.1004090.

Shipp, S., 2016. Neural elements for predictive coding. Front. Psychol. 7, 1792. https:// doi.org/10.3389/fpsyg.2016.01792.eCollection2016.

Shipp, S., 2024. Computational components of visual predictive coding circuitry. Front. Neural Circ. 17, 1254009 https://doi.org/10.3389/fncir.2023.1254009. eCollection2023.

Shipp, S., Friston, K., 2023. Predictive coding: forward and backward connectivity. In: Usrey, W.M., Sherman, S.M. (Eds.), The Cerebral Cortex and Thalamus. Oxford University Press, New York, pp. 436–445.

Siegel, M., Körding, K.P., König, P., 2000. Integrating top-down and bottom-up sensory processing by somato-dendritic interactions. J. Comput. Neurosci. 8, 161–173.

Solomon, S.S., Tang, H., Sussman, E., et al., 2021. Limited evidence for sensory prediction error responses in visual cortex of macaques and humans. Cerebr. Cortex 31, 3136–3152.

Spratling, M.W., 2008. Predictive coding as a model of biased competition in visual attention. Vis. Res. 48, 1391–1408.

Spratling, M.W., 2017. A review of predictive coding algorithms. Brain Cognit. 112, 92–97.

Spratling, M.W., 2019. Fitting predictive coding to the neurophysiological data. Brain Res. 1720, 146313.

Sprevak, M., 2021. Predictive coding I: Introduction. PhilSci Archive. http://philsci-arch ive.pitt.edu/id/eprint/19365.

Sun, Z., Firestone, C., 2020. The dark room problem. Trends Cognit. Sci. 24 (5), 346–348.

Takahashi, N., Oertner, T.G., Hegemann, P., et al., 2016. Active cortical dendrites modulate perception. Science 354, 1587–1590.

Tantirigama, M.L.S., Zolnik, T., Judkewitz, B., Larkum, M.E., Sachdev, R.N.S., 2020. Perspective on the multiple pathways to changing brain states. Front. Syst. Neurosci. 14, 23.

van Versendaal, D., Levelt, C.N., 2016. Inhibitory interneurons in visual cortical plasticity. Cell. Mol. Life Sci. 73, 3677–3691.

Walsh, K.S., McGovern, D.P., Clark, A., et al., 2020. Evaluating the neurophysiological evidence for predictive processing as a model of perception. Ann. N. Y. Acad. Sci. 1464, 242–268.

Wang, X.-J., Yang, G.R., 2018. A disinhibitory circuit motif and flexible information routing in the brain. Curr. Opin. Neurobiol. 49, 75–83.

Wibral, M., Lizier, J.T., Priesemann, V., 2015. Bits from brains for biologically inspired computing. Front. Robot AI 2. https://doi.org/10.3389/frobt.2015.00005.

Granato A, Phillips WA, Schulz J, Suzuki M, Larkum ME. Sumbitted. Cellular Mechanisms of Neurodevelopmental Learning Disabilities.

Williams PL, Beer RD. Nonnegative decomposition of multivariate information. arXiv: 1004.2515 [cs.IT]..

Williams, S.R., Fletcher, L.N., 2019. A dendritic substrate for the cholinergic control of neocortical output neurons. Neuron 101, 486–499.e4.

Wittgenstein, L., 1953. Philosophical Investigations. Blackwell, Oxford.

Xu, N., Harnett, M.T., Williams, S.R., et al., 2012. Nonlinear dendritic integration of sensory and motor input during an active sensing task. Nature 492, 247–251.

Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.C., Jenvay, S., Miyamichi, K., Luo, L., Dan, Y., 2014. Long-range and local circuits for top-down modulation of visual cortex processing. Science 345 (6197), 660–665.