Is Predictive Coding Falsifiable?

H. Bowman[1,2,3], D. J. Collins[1], A. K. Nayak[2], D. Cruse[2]

1. School of Computing, University of Kent;

2. School of Psychology, University of Birmingham;

3. Wellcome Centre for Human Neuroimaging, UCL [honorary Professor]

Abstract

Predictive-coding has justifiably become a highly influential theory in Neuroscience. However, the possibility of its unfalsifiability has been raised. We argue that if predictive-coding were unfalsifiable, it would be a problem, but there are patterns of behavioural and neuroimaging data that would stand against predictive-coding.

Contra-predictive patterns are those in which the more expected stimulus generates the largest evoked-response. However, basic formulations of predictive-coding mandate that an expected stimulus should generate little, if any, prediction error and thus little, if any, evoked-response. It has, though, been argued that contra-predictive patterns can be obtained if precision is higher for expected stimuli. Certainly, using precision, one can increase the amplitude of an evoked-response, turning predictive into contra-predictive pattern.

We demonstrate that, while this is true, it does not present an absolute barrier to falsification. This is because increasing precision also reduces latency and increases the frequency of the response. These properties can be used to determine whether precision-weighting in predictive-coding justifiably explains a contra-predictive pattern, ensuring that predictive-coding is falsifiable.

Introduction

Predictive coding (Friston, 2018; Rao & Ballard, 1999; Clark, 2013) has proved to be one of the most influential theories in cognitive neuroscience, with many authors identifying brain responses that are consistent with the theory (e.g. Brodski-Guerniero et al, 2017; Den Ouden et al, 2012; Garrido, Kilner, Stephan & Friston, 2009; Bekinschtein et al, 2009; Shirazibeheshti et al, 2018; Witon et al, 2020). The most basic (vanilla) predictive coding

theory makes a particularly clear claim concerning the nature and presentation of evoked brain responses. We call this basic claim, the *predictive pattern*, and state it as follows: the brain's (bottom-up) evoked response to a stimulus should reflect prediction errors. That is, the size of this bottom-up evoked response should reflect the size of the prediction error, i.e. completely unexpected stimuli should generate the largest evoked response, and stimuli that are "in all senses" expected should not generate an evoked response (we give further justification that this position is prominent in the field in **Appendix 1: Further Justification of PC-Evoked model**; see inline heading *Evoked Response as Prediction Error*). Consistent with this, there are many event-related potential responses that increase in size as a stimulus becomes more unexpected: classic examples are the mismatch-negativity (Näätänen, 1995; Garrido, Kilner, Stephan & Friston, 2009), the Odd-ball P3 (Donchin & Coles, 1988) and the N400 semantic anomaly (Kutas & Federmeier, 2011). Basic predictive coding beautifully explains these phenomena.

One can interpret this link between predictive coding and evoked responses as a neuro-biological realisation of Shannon's efficient coding scheme (Shannon, 1948), a key characteristic of which is that to optimally compress communication, the more unlikely a message is, the longer/ more complex the code representing it should be. In other words, shorter codes should be reserved for more frequently occurring stimuli. If one relates the size of an evoked response to the code length, i.e. a larger, or perhaps more complex, evoked response corresponds to a longer/ more complex message being sent up the sensory processing pathway, the evoked response should be bigger/ more complex for more unexpected stimuli. This is the basic (vanilla) predictive coding theory of evoked response amplitude/ form.

However, although predictive evoked response patterns are very common, contra-predictive patterns (or strictly, contra *vanilla* predictive patterns) can also be observed – i.e. where the largest evoked response is generated by the most expected stimulus (e.g. Vidal-Gran, Sokoliuk, Bowman & Cruse, 2020; Banellis, Sokoliuk, Wild, Bowman & Cruse, 2020). As highlighted in (Bowman, Filetti, Wyble & Olivers, 2013a), a case in point is pop-out/ breakthrough effects in M/EEG studies of conscious perception (Bowman, Filetti, Wyble & Olivers, 2013a; Bowman et al, 2013 & 2014; Banellis, Sokoliuk, Wild, Bowman & Cruse, 2020). In this context, the brain is faced with stimuli presented on the threshold of

awareness, perhaps because noise has been overlaid (Davis et al, 2005) or because the brain is being deluged with fleeting stimuli (Potter & Levy, 1969; Vul, Hanus & Kanwisher, 2009; Bowman & Avilés, 2021). Here, the perceptual system is attempting to select "salient" stimuli (where the term salient is broadly defined) from amongst the noisy or overloaded background, and stimuli are perceived as a "pop-out"/ breakthrough into awareness event (Davis et al, 2005; Banellis, Sokoliuk, Wild, Bowman & Cruse, 2020; Alsufyani et al, 2019; Bowman, Filetti, Alsufyani, Janssen & Su, 2014; Harris, Miller, Jose, Beech, Woodhams, & Bowman, 2021; Alsufyani, Harris, Zoumpoulaki, Filetti, & Bowman, 2021).

A common way to incorporate these contra-predictive evoked response patterns into the predictive coding framework is to use top-down modulated precision-weighting of prediction errors, giving a refinement of vanilla predictive coding, which we call *precision-modulated Predictive Coding* (or *pmPC-Evoked*)[1]. More specifically, if one argues that expected stimuli (e.g. standards in a mismatch paradigm) are treated as higher precision, perhaps because they engage attention more strongly, then one can generate larger responses for expected stimuli, essentially because the system has more "confidence" in their processing (Kok, Rahnev, Jehee, Lau & De Lange, 2012). Indeed, such an extension of the vanilla predictive coding framework is essential in order to reflect the strong top-down attentional effects that the brain exhibits. For example, a phenomenon such as Inattentional Blindness (Simons & Chabris, 1999) seems highly contra-predictive: a man jumping in a black gorilla's costume in the middle of a basketball game would seem to be a clear prediction error, but it is not noticed by those counting passes between players in white. In order to accommodate this phenomenon, one has to assume that a strong task set turns black feature detectors right down, which, within the prediction framework, would amount to an

[1] Even in the early presentation of predictive coding by Rao and Ballard (1999), prediction errors were weighted with precisions. However, the Rao and Ballard precisions just reflected noise, specifically being the reciprocal of the sensory noise/variance; see also (Feldman & Friston, 2010). That is, precisions were not seen to be manipulable by top-down feedback. Our real interest in this paper is with top-down manipulation of these precisions; the term precision-modulated is specifically introduced to describe such top-down control of precision.

The full predictive coding theory incorporates precision weighting both on prediction errors (i.e. likelihoods) and on priors. Indeed, a good deal of the "richness" of the theory's capacity to explain psychiatric conditions is associated with relative weighting strengths of these two classes of precisions (Yon & Frith, 2021). Additionally, even in Rao and Ballard (1999), precisions were present on both prediction errors (i.e. likelihoods) and priors. However, for simplicity of presentation in this paper, we focus exclusively on precision modulation of prediction errors.

extremely low precision on black, quenching any prediction error that the gorilla may induce.

A typical precision-modulation interpretation of attention can be found in Feldman and Friston (Feldman & Friston, 2010); further justification that this position is prominent in the literature can found in **Appendix 2: Precision, Gain and Attention**. This elegantly accommodates contra-predictive response patterns with the perception-as-inference perspective (Knill & Richards, 1996; Boring, 2008) that is central to predictive coding. For example, ignoring degrees of freedom, a two-sample t-test from inferential statistics can be expressed as a product of a prediction error term (difference of means) and a precision term (reciprocal of the standard deviation of the difference of means)[2].

Bowman, Filetti, Wyble & Olivers (2013a) and more recent papers (Banellis, Sokoliuk, Wild, Bowman & Cruse, 2020, Heilbron & Chait, 2018) raised the possibility that using precision to re-weight predictive patterns to turn them into contra-predictive patterns offers considerable degrees of freedom to the theory, indeed, running the risk of generating an unfalsifiable theory[3]. In other words, predictive coding becomes tautological: any evoked response pattern can be accommodated by the theory and no experiment can be run that would produce a pattern of data that would ever stand against it.

We consider predictive coding's susceptibility to unfalsifiability here. We do this with simple neural simulations of evoked response patterns, where, under Occam's Razor, we consider this simplicity to be an advantage. On the basis of these simulations, we then discuss how the field should effectively go forward in a fashion that could allow the possibility of falsification.

In this way, we seek to differentiate between two claims: 1) predictive coding explains a large part of the behaviour of the brain; and 2) predictive coding explains *all* of the behaviour of the brain. A positive response to the first of these seems difficult to argue against – there is a substantial extent to which the brain seeks to predict the world. This

[2] In fact, Cohen's d would be an exact statistical analogue of this concept.

[3] Essentially, one has a case of the law of the excluded middle in classical logic, that is, $\vdash P \vee \neg P$, i.e. for any proposition P (which here would be a predictive evoked response pattern), the logical statement P or not P is true.

paper specifically considers whether the second of these claims is supported, or at least lays a foundation for how to empirically test it.

Methods

Neural Simulations

Our simulations use a simple predictive coding model, called *PC-evoked*, which focusses on the mechanisms that directly drive the evoked response. This is depicted in figure 1 and described in the caption; full details can be found in **Appendix 5: Details of PC-Evoked model**. In these first simulations, we are interested in the first evoked transient following the onset of a stimulus. One reason for focussing on this is that it is the brain response that can most easily be studied, as it is not contaminated by overlaid feedback components. Although, we will add a second (higher-level) circuit to this model later in the paper.

We would argue that the simplicity of our model is a virtue. Importantly, our modelling objective is not to build a neural network model that can classify stimuli, predict on a variable or even implement a working generative model. Rather, our objective is to illustrate canonical patterns of brain responses. From an Occam's Razor perspective, the simpler the model that enables you to do this, the better, i.e. if one can differentiate amongst key hypotheses with a simple model, one should prefer that. Additionally, we relate our model to the classic model by Rao and Ballard (1999) in **Appendix 1: Further Justification of PC-Evoked model**.

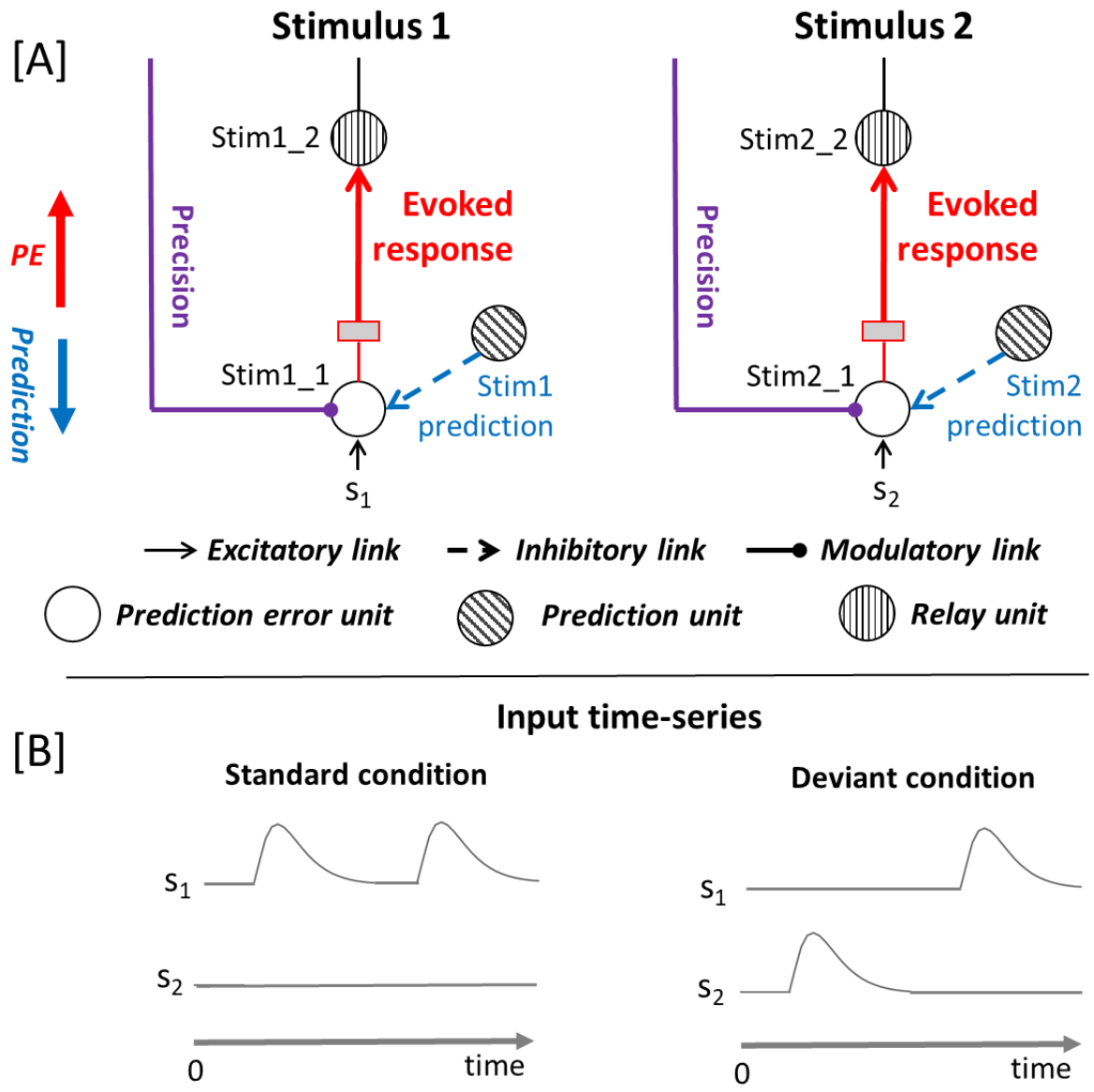

Figure 1: depiction of PC-evoked model: [A] two stimulus circuits are included, which are called Stimulus 1 and Stimulus 2. The first level contains an early prediction error unit (Stim1_1 or Stim2_1), which is excited by the input, but inhibited by a prediction unit. Thus, the activation of a prediction unit reflects how expected that stimulus is (according to recent presentations), and the activation of a first level prediction error unit can be quenched through inhibition, if it is strongly predicted. However, this activation is also modulated by a top-down precision signal, which adjusts the gain on first level prediction error units. The evoked response is modelled as the post synaptic activation entering the second level relay unit (Stim1_2 or Stim2_2). This is an analogue of the dendritic currents that are known to underlie the M/EEG signal (da Silva, 2004; Murakami & Okada, 2006). [B] Input time-series

are injected into the Stimulus 1 and Stimulus 2 circuits at s_1, respectively s_2, with gamma shaped stimulus deflections. We show the stimulus presentation associated with a Standard condition: Stimulus 1 presented twice (at s_1) and no Stimulus 2 (at s_2). The presentation for a Deviant condition involves an initial presentation (at s_2) of Stimulus 2 (deflection earlier in time) and then (at s_1) of Stimulus 1 (the deviant).

The activation equations we use are inspired by those in O'Reilly and Munakata (2000), which have similarities to those introduced by Grossberg (Ellias & Grossberg, 1975) and to Hodgkin-Huxley equations (Ermentrout et al, 2010).

Membrane potential: the membrane potential is the key measure of how excited a neuron is; its dynamics are described by the following ordinary differential equation:

$$\dot{V}(t) = \rho(t) \cdot I_{net}(t)$$

where $t \in \mathbb{R}^{\geq 0}$. Here, $\dot{V}$ is the first time-derivative of the membrane potential, $\dot{V}$; ρ is a (time-varying) neural responsiveness, and I_{net} is the net current. For simulation, the equation was discretised and numerically integrated using a 4th order Runge-Kutta method. For simplicity, neurons have the identity function as the output mapping, i.e. it is this membrane potential that is output.

Net current: the net current is a sum of excitatory, inhibitory and leak currents:

$$I_{net}(t) = I_e(t) + I_i(t) + I_l(t)$$

Individual currents: equations for the contributing currents have the same basic form:

$$I_e(t) = g_e(t) \cdot Ge \cdot (Rev_e - V(t))$$

$$I_i(t) = g_i(t) \cdot Gi \cdot (Rev_i - V(t))$$

$$I_l(t) = g_l(t) \cdot Gl \cdot (Rev_l - V(t))$$

where, first considering constants, Ge, Gi and Gl are maximum conductances, one for each channel, reflecting the maximum extent that a channel can be open, and Rev_e, Rev_i and Rev_l are reversal potentials (also called driving potentials or equilibrium channel potentials), one for each channel. $g_e(t)$ is the extent to which the excitatory channels are open at time t and mediates the action of excitatory inputs, such as those from the stimulus or a pre-synaptic unit. Similarly, $g_i(t)$ is the extent to which the inhibitory channels are

open and mediates the action of inhibitory inputs coming from the prediction units. $g_l(t)$ models the opening of leak channels, which are, in fact, always fully open, and so for all t, $g_l(t) = 1$. The reversal potentials bound the values that the membrane potential can take, with $Rev_e = 1$, giving the top of the range and $Rev_i = Rev_l = 0$, the bottom. Thus, the $\big(Rev_e - V(t)\big)$ term ensures that excitation drives the membrane potential up towards the top of its range, while $\big(Rev_i - V(t)\big)$, respectively $\big(Rev_l - V(t)\big)$, ensure that inhibition, respectively leak, drives it down to the bottom. Thus, the excitatory current has a positive polarity, while inhibition and leak are negative.

Time-dependent conductances: Additionally, the excitatory and inhibitory time-dependent conductances are set to be sums of weighted inputs. Thus, the extent to which a conductance channel is open at a particular time point, is determined by the efficiency of the synapses containing the channel and the corresponding presynaptic activations. Neuro-physiologically, the product of the presynaptic activations and their synaptic efficiencies determines the quantity of the corresponding neurotransmitter (e.g. glutamate for excitation and GABA for inhibition) that is released into the synaptic cleft, thereby opening ion channels. This electrochemical process is abstracted away from, by simply setting time-dependent conductances to sums of weighted inputs, e.g. with neuron indices added to our notation (j for the current unit and k for pre-synaptic units) for excitation,

$$g_{e,j}(t) = \sum_k w_{kj} A_k(t), \qquad where\ A_r(t) = V_r(t)$$

and similarly for $g_{i,j}(t)$, the time-dependent Inhibitory conductance. As previously discussed, for simplicity we do not include an activation function and thus, the output activation of a unit is simply its current membrane potential.

Neural Responsiveness: ρ, which in electrical terms could be related to the reciprocal of the capacitance, is defined as follows:

$$\rho(t) = \tau + (1 - \tau) \cdot \left(1 - {}^{1}\!/_{e^{\pi(t)}}\right) \qquad \text{(Eqn Responsiveness)}$$

where τ ($0 < \tau \leq 1$) is a time-constant, and $\pi(t)$ is a time-varying precision, which is subject to the constraint that $\forall t \in \mathbb{R}^{\geq 0} \cdot \pi(t) \geq 0$. Thus, the time-constant provides a basic responsiveness, i.e., update rate, but this increases as precision, π, increases, as one would

expect from an increase of gain. The relationship between precision and responsiveness is shown in figure 2, and the association of precision with responsiveness and gain is further justified in **Appendix 2: Precision, Gain and Attention**, with formal justification in **Appendix 3: Mathematical Definition of Responsiveness**.

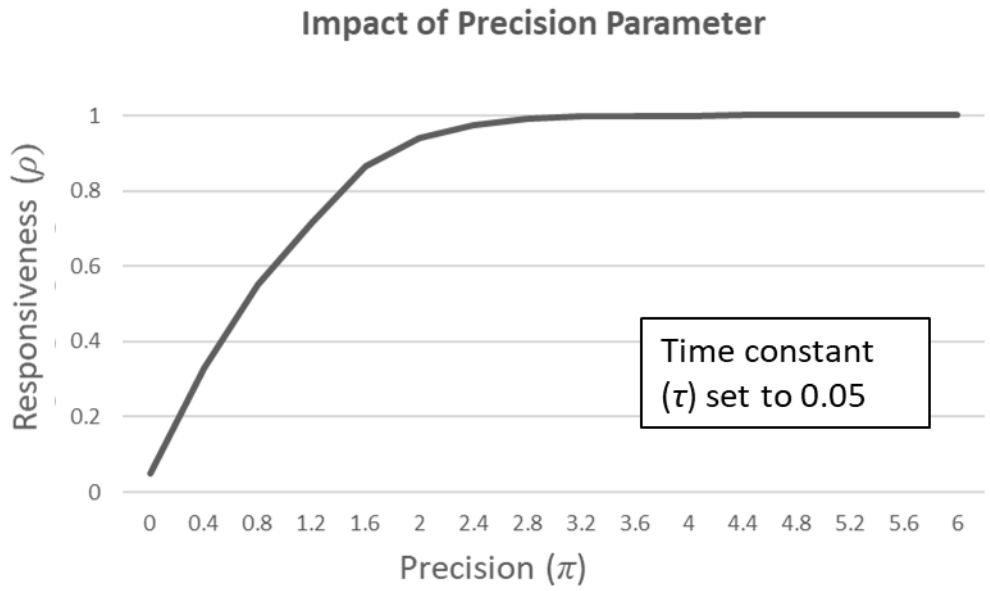

Figure 2: neural responsiveness by precision: precision (π) is shown on the x-axis and responsiveness (ρ) on the y-axis. The (basic) time constant (τ) is set to 0.05*. As a result, responsiveness is* 0.05*, when precision is zero. Responsiveness rises as precision increases, asymptotically approaching 1 for large precisions.*

Evoked response: the M/EEG signal originates from dendritic currents (da Silva, 2004), the closest analogue of which is the net current, I_{net}. Thus, the evoked response is defined as follows,

$$Evoked(t) = C \cdot I_{net}(t)$$

where, $C = -10$ scales the net current, which flips polarity, in order that our model can be related to mismatch *negativity* data[4].

[4] Physiologically, the fact that we set C to a negative number reflects the orientation of the electrical dipole in the brain to the electrode at which the component is recorded from. At the electrode the mismatch negativity is typically recorded from, it manifests as an initially negative-going component. If we could place an electrode on the other side of the electrical dipole, it would be initially positive-going. This mapping from brain dipole to electrode would be reflected in a forward/lead-field model in source localisation algorithms, which exactly map from time-series in the brain to a response at the sensor (i.e. electrode) level.

Ensemble response: one interpretation of the M/EEG signal is that it is the result of averaging over the dendritic currents of large neuronal populations. More or fewer neurons may be active in these ensembles at any one time, leading to additive effects on the measured current, $Evoked(t)$. We therefore define an ensemble response as such,

$$Ensemble(t) = Evoked(t) \cdot I_c$$

where $I_c \in \mathbb{R}^{\geq 0}$ is a scaling constant. The additive ensemble response provides an alternative response pattern to the multiplicative effects of precision realised as neural responsiveness. In this way, this ensemble response will serve as a contrast condition to which the multiplicative effects can be juxtaposed.

Running Model: When a simulation is run, all constants are set by hand and all time-varying parameters are initialized at zero. Since we are generating Event Related Potentials, we sum the activation of the Evoked response from the two stimulus circuits.

Time-frequency analysis

For our time-frequency analysis, we obtained the power of the evoked response through a Morlet wavelet transform of the data. The wavelets were defined as such:

$$\Psi(f,t) = exp(2i\pi ft) \cdot exp\left(-\frac{t^2}{2\sigma^2}\right)$$

Where f denotes the frequency of interest, t denotes time, i is the imaginary unit and σ is the standard deviation of the Gaussian envelope, defined using the (frequency-varying) wavenumber *k*:

$$\sigma = \frac{k(f)}{2\pi f}$$

We analysed 50 linearly spaced frequencies ranging from 1-40Hz. The wavenumber ranged from 4 to 10 cycles and was increased logarithmically with the frequency of interest to ensure greater temporal precision of low-frequency signal components. Wavelets were convolved with the evoked signal ($Evoked(t)$) via frequency-domain multiplication after being passed through a Fast Fourier Transform (FFT). An inverse FFT was used to recover the time-domain signal and power was extracted by taking the squared absolute of this signal.

Results

Simulations

Vanilla predictive pattern

Figure 3A presents a classic predictive evoked response pattern. The PC-evoked model was run with precision set to zero. Thus, we are observing vanilla prediction errors, without precision-modulation. The response to the repeated (standard) stimulus is lower amplitude (i.e., less extreme from zero) than the response to the non-repeated (deviant) stimulus. This is caused by the inhibitory projection from the Stim1 prediction unit, which is strongly active for the second presentation in the standard condition, because stimulus 1 was previously presented, but not in the deviant condition.

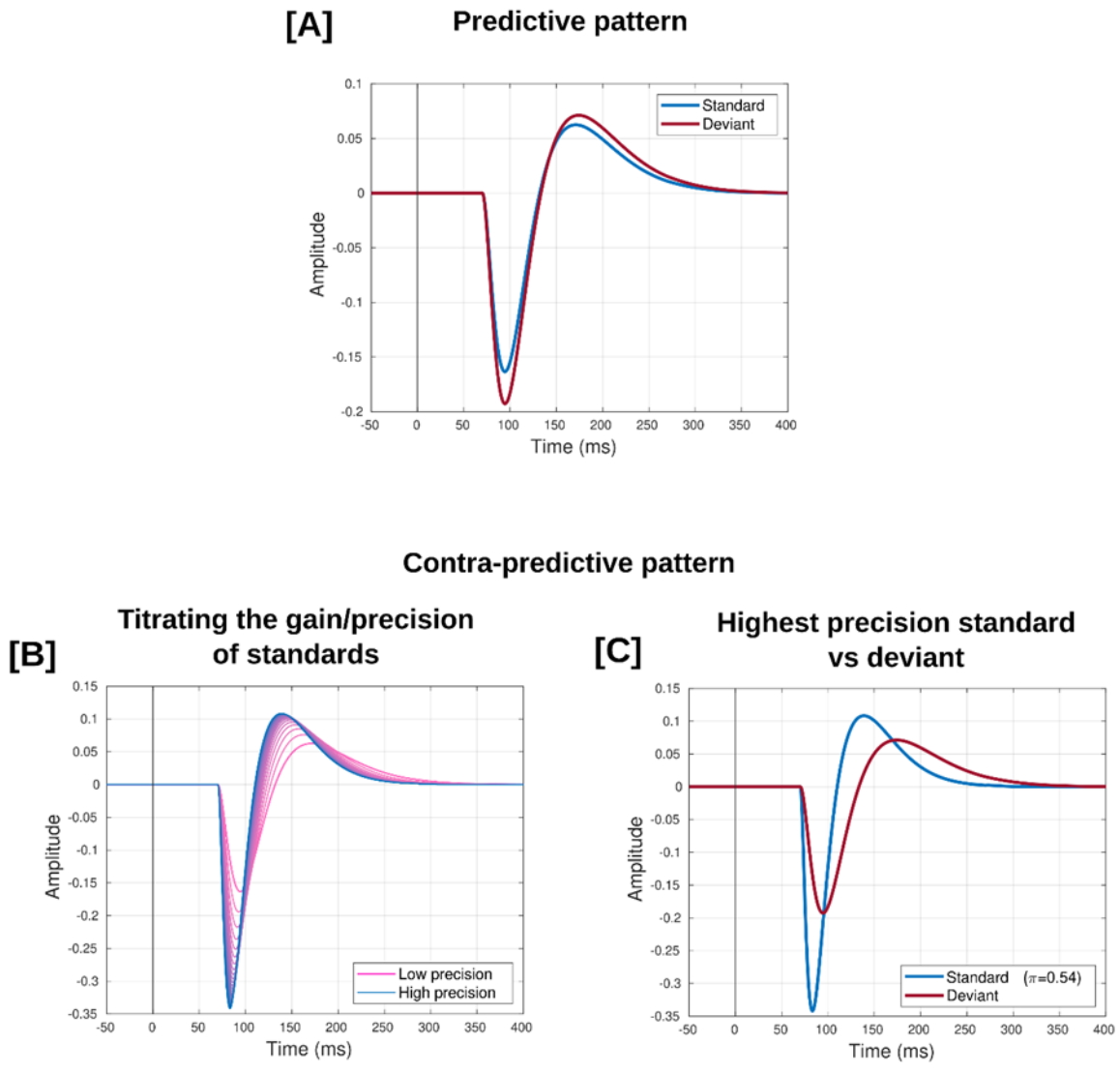

Figure 3: predictive and non-predictive patterns from PC-evoked model: in all cases, we are showing the response to the second (always Stimulus 1) of two stimulus presentations. In the Deviant case, Stimulus 2 was previously presented; in contrast, in the Standard case, it was Stimulus 1. [A] Predictive pattern, with precision parameter, π, set to zero (see Simulation 1, Appendix 5). [B] Contra-predictive pattern generated using precision parameter. By increasing precision (the π parameter) onto Stimulus 1 (but not Stimulus 2), the standard can be made higher amplitude (see Simulation 3, Appendix 5). In this way, a predictive pattern can be turned into a contra-predictive pattern, ultimately, [C] with standard substantially higher in amplitude (i.e. more extreme from zero) than deviant when precision is 0.54 (see Simulation 2, Appendix 5).

When configured with a precision of zero, the response to the Deviant will not be smaller than for the Standard, i.e., a contra-predictive pattern cannot be generated.

Contra-predictive pattern

However, by increasing the precision parameter (π), one can obtain a contra-predictive pattern from the PC-evoked model. This is shown in figure 3[B&C], where increasing precision can generate an evoked response for the standard that is larger in amplitude than the evoked response for the deviant. This is because precision is a gain parameter, which can be used to "turn-up" the evoked response. Thus, once precision is added into the PC-evoked model, and precision-modulated prediction errors are being generated, both classic predictive patterns (low or zero precision), as well as contra-predictive patterns (large precision) can be generated from the model.

Of course, there is one situation in which increasing precision would not be able to turn a predictive into a contra-predictive pattern. This is if the Standard generated zero prediction error, i.e. the stimulus was completely expected. In this situation, it does not matter how large precision (π) is, since it is multiplied with nothing, the precision-modulated prediction error (i.e. the evoked response) will be zero.

However, from a philosophical perspective, it may be argued that perfect prediction is impossible, i.e. there is always a prediction error, even if it is extremely small. Indeed, the presence of noise in the brain, might be argued to prevent a brain signal from ever perfectly matching the expected signal.

Thus, this capacity to generate both predictive and contra-predictive patterns from a theory based upon precision-modulated prediction errors, does raise the possibility that predictive coding becomes unfalsifiable. That is, one arrives at a situation in which, whatever pattern any experiment generates, it can be accommodated within the theory, i.e. there is no experiment that can (at least qualitatively) be run that could find definitive evidence *against* the theory.

However, the results in figure 3[B&C] suggest that this absolute unfalsifiability may not in fact be the case. Specifically, if precision-modulation is used to generate a contra-predictive pattern, it implies a latency change; that is, one can make the standard bigger than the deviant by increasing precision, but that has the knock-on consequence that latency shortens. Very simply, this arises from the link between precision and gain: increasing gain, increases neural responsiveness, and increased responsiveness implies reduced latency, as well as increased amplitude.

Thus, a finding of a contra-predictive pattern (evoked-standard larger than evoked-deviant) in which the latency of the standard is not less than the latency of the deviant, would stand against predictive coding.

Characteristics of Contra-predictive Pattern

Modulation of Latency: As just indicated, perhaps our main contention is that, while a contra-predictive pattern can be generated from a predictive coding model by titrating precision modulation, suggesting unfalsifiability, that titration does have consequences. These consequences yield a new set of predictions that could be the focus of further empirical work. We explored these consequences in the PC-evoked model. As shown in figure 4[A], as precision is increased, amplitude increases (more negative for a negative component). This is the basic mechanism that enables a contra-predictive evoked response to be generated from a predictive-coding model and is evident in figure 3[B&C]. Also, we can clearly see that the pattern observed is non-linear, reflecting the fact that precision is a multiplicative term in the activation equations. This non-linearity is also shaped by progression towards saturation.

Importantly, also evident in figure 3[B&C], is a reduction in latency with increasing precision. This relationship is characterised in figure 4[B]. Indeed, this reduction in latency exhibits a very similar characteristic pattern to the increase in amplitude (for a negative component).

These coincidental increases in amplitude and reduction in latency arise simply because an increase in precision is really an increase in gain. If one pushes the gain up, a system will respond both more quickly and with greater strength. This is shown by the near linear relationship between amplitude and latency observed for this particular formulation of predictive coding in figure 4[C].

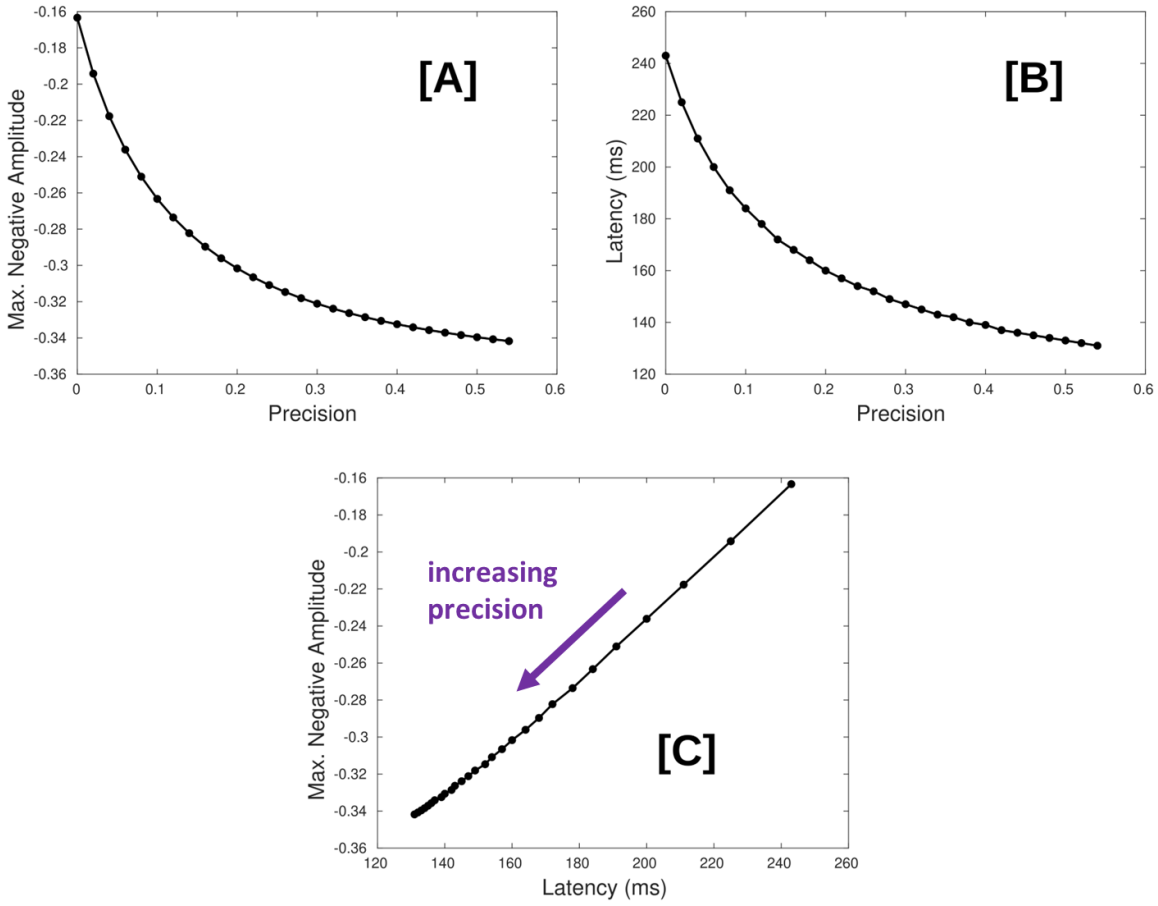

Figure 4: results of running PC-evoked model (see Simulation 3, Appendix 5) to characterise properties of contra-*predictive pattern. As precision increases, [A] component amplitude (here of first negativity) increases (down on y-axis), [B] latency of component decreases, and [C] for this configuration of the model, an almost linear relationship between amplitude and latency is observed.*

Thus, the relationships characterised in figure 4, suggests a constraint on the contra-predictive pattern that can be generated by the predictive coding framework. That is, if a claim is made that precision is enabling an empirically observed contra-predictive pattern to be viewed as consistent with predictive coding, then that argument can only be sustained if latencies reduce (or at least do not increase) with the putative increase in precision.

Evoked Frequency Characteristics: the contra-predictive pattern shown in figure 3[B&C] also generates characteristic evoked patterns in the frequency domain; see figure 5. This is nothing more than a change of the data feature space. However, it may be that time-frequency plots offer a particularly clear representation of the discriminating features of the contra-predictive pattern. In particular, we can identify the following characteristics of the time-frequency plots for a Standard, as precision increases.

1. The maximum of the power feature moves to higher frequencies as precision increases; see figure 5, particularly panels C, and D. The former of these shows the qualitative change in the frequency feature with amplitude differences normalised away. Changes to precision, and thereby to the gain, can also be viewed as adjusting the effective time constant. Increasing the time constant makes the neuron more responsive; that is, in response to stimulation, the neuron will increase its membrane potential more rapidly, as well as, decaying faster when driving input is removed. The resulting change in the shape of the evoked components, which can be seen in figure 3[B&C], generate the increase in maximum frequency.
2. As evident in figure 4, increasing precision, increases amplitude (more extreme from zero) and reduces latencies. In the frequency domain, this manifests as an increase in power (not shown in figure 5 due to normalisation) and reduction in latency of the point of maximum power (see figure 5[C,D]).

Figure 5[D] is probably the best summary of the changes in time-frequency features we are proposing to accompany the generation of a contra-predictive evoked response pattern from predictive coding. It can be clearly seen that the increase in precision causes a simultaneous reduction in latency and increase in frequency, here with a linear trajectory.

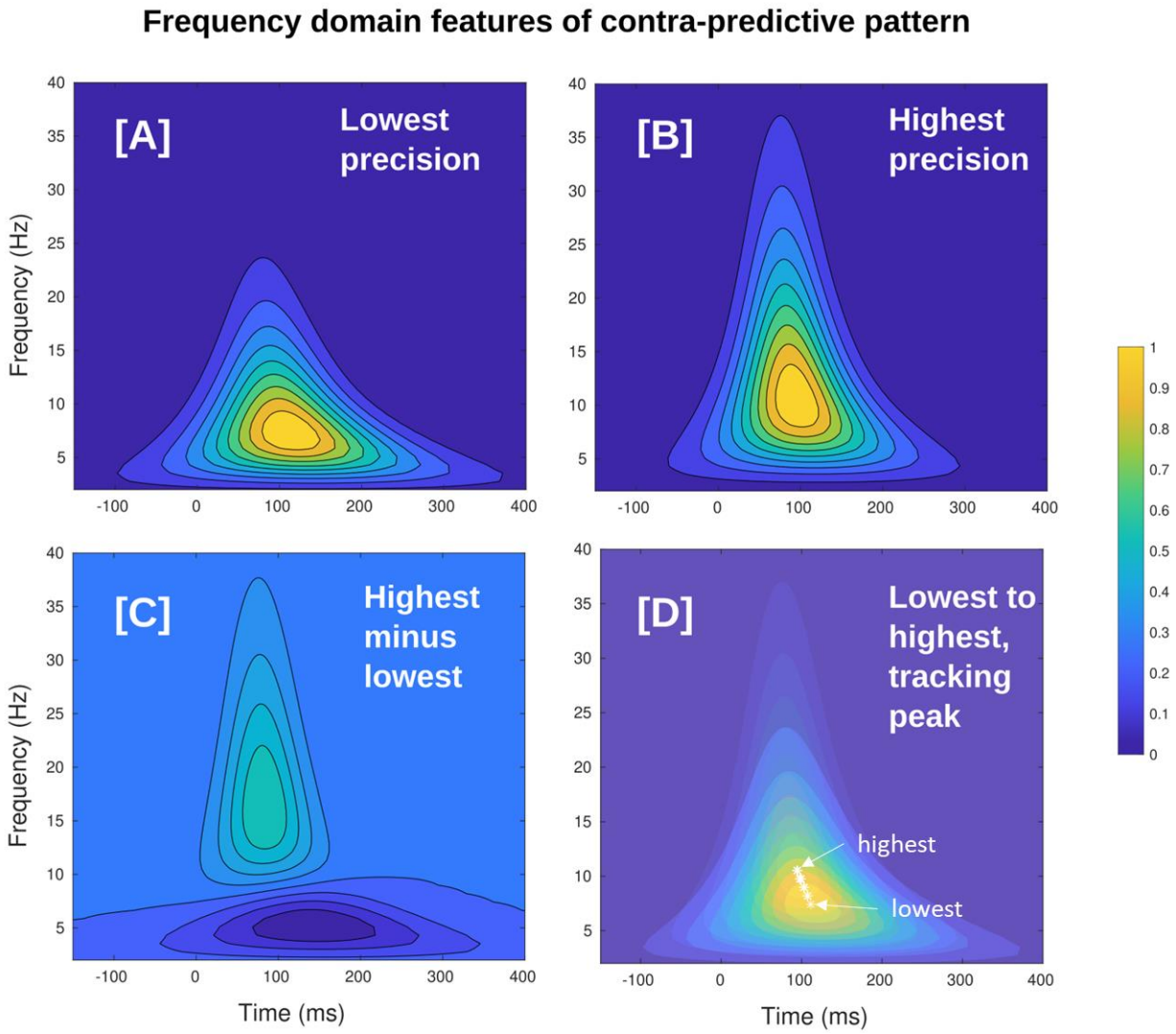

Figure 5: frequency domain features of contra-predictive pattern obtained from PC-evoked model (see Simulation 3, Appendix 5). Panels A, B and C are simple time-frequency plots; panel D contains four such plots that are overlaid on top of each other, with some transparency added to each constituent plot. [A] time-frequency feature obtained when precision is low. [B] time-frequency feature obtained when precision is high. Panels A and B have been amplitude-normalised, such that the maximum power was one and the minimum power zero in both plots. This allows one to see qualitative changes in signal, unobscured by amplitude differences. [C] panel B minus panel A. [D] time-frequency plots for four values of precision overlaid on one another, with the time-frequency point of maximum power indicated for each plot. Clearly, as precision increases, the point of maximum power moves simultaneously earlier and to higher frequencies, here following a linear trajectory.

Contrast with Additive Ensemble Effects: Are there ways of producing contra-predictive patterns that lack the empirical foothold of latency modulation? It is possible to produce contra-predictive patterns without twin amplitude-latency effects by titrating a scaling constant of the evoked response. The scaling modulation acts to model additive ensemble effects, i.e. the recruitment of a different quantity of neurons in the response. As evoked potentials find their origin in current summation over the dendrites and soma of responding cells, the ensemble effect is additive and only modulates amplitude. We can therefore contrast this response pattern to the contra-predictive pattern produced by precision-modulation.

As shown in Figure 6, titrating the scaling constant allows one to produce contra-predictive evoked responses. Increasing the constant will increase the amplitude, as is the case with precision-modulation. However, whereas the effects of precision modulation are non-linear and saturating, (as implemented in PC-evoked) the effects of scale modulation are linear and non-saturating. Most importantly, there is no reduction in latency with increased scaling. The system is not responding with greater speed, only greater strength. This strength can be smoothly and linearly modulated, in the context of the model, to achieve any desired amplitude pattern – predictive or contra-predictive.

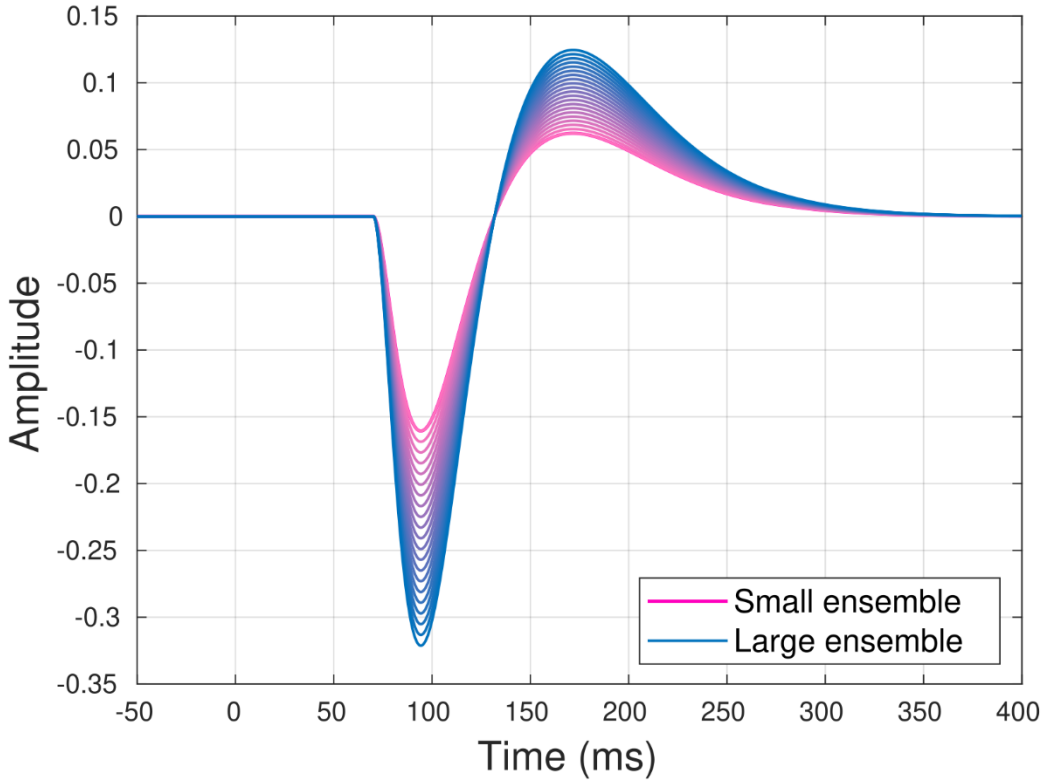

Figure 6: contra-predictive pattern generated using the scaling parameter (see Simulation 2, Appendix 5). As in figure 3[B&C]*, we take a standard response (to Stimulus 1 after a previous Stimulus 1 presentation) and titrate the value of a parameter in order to increase the amplitude of the evoked response. Here, we have increased the value of the scaling parameter,* I_c*, from 1 to 2 in steps of 0.05.*

The frequency characteristics of ensemble-modulation (found in Figure App 5[A,B,C] in appendix 6) also differ from those of precision-modulation. An ensemble-modulation increases the power across the component, and particularly at the point of maximum power. The component becomes broader both in time and frequency. However, the peak of the component remains stationary as the scaling constant is increased. Thus, the peak frequency does not change.

This ensemble-modulation hypothesis corresponds to one of the two most prominent theories of how ERP components arise in the brain: a pure power increase rather than a phase-reset (Fell, Dietl, Grunwald, Kurthen, Klaver, Trautner, ... & Fernández, 2004; Min, Busch, Debener, Kranczioch, Hanslmayr, Engel & Herrmann, 2007). That is, an ERP component could increase simply because more neurons (of the same basic kind) are

activated in response to a stimulus presentation, generating a simple increase in power, without a corresponding increase in phase consistency across trials (the marker of a phase-reset).

In the context in which we are considering this additive ensemble increase to happen, i.e. when a stimulus is expected, one obtains a theory quite different to predictive coding (see figure App 5 in appendix 6), regardless of whether prediction errors are vanilla or precision-weighted. That is, the more expected a stimulus is, the more neurons become excited, and indeed, we will argue, see subsection "The P3 in Rapid Serial Visual Presentation (RSVP)", that presenting stimuli on the fringe of awareness may be a way to elicit higher amplitude responses for expected stimuli.

This "more neurons for more expected stimuli" hypothesis contrasts with what one would expect from the Shannon efficient coding theorem (Shannon, 1948), which would suggest that more neural/ representational resource should be deployed to represent more *un*expected stimuli. Indeed, notwithstanding the discussion early in the section "Contra-predictive pattern", if it were possible for a stimulus to be 100% expected, there would be no need for any prediction error neurons to be active.

Sustained Prediction

We can also ask whether the PC-evoked model makes predictions about predictive coding more broadly – predictions that could be used to test the veracity of both vanilla and precision-modulated predictive coding. The common denominator between the two variations is the suppression of prediction error units via top-down inhibition[5] and the propagation of a prediction error through the cortical hierarchy. We could ask: what happens to the prediction error as the stimulus approaches complete predictability?

We therefore presented the PC-evoked model with 45 repeated stimuli in close succession. Predictive coding might be considered to suggest that – as the stimulus becomes more and more predictable – the prediction error will tend toward zero. In other words, the stimulus

[5] Although, see Rauss and Pourtois (2013) for an alternative view of the use of the terms top-down and bottom-up in predictive coding.

will become completely predictable, and so, one might expect that the prediction units will perfectly inhibit the prediction error units, generating a null response.

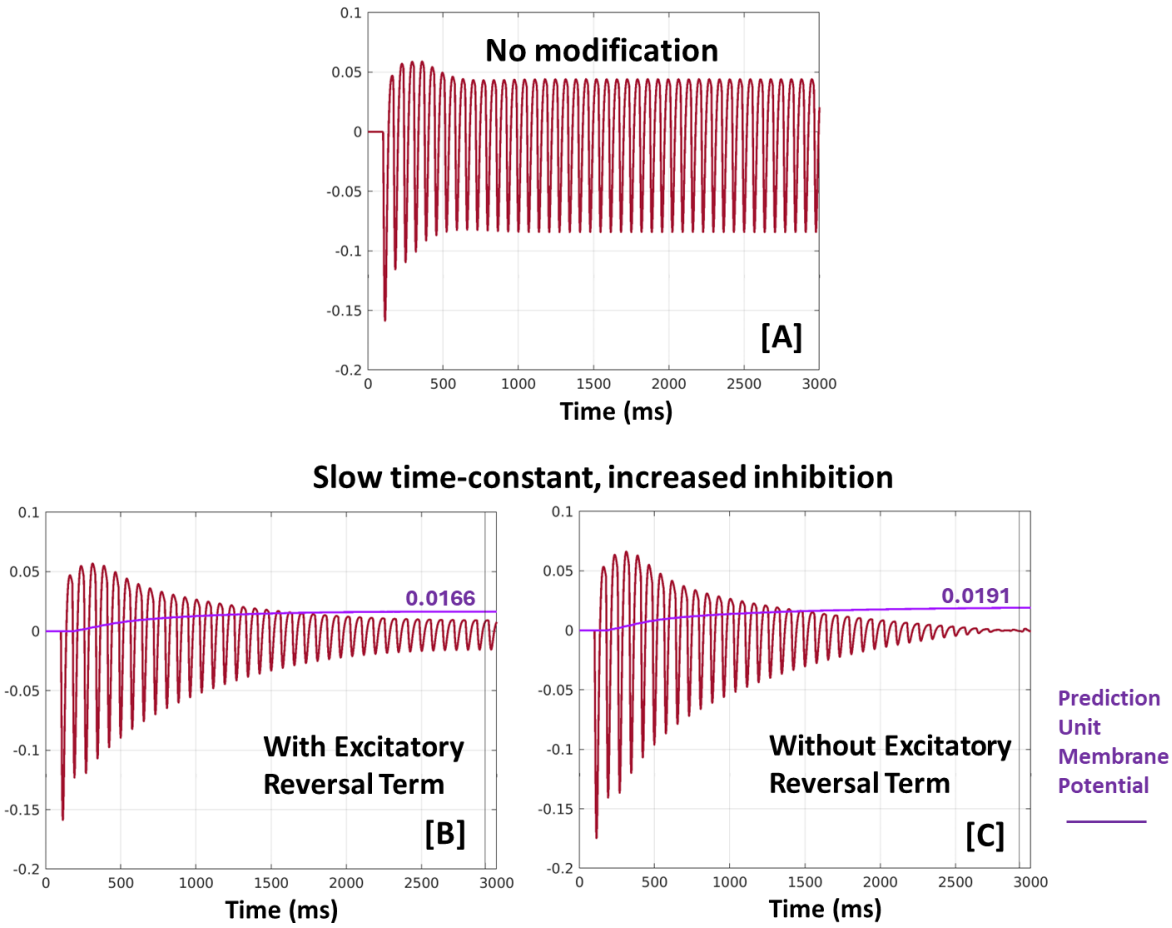

Figure 7: evoked responses from PC-evoked model under repetition of stimuli (see Simulation 4, Appendix 5). [A] evoked responses to repetitive stimuli with no modifications to the model from earlier sections. [B] evoked responses to repetitive stimuli after modifications to model. The time-constant, τ_p, for the Stim1 prediction unit was reduced from 0.04 to 0.005, i.e. stimuli induce more temporally sustained predictions. The weight from the prediction unit to the prediction error unit was increased from 14.5 to 100, i.e. much stronger suppression of predicted stimuli. [C] same settings as [B] in all respects apart from removal of Excitatory Reversal term (see Simulation 5, Appendix 5). Purple line in [B] and [C] is the membrane potential of the prediction unit, showing that it ends up higher in [C] than in [B]. This is due to the removal of the Reversal term from the prediction unit, which, in [B], limits the excitatory drive, i.e. constrains how excited the prediction unit can get, and thus how much it can suppress the bottom up response.

What we find in the PC-evoked model (Figure 7[A]) is a large prediction error for the first presentation followed by a rapid stabilisation of response amplitude to the successive presentations of the stimuli. The prediction error amplitude stabilises by around the seventh or eighth presentation and is not much lower than the onset transient at the start of the stream.

In the attempt to generate a null response (zero prediction error), we modified the PC-evoked model in two ways: reducing the value of the time constant and increasing the weight of the prediction unit's projection. This has the effect of allowing prediction to 'stack-up' more effectively over time by slowing down the return of the membrane potential of the prediction unit to its resting value. Additionally, the increased inhibitory weight increases the suppression of excitation in the prediction error units. These modifications result in a prolonged decrease of the prediction error with each stimulus presentation (Figure 7[B]), leading to a very substantial reduction in amplitude relative to the onset transients, although the attractor dynamics in the PC-evoked model mean that the prediction error is never completely quenched.

We then went further in attempting to reach an absolute quenching of the prediction error response (Figure 7[C]). We removed the reversal term in the calculation of excitatory currents. The relevant term in our equations is as follows:

$$I_e(t) = g_e(t) \cdot Ge \cdot (Rev_e - V(t))$$

We change this term to the following:

$$I_e(t) = g_e(t) \cdot Ge$$

This change brings the model more into line with Rao & Ballard's equations (see subsection *Removal of Excitatory Reversal Term* in appendix 1, Further Justification of PC-Evoked model). Our simulations show that with this change, one can obtain a full quenching to zero. This is because the upper bound on excitation (which is present with our basic equations) has been removed and the prediction unit can become more excited and clamp down further on the prediction error unit.

In conclusion, perhaps surprisingly, the combination of the attractor dynamics of the predictive coding circuit and the possibility that predictions may evaporate rapidly, means

that predictive coding, as realised in the (basic) PC-evoked model, does not definitively imply the possibility to completely quench the evoked response. It is within the parameter space of the model to obtain a substantial (even complete if the excitatory reversal potential is removed) quenching, but it is not mandated.

Informal Predictions of Contra-predictive Pattern

Taking inspiration from the PC-evoked model, we can also highlight some informal predictions. Notably, these informal predictions are in some sense more general then our previous predictions, since they are not dependent upon our model of evoked responses.

Sensory Noise vs Attention

One aspect of a basic predictive coding theory is that noise (for example, sensory noise) and attention act on the same variable, i.e. precision. One could parametrically manipulate the sensory noise on its own and attention on its own and ask whether these two manipulations have the same effect on the features of the evoked response. If they do have different effects on these features, it suggests they are not mediated by a common mechanism, which would be precision. For example, the amplitude of the response might change linearly with one, but logarithmically with the other. One could plot the panels in figures 5 and 6 for separate manipulations of attention and sensory noise, and ask the question, do these exhibit the same relationships with latency, amplitude and frequency?

Counter-intuitive Prediction

The most telling predictions that can be made by a theory are those that would only be true if the theory were true. In this sense, you could think of such predictions as "counter-intuitive" from the perspective of all other theories. If such a counter-intuitive prediction is demonstrated, it provides strong evidence for the theory. A good example of this would be the empirical effort to verify General Relativity by observing the position of stars during an eclipse in order to measure the gravitational deflection of starlight passing near the Sun (Coles, 2001).

We tentatively offer the following prediction.

Shared channel saturation effect

Precision-modulated predictive coding would seem to imply that sensory noise and attention act on the same variable, i.e. precision. In this sense, the theory suggests that sensory noise and attention share the same "channel". This suggests that they share a ceiling. Thus, when each is manipulated alone to improve performance, they should asymptote at the same performance level.

This shared-channel also suggests the presence of an interaction.

Interaction between sensory noise and attention: as shown in figures 4 and 5, the model suggests a saturation effect on precision. Since sensory noise and attention share the same channel, elevation of precision through reduced sensory noise, should reduce the effect of attention, simply because there is less dynamic range of the precision variable for attention to act on as saturation is approached. Figure 8[B] shows a potential interaction pattern that would reflect this shared-channel saturation effect. For example, for behavioural accuracy or amplitude of a positive going magneto-electrophysiological component, the effect of attention should be reduced as sensory noise reduces. Such an interaction could be tested with a range of behavioural and physiological measures, although, the direction of the dependent variable axis would change if, for example, reaction times, component latency or amplitude of a negative going component was under consideration.

One could push this interaction phenomenon to its limit and completely quench any effect of attention. That is, one could experimentally reduce sensory noise to a point of saturation of the precision parameter, i.e. where further reduction in sensory noise has no impact on the dependent variable. If sensory noise and attention share the same channel, at this saturation point, attention should have no effect.

These interaction effects would be especially compelling if one could also show that the observed interaction is not caused by an absolute "overall" ceiling. There will be ceilings to all dependent variables, but we are specifically interested in one associated with attention and the impact of noise. So, the demonstration would be particularly telling if it were possible to demonstrate that the observed ceiling effect is specific to the precision channel and that manipulation of other variables could place performance beyond that obtained through its manipulation.

This prediction could be explored in a simple behavioural-MEEG experiment in which attention is manipulated through (highly predictable) spatial cueing (see, for example, the Posner Cueing task (Posner, 1980)) and random visual noise is overlaid on the stimuli. These predictions would suggest that as the environment becomes more reliable (i.e. less noisy), the impact of attention reduces. Indeed, the prediction might suggest that cueing has its biggest effect when sensory noise is at its highest, e.g. when it is most difficult to detect the cue and the target from amongst noise. These might be considered counter-intuitive predictions, simply because one may believe that attention would have more effect when stimuli are more easily discriminated, i.e. the environment has the least sensory noise.

Another way to think about this experiment is that it is considering whether sensory noise and attention have the same or different saturation ceilings. If they have different saturation ceilings, then they represent different variables. To be clear, this finding would not necessarily stand against the notion that attention controls gain, a notion that attention theorists have subscribed to for a very long time (e.g. Cave, 1999; Mozer & Baldwin, 2007 and Wyble, et al, 2009; and see subsection “Confidence, Attention and the Predicted” in the Discussion section). Rather, it would suggest that attention does not act upon the precision variable originally conceptualised in predictive coding theories (Rao & Ballard, 1999), as variability due to noise.

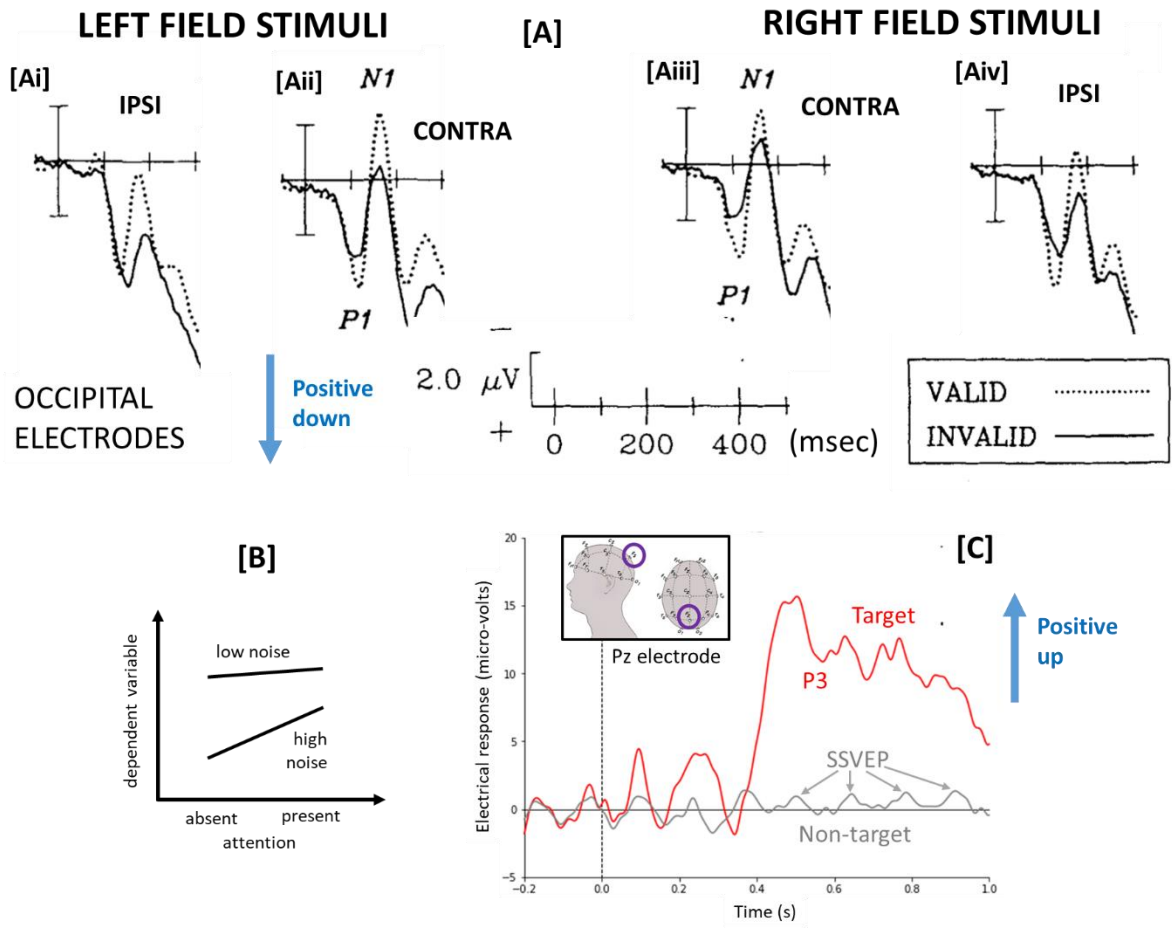

Figure 8: [A] ERPs of Posner task from Mangun & Hillyard (1991). Occipital electrodes are shown. Time-series are shown from the onset of the target (following a central cue, which pointed with equal probability towards either the left or the right). The target could appear in either visual field, giving ipsilateral and contralateral evoked responses for target in left or in right visual fields. Positive is plotted down. The clearest pattern is one whereby both P1 and N1 are higher amplitude for valid (i.e. expected) trials. [B] potential (counter-intuitive) interaction emerging from shared-channel saturation effect: a sensory noise manipulation is crossed with an attention manipulation. Due to proximity to ceiling for the precision variable, the high noise condition (small precision) should exhibit a stronger effect of attention then the low noise condition (high precision). The dependent variable could be behavioural, e.g. accuracy, or physiological, e.g. component amplitude (positive going). Reaction time or component latency could also serve as the dependent variable, but with the direction of the dependent variable reversed. [C] A typical RSVP experiment, with positive plotted upwards. Each individual distractor appears very rarely (once or twice), while pre-

specified Targets appear frequently. A large evoked response is observed for the Target (a P3 component), but effectively, no such response is elicited for distractors, save for the much lower amplitude steady-state visually evoked potential, which oscillates at the frequency of the stream (7.5 hz). A control condition (here called Irrelevant) is also displayed, in which a task-Irrelevant stimulus is presented as many times as the Target. This does not induce a P3, since the Irrelevant is not being searched for. However, of most relevance here, this condition shows the sequence of transients set-up by distractors, unaffected by the occurrence of a P3, with their low amplitude relative to the P3 being evident.

Empirical Evidence

We present the following pieces of empirical evidence related to the predictive vs contra-predictive question. Importantly, our objective in this paper is not to definitively disprove predictive coding, but rather lay down an experimental framework in which it could be disproved. Part of the reason for an "Empirical Evidence" section is to highlighted published experiments that could be adjusted to become valid tests of the predictions we have identified.

Contra-predictive pattern in Posner task

Mangun & Hillyard (1991) observed a strongly contra-predictive ERP pattern for early transients on the Posner task. Figure 8A reproduces their data, in general showing a much larger P1 (and N1) transient for the validly cued target. This is the opposite pattern to that expected by vanilla predictive coding. Typically, the increased amplitude for valid cuing looks most like a scaling (additive ensemble) effect, apart from Ipsilateral in the right visual field (see panel Aiv), which might be exhibiting a pattern consistent with an increased gain.

The P3 in Rapid Serial Visual Presentation (RSVP)

Bowman, Filetti, Wyble & Olivers (2013a) highlighted the P3 evoked response in RSVP streams as a contra-predictive pattern. Importantly, the P3 may behave quite differently in the context of conscious break-through experiments compared to experiments in which all stimuli are presented clearly above the threshold of awareness; see discussion section of Pincham, Bowman & Szucs (2016). In particular, classic Odd-Ball experiments, where conscious break-through is not an issue, elicit canonical (vanilla) predictive patterns (Polich, 1986; Donchin & Coles, 1988).

However, in RSVP search experiments, participants look for and find the same target in very many trials, thus the target stimulus becomes highly predictable. Nonetheless, contrary to a vanilla predictive pattern, the target elicits a very high amplitude evoked response.

The ERP in figure 8[C] shows this phenomenon. RSVP streams were of faces presented at an SOA of 133ms, i.e. with a presentation frequency of 7.5 hz. The band passed by the filter was 0.1 to 30 hz. Within a block, a single Target was searched for, which was a famous face, e.g. the face of Donald Trump. The target was presented 12 times during a block. Distractors were sampled at random (with replacement) from a large database of (560) faces. Within a block, most distractors that occurred, were only presented once. Thus, they are much less expected than targets. For more details see, Aviles, Anderson, Orun, Gibson, Solomon, Via, Bowman (2023).

Thus, this is a contra-predictive pattern. Indeed, even though they are highly unexpected, distractors neither attract attention, in fact, they are largely rejected subliminally (Avilés, Bowman & Wyble, 2020; Bowman & Avilés, 2021; Bowman, Filetti, Alsufyani, Janssen & Su, 2014), or generate a substantial evoked response, unlike the (highly predicted) target. This data and that presented in the previous subsection “Contra-predictive pattern in Posner task" are the sort of data that precision-modulation is required to explain.

Steady State Response

As discussed earlier, one might think that the evoked response should reduce in amplitude very substantially if one continued to present the stimulus. As shown in Figure 7, this could be the case (panels [B] and [C]), but it does not have to be (panel [A]). In fact, evoked responses to long trains of repeating stimuli have been extensively explored. A typical pattern of data is shown in figure 9, where the onset of the stream of stimuli generates a transient response, which might be related to a prediction error. However, after a number of repetitions, the evoked response settles into a relatively stable oscillation at the frequency of the visual stimulation, which has an amplitude not much lower than the evoked (onset) transient. In particular, there is little evidence of substantial attenuation of the response. Thus, the data looks more like figure 7[A] than figure 7[B or C].

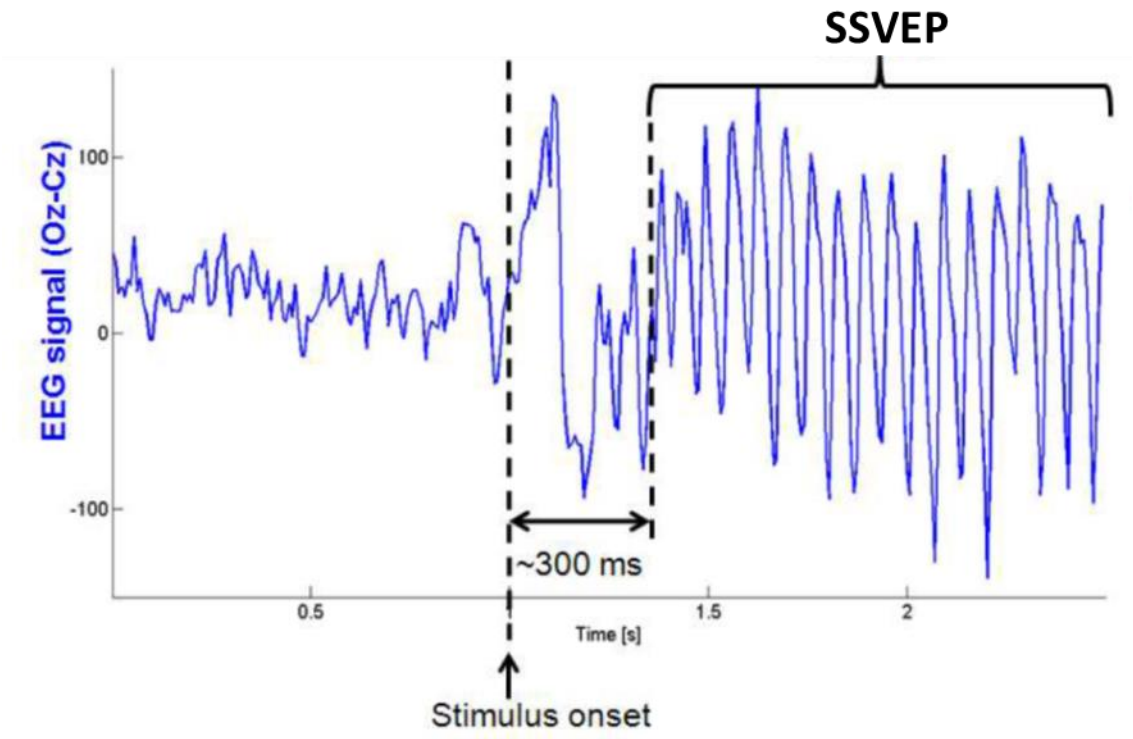

Figure 9: steady state visually evoked potential from (Garcia-Molina & Milanowski, 2011). A visual stimulus was repeated at 15 Hz.

Shared-channel noise versus attention prediction

Interestingly, there is a literature focused on the impact of noise on higher cognition, e.g. Moss et al (2004) and some of this has considered the interplay between sensory noise and spatial attention, e.g. Dosher & Lu (2000) and Herweg & Bunzeck (2015). These studies could inform our shared-channel noise versus attention behavioural prediction. The most relevant study is Dosher & Lu (2000), since they manipulated noise in the same modality as attention (Herweg & Bunzeck (2015) added *auditory* noise to a *visual* Posner task).

One might believe that the more direct test of the shared-channel noise vs attention prediction would be when noise is applied in the same modality in which attention is manipulated. Dosher & Lu (2000) employed a form of Posner task that tested the effect of overlaid visual noise on an orientation judgement task. We reproduce their key finding in our Figure 10.

They observed a pattern consistent with our prediction, with a strong attentional effect with high sensory noise, but no attentional effect in the absence of sensory noise (in fact, this no-

sensory noise condition exhibited a strong ceiling effect; see our Figure 10 and also Figure 2 in Dosher & Lu (2000)). Remember, our prediction was that, if attention and levels of sensory noise share the same channel (i.e. precision), the effect of attention should increase as sensory noise increases (i.e. induced precision reduces), since the resulting greater distance to ceiling, would give more room for attention to act.

Dosher & Lu (2000) fitted their perceptual-template model (PTM) to the full data pattern and explained the findings in terms of perceptual-template sharpening. A very interesting direction for further work would be to formulate a model comparison between the Dosher & Lu perceptual-template model and a shared-precision predictive coding model on results of their experimental paradigm.

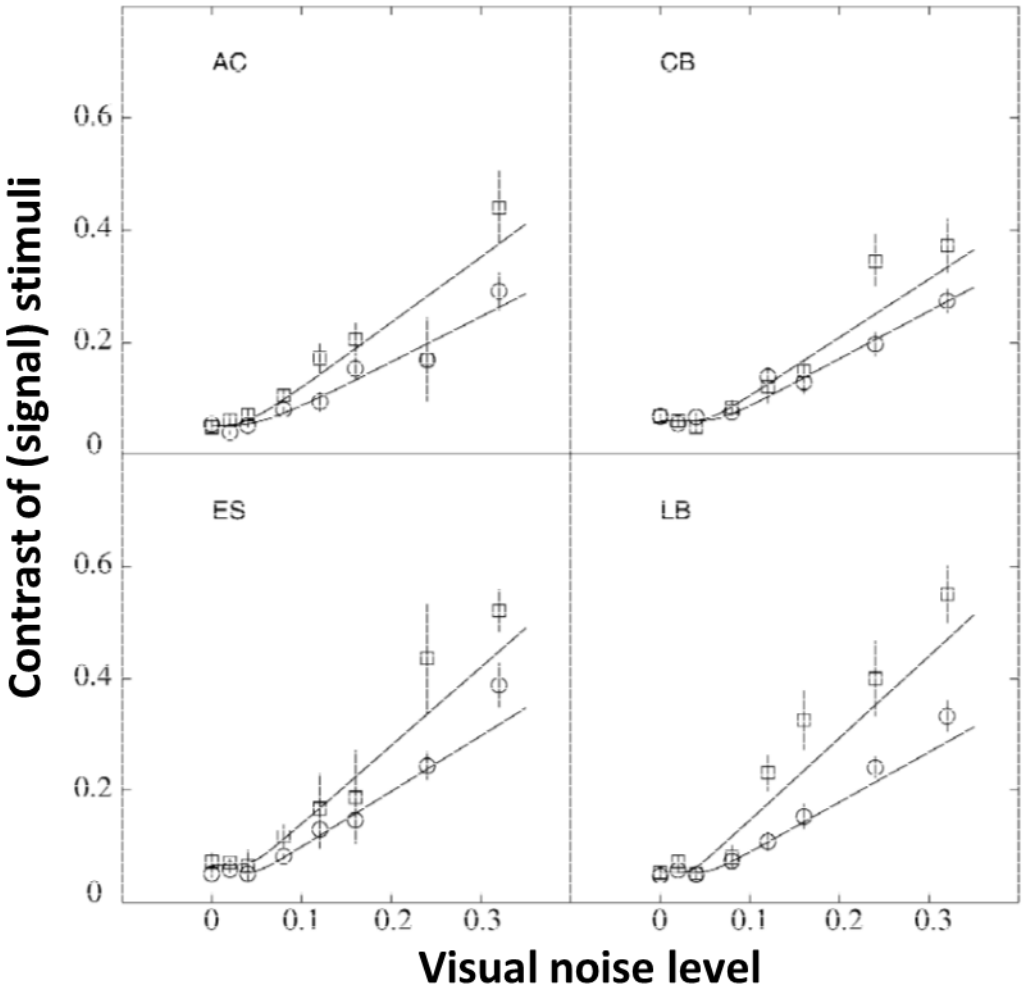

Figure 10: Results from Dosher & Lu (2000) (their figure 3). Data for the four participants are shown, one per quadrant. X-axis is the contrast level of the overlaid visual noise, i.e. noise

level increases from left to right. Y-axis shows the contrast of the (signal) stimuli, i.e. stimulus strength increases from bottom to top. Data points are level of stimulus contrast at which 62.5% performance accuracy is obtained for a particular noise level. Thus, a data point higher on the y-axis indicates worse performance, i.e. that a stronger stimulus was required to obtain the 62.5% performance level, for a given level of noise. In all quadrants, the higher curve (squares) is for invalid cuing (i.e. unattended) and the lower curve (circles) is for valid cuing (i.e. attended). This, then, is the same pattern as our interaction prediction (figure 8[B]), but with y-axis reversed, since stimulus strength required to reach a performance threshold is plotted (i.e. lower is better performance). That is, we observe a large effect of attention when sensory noise is high and a small effect when noise is low.

Finally, although not explicitly prediction experiments, there has been neuroimaging work crossing manipulation of sensory noise (clear speech versus noise vocoded) and attention (dichotic listening) (Wild, Yusuf, Wilson, Peelle, Davis & Johnsrude, 2012; Rimmele, Golumbic, Schröger & Poeppel, 2015). Adding a prediction component to these paradigms could be a fruitful direction for research.

2nd-level Prediction Circuit: the Breakthrough-P3

We focus here on modelling Rapid Serial Visual Presentation (RSVP) and specifically, the SSVEP and P3 observed in that context. This follows from Bowman, Filetti, Wyble & Olivers (2013a) who highlighted the P3 evoked response in RSVP streams as a contra-predictive pattern. Importantly, as previously discussed, the insensitivity to prediction of the P3 in RSVP should be distinguished from that classically observed for the odd-ball P3 (Donchin & Coles, 1988), which exhibits a pattern of data that is much more predictive in nature. Thus, what follows is only intended to obtain for the Breakthrough-P3. Additionally, what we present is not in any sense, a complete treatment of the Breakthrough-P3, rather our interest here is to show specific data patterns by building on top of the predictive coding framework.

Reasons for focusing on the Breakthrough-P3 are as follows. 1) It is a classic example of a breakthrough (into consciousness) component, which we have argued earlier may reflect a situation in which predictive coding is not strongly at play. In RSVP search experiments, participants look for and find the same target in very many trials, thus the target stimulus

becomes highly predictable. Nonetheless, contrary to a vanilla predictive pattern, the target elicits a very high amplitude evoked response. Additionally, Wierda, Taatgen, van Rijn & Martens (2013) found little evidence of an effect of (pre-experimental) word-frequency on the evoked response during an (RSVP) attentional blink experiment. 2) It may be that there is more "room" to observe precision/gain bringing the component earlier in time with the P3, rather than the N1/P1 (the components we have focussed on so far in this paper), which, as early components, are closer in latency to the physiologically minimum possible latency (thus, if such a latency decrease is not observed, it may represent a more compelling finding for the P3, since there was considerable "room"/potential for it to be observed). 3) For the P3, a model exists that proposes an additive effect of attention, viz the blaster response in the Simultaneous Type/ Serial Token model (Bowman & Wyble, 2007). This model has credence because of the spectrum of effects it successfully simulates, giving the additive ensemble hypothesis credibility in the context of the Breakthrough-P3 (e.g. Bowman & Wyble, 2007; Bowman, Wyble, Chennu & Craston, 2008) including modelling of the P3 (e.g. Chennu, Craston, Wyble & Bowman, 2009; Craston, Wyble, Chennu, & Bowman, 2009).

Here, then, we take the *Predictive Coding-Evoked* (*PC-Evoked*) model and simulate RSVP, the Breakthrough-P3, as well as the SSVEP, in what we call the *Hierarchical-PC-Evoked* model. We relate this revised model to Rao & Ballard's classic model under inline heading *Hierarchical Model* of **Appendix 1: Further Justification of PC-Evoked model**.

We perform two sets of simulations with the *Hierarchical-PC-Evoked* model.

Early Circuit Simulations

We make the following changes to the simple PC-Evoked model, focussed on up to this point in this paper.

1) We repeat the prediction circuits, since there are many distractors, giving us one per distractor or target. These all have absent precision, since it is not manipulated in the early circuit in these simulations. The RSVP SSVEP arises from these early circuits.

2) The target is presented more often than the distractors, but only ever across streams. We model this by having relatively small time-constants at early circuit

prediction units, enabling residual activation (in prediction units) in different circuits to build-up and last across RSVP streams.

Thus, these simulations reflect vanilla predictive coding in the early circuit, i.e. without any difference in precision between distractors and targets.

Our main finding is that once prediction has built-up across streams, the model generates a weaker response for the target, which is suppressed by residual prediction from earlier presentations in previous streams; see figure 11. Thus, this suggests that, assuming relatively long prediction dynamics, vanilla predictive coding, generates reductions in amplitudes of targets in SSVEPs once expectations have accumulated sufficiently. We are not aware of any reports of such a phenomenon in RSVP experiments, although of course, this may be because the effect is small and nobody has determinedly looked for it. This is a good focus for future work.

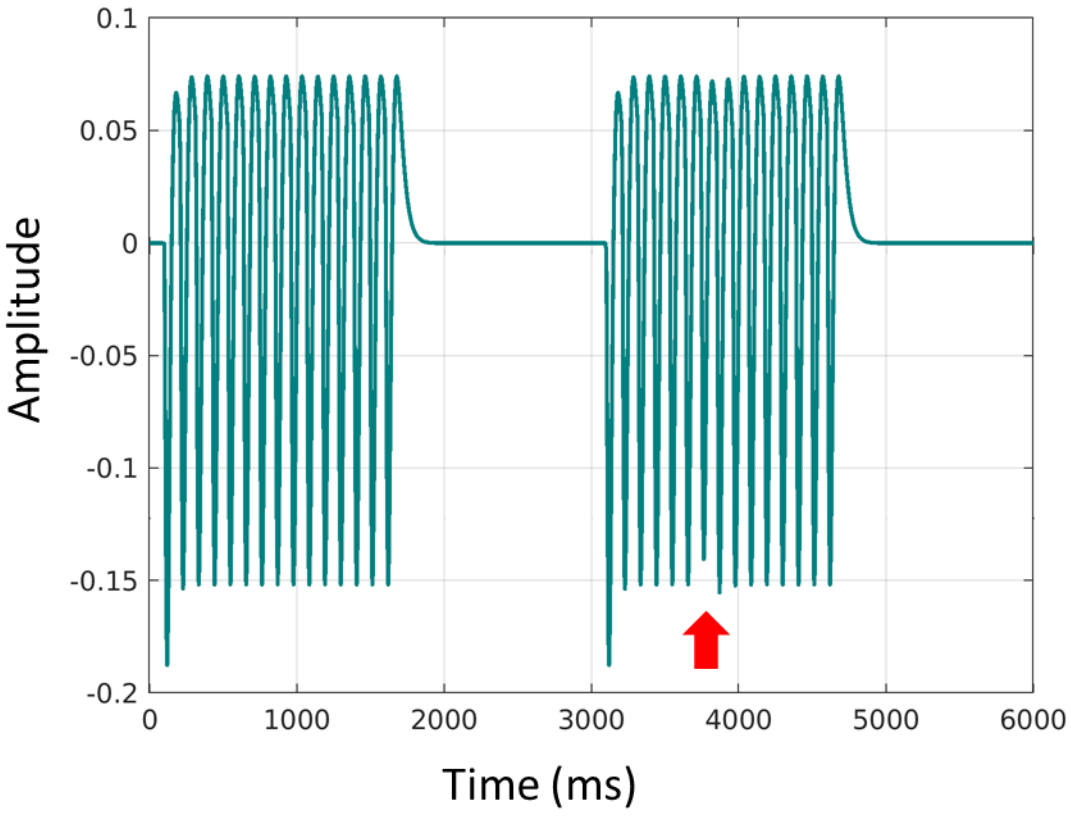

Figure 11: Early Circuit response to RSVP stimulation (see Simulation 6 in Appendix 5). Two RSVP streams (each of 15 items) are presented in succession to the model. Since an expectation carries over from the first stream to the second, the amplitude of the target is

reduced in the second stream; see red arrow. This is because distractors are not repeated across streams, but the target is.

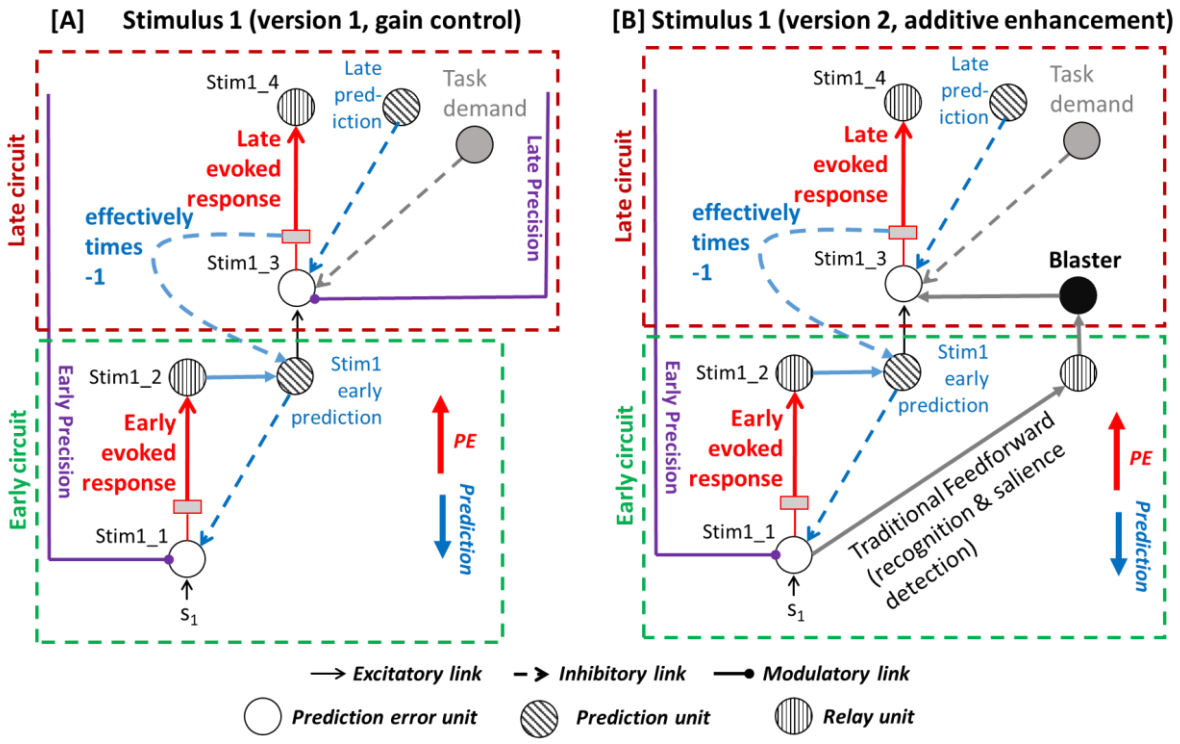

Figure 12: Hierarchical PC-Evoked models. A late circuit is added to the (early) circuit of the basic PC-Evoked model. We only depict the Stimulus 1 part of the full model. The basic Late circuit has the same general form as the Early circuit, apart from the addition of a task demand system (solid grey node), which backgrounds Distractors, preventing them from being able to generate activation in the Late circuit. We show two versions of this model. [A] (version 1, gain control) mirrors the early precision pathway, with a late precision pathway. [B] (version 2, additive enhancement), taking inspiration from the blaster in the STST model, a transient attentional enhancement mechanism is implemented, which amplifies on detection of a salient (e.g. task-relevant, emotionally salient, personally salient) stimulus. This mechanism is realised with a separate pathway from the stimulus, which is assumed to be more like a traditional recognition system that is seeking to detect stimuli that are salient to the organism.

Late Circuit Simulations

As previously discussed, see Figure 8[C], the really big response in RSVP experiments is the P3 for targets. This is even though, as just emphasized, targets are more expected than

distractors, which only elicit a single deflection in the (low amplitude) SSVEP and no P3 at all. To explore such (contra-predictive) P3s, we add a late circuit to our model; see figure 12. Specifically, we make the following changes to our model:

1) We add later higher level circuits, one for each distractor and target, but where task set ensures that only targets become active. Specifically, the inhibitory link from the task demand unit shown in figure 12 (see grey dashed arrow) is set to 0.13 for distractors and zero for targets. This blocks activation from passing into the Late circuit for Distractors, on the grounds that Distractors are not being "searched for" by participants.

2) Version 1 (gain-control): our first version of the high-level circuit (see figure 12[A]), reflects that modulation by precision works similarly for this second level response as it does at the first level (N1/P1). We show this by changing precision (late precision in figure 12[A]) in this higher-level circuit in a similar way to our manipulation of the N1/P1.

3) Version 2 (additive ensemble): as a reflection of the additive ensemble theory (see inline heading "Evoked response" of the Methods section), we import a simplified version of the transient attentional enhancement (the blaster) introduced in the Simultaneous Type/ Serial Token model (Bowman & Wyble, 2007)[6], in which the P3 scales with the level of attention, i.e. blaster firing. This mechanism amplifies on detection of a salient (e.g. task-relevant, emotionally salient, personally salient) stimulus, being realised with a separate pathway from the stimulus (see figure 12[B]). This pathway is assumed to be more like a traditional recognition system (what we call Brain as Recognizer in the Discussion) that is seeking to detect stimuli that are salient to the organism.

Version 1 of this second simulation shows that the P3 exhibits the same pattern as the N1/P1 when precision is titrated; see figure 13[A]. Thus, the large amplitude P3 observed in RSVP experiments could be obtained by precision weighting, but this second-level response

[6] Although, this did change in the eSTST model, where a multiplicative attentional enhancement was implemented (Wyble, Bowman & Nieuwenstein, 2009).

(i.e. the P3) changes in the manner suggested by our precision-weighted effects, i.e. as precision increases, the response is higher amplitude, higher frequency and earlier.

Version 2 shows a plausible alternative to a predictive coding explanation of the Breakthrough-P3. That is, this component may be generated by an additive attentional enhancement, whereby the attentional enhancement increases with stimulus salience, but with an additive, rather than multiplicative, effect; see figure 13[B] and compare to figure 6, but now, of course, with polarity reversed, since we are considering an initially positive going, rather than negative going effect[7].

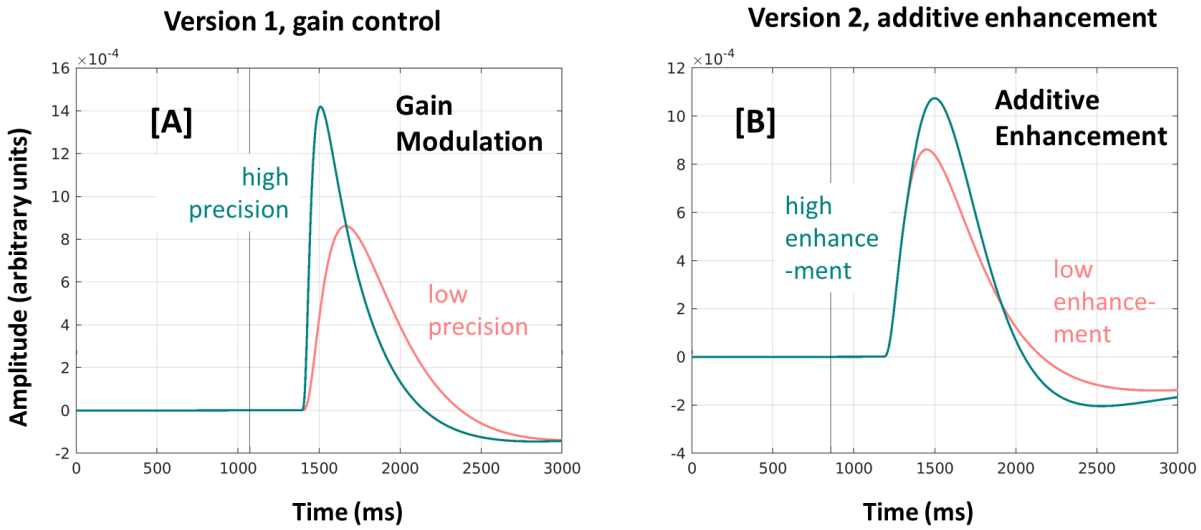

Figure 13: Competing predictions for the Breakthrough-P3 from Hierarchical PC-Evoked model. Responses correspond to late evoked response in figure 12. [A] Prediction from precision-weighted prediction error theory, in which precision/gain (π_E) is modulated by stimulus salience/engagement of attention, low precision was 0 and high precision was 0.54 (see Simulation 6, Appendix 5). As expected, increasing gain, increased amplitude, increased component frequency and reduced latency. [B] Prediction from additive (ensemble) enhancement theory, where an additive effect of enhancement is obtained (Blaster weight, W_{bB}, increased from 0 to 0.02; *see Simulation 7, Appendix 5). Note, the change of y-axis scales between these two plots.*

Summary of Predictions

[7] One could also drive the blaster from the early precision unit and make the P3 bigger when it is expected, rather than when it is salient.

In summary, three predictions that experimentalists can explore, have been identified from this section.

1) Can one observe a reduction in SSVEP-deflection amplitude once a frequently presented stimulus (particularly a target in a classic RSVP experiment) has become expected? Vanilla predictive coding suggests this could be present.

2) Does the high-amplitude P3 observed for targets in RSVP, behave in a fashion consistent with precision modulation, i.e. becoming earlier, higher amplitude and higher frequency as attentional enhancement increases? This would be consistent with precision-modulated predictive coding.

3) Does this P3 behave according to an additive enhancement, whereby, most importantly, the component does not become earlier as attentional enhancement increases? This would be inconsistent with precision-modulated predictive coding.

Further Empirical Evidence: Latency in a Contra-predictive Evoked Response Pattern

As just illustrated, in order to generate a contra-predictive Breakthrough-P3 evoked response pattern, precision needs to be used to elicit the larger amplitude for the unpredicted condition. As demonstrated in the Hierarchical PC-evoked model simulations, this should reduce the latency of the evoked response. Accordingly, we present an ERP approach that is relevant to testing this hypothesis.

Within the field of psycholinguistics, there is considerable evidence that successful comprehension of degraded speech relies on active predictions generated by the listener (e.g., Davis et al., 2005; Sohoglu et al., 2012; Wild et al., 2012). Consistent with a vanilla predictive coding account, we recently observed a predictive pattern in the magnitude of the ERPs elicited by degraded speech at approximately 200-250ms post-stimulus – i.e. more extreme values for unexpected stimuli relative to expected stimuli (Banellis et al., 2020). However, the subsequent ERP component, between approximately 250-350ms post-stimulus, exhibited a contra-predictive pattern when participants were actively attending to the speech stimuli – i.e. more extreme values for expected stimuli relative to unexpected. Furthermore, when participants were distracted from the speech stimuli, the ERPs in that same time-window maintained the earlier predictive pattern.

As in the RSVP data above, to explain this contra-predictive pattern within a prediction error framework, one must appeal to precision-modulation. As detailed above, such differential precision-modulation will also affect the latencies of the ERP components, with a shortening of component latency under high precision. Nevertheless, in Banellis et al. (2020), we found no evidence of an interaction between prediction and attention for the latency of this contra-predictive ERP component (nor any other component). Furthermore, using Bayesian equivalent analyses, the latencies of the ERP components were between approximately 2- and 4-times more likely under a model containing no interaction term – a result that is inconsistent with a role for precision-modulation in this contra-predictive ERP.

We acknowledge that the above Bayes Factors, while in the direction of the null, are relatively small. Furthermore, our original experiment was not designed to explicitly test for latency effects. However, incorporating manipulations of prediction and attention alongside Bayesian analyses in this way is one possible principled means for future targeted efforts to falsify a precision-modulated prediction error characterisation of ERP components.

Discussion

Falsifiability vs falsification

It is important to differentiate unfalsifiability from failure to falsify. A theory is unfalsifiable if there is no experiment that could be run that could come out in a way that stands against it, meaning that the theory is tautological. In contrast, a failure to falsify a theory, simply means that no experiment has been run (to date) that provides evidence against it, but this does not mean that no experiment exists that could falsify the theory. In particular, a failure to falsify does not imply unfalsifiability, and falsifiability does not imply falsification. That is, the fact that a theory is falsifiable does not mean that it will be falsified. Taking Physics as an example: there are many theories that have *not* been falsified, but this does not mean that they were tautological (i.e. unfalsifiable), it just means that attempts to falsify them failed. Take the laws of thermodynamics as examples (Cengel, Boles & Kanoğlu, 2011). If experiments had come out differently, they could have been falsified. Thus, the laws of thermodynamics have not been falsified, but they are falsifiable.

The issue with predictive coding with precision-modulation and a purely amplitude-oriented means of discriminating evoked responses is that it really does risk becoming unfalsifiable. That is, it can generate both predictive and contra-predictive patterns, raising the possibility that no empirical scientist can inform the correctness of the theory, i.e. there is no point in running experiments to test the theory.

Bayesian Approaches

The main focus of this paper is to identify properties that can *qualitatively* differentiate theories, i.e. properties that one theory can exhibit, but the other cannot (for any setting of its parameters). However, two models that can both generate the observed data can be differentiated using Bayesian techniques, by considering how likely the data is given the range of possible parameter settings of each model. This is certainly a strategy that could be employed to assess the validity of predictive coding. For example, one could pit the additive ensemble theory of contra-predictive evoked responses (i.e. higher amplitude responses to expected stimuli), against the PC-evoked model in a Bayesian model comparison.

However, while Bayesian approaches could be used to *quantify* differences between theories, if qualitative predictions are available then they are the most useful to experimentalists, providing the strictest, most incontrovertible, falsification. Accordingly, ways to *qualitatively* differentiate between theories is our main focus in this paper.

Latency, frequency-domain features and precision-modulation

The modelling in this paper can be considered scientifically positive, since it more strongly constrains the claims made by predictive coding with precision-modulation, providing a target for empirical scientists. If the latency of evoked responses is considered in addition to amplitude, an experimentalist can find evidence against predictive coding. This would occur if a standard condition having a larger amplitude than a deviant condition (i.e. a contra-predictive pattern) is not associated with a shorter latency. There may be the necessity to find evidence for the null, but Bayes can be used for that (Dienes, 2014).

In fact, since precision is a gain parameter, as demonstrated in figures 4, 5 and 6, obtaining a contra-predictive pattern from predictive coding, implies a change in form of the evoked response. These characteristics might be most easily observed in the frequency domain (see

figure 5), where the frequency of the evoked transients increases as precision/ gain is increased.

These frequency domain features are focused on the evoked response. However, even though the PC-evoked model cannot inform such features, one can also make an argument about induced-responses in the frequency domain, since also in this case, increased precision should move power to higher frequencies. This is essentially because increasing effective time-constants in an oscillator, would cause the (oscillator) circuit to be traversed more quickly, increasing its intrinsic frequency.

These kinds of changes of timing and frequency dynamics can be investigated with Dynamic Causal Modelling (Kiebel, Garrido, Moran & Friston, 2008). In particular, self-loops in DCM micro-circuit models really act as gains on neural responses. Thus, fitting DCM micro-circuit models to experimental manipulations of deviance and interrogating the strength of self-loops, offers one approach to testing the amplitude and latency claims of precision-weighted predictive coding. (Additionally, Bastos, Usrey, Adams, Mangun, Fries & Friston (2012) present a Canonical MicroCircuit (CMC) model that realises predictive coding concepts in a neurophysiologically detailed mannner, including with a realisation of precision. Changes of timing and frequency dynamics could also be explored with this CMC model.)

Confidence, Attention and the Predicted

Does precision become an overloaded concept in precision-weighted Predictive Coding? We highlighted three different flavours by which precision may enter the theory.

1) *Confidence*. This is the root definition of precision. Consider, for example, the equations of Rao and Ballard (Rao & Ballard, 1999), a highly influential, formulation of the framework. Precision terms (reciprocal of standard deviation) appear in these equations, which at the sensory level reflect the noise in the sensory input, and thus the confidence that the system has in the prediction errors it will pass up the sensory pathway. Thus, for example, high precision (i.e. low variance) will correspond to high certainty and thus, to high confidence. This link between precision, certainty and confidence has been frequently made in the literature, e.g. Clark (2015), Allen et al. (2016), Spence et al. (2016) and Boldt et al. (2017).

2) *Attentional control*: Friston has argued that attentional signals can be realised in predictive coding using precision modulation (Friston & Feldman, 2010). This position associates attentional control with modulation of gain, e.g. if attended, the gain at a particular position in space is increased.

3) *Predicted*: more speculatively, as a by-product of the association of attention with precision, is there a sense to which precision comes to be positively correlated with prediction, i.e. it increases when a signal is expected and decreases when it is not?

Considering these three flavours of precision, the first seems uncontroversial: prediction error uncertainty arising due to noise needs to be reflected in the model. However, there may be more to discuss about the other two.

Attentional Control: Firstly, while precision as attentional gain is a theoretically elegant approach, it does not fully answer the question of the mechanics by which the brain engages in top-down (and bottom-up) attentional control, and perhaps particularly the implementation of feature-based attention (Bowman et al, 2013; Ranson & Fazelpour, 2015; Ransom, Fazelpour & Mole, 2017; Ranson & Fazelpour, 2020). Indeed, many attention researchers would agree that attention can modulate the sensory pathway with gain control (see, for example, Experience-Guided Search (Mozer & Baldwin, 2007); the blaster in the eSTST model (Wyble, et al, 2009) (although, as previously discussed, STST employed an additive enhancement (Bowman & Wyble, 2007; Bowman et al, 2008)); the FeatureGate model (Cave, 1999); etc), but the further question is the overall architecture and associated "wiring" by which it does this, proposals for which have been made in a range of computationally instantiated models, e.g. RAGNAROC (Wyble et al, 2020); STST (Bowman & Wyble, 2007; Wyble et al, 2009; Bowman et al, 2008); Saliency-map model (Itti, Koch & Niebur, 1998; Itti & Koch, 2000); Experience-Guided Search (Mozer & Baldwin, 2007); and Neural Theory of Attention (Bundesen, Habekost & Kyllingsbæk, 2005).

Moreover, in and of itself, a finding that attention does act as gain control is certainly required for the precision-weighted prediction error theory to be supported, but it would not definitively verify it. This is because, as just highlighted, there are many extant theories that predict the same (i.e. attention as gain-control), without being predictive in nature. So, it would be an important demonstration for predictive coding, but it would not be

conclusive. In contrast, identification of a non-gain pattern would stand against predictive coding.

Neuromodulatory mechanisms are also candidates for adjusting precision in order to realise attentional control. For example, Friston and co-workers have proposed that dopamine might play this role (Friston et al, 2012; Friston et al, 2014); see also (Dayan & Yu, 2002; Dayan & Yu, 2005) for proposals concerning neuromodulators and uncertainty.

Secondly, is it a problem that precision represents both attention and confidence? The link between confidence and attention seems to be strong in Cueing experiments (Feldman & Friston, 2010). That is, the cue typically directs attention to a spatial location, and if a stimulus appears there, confidence as to whether a prediction error has occurred is indeed likely to be high, since the stimulus falls in the focus of (previously cued) covert attention.

However, there certainly are situations in which we can attend to low-confidence stimuli. For example, in some situations when driving, we may monitor the pavement for pedestrians crossing the road, a region that would (let's hope) be in the periphery of our vision. Thus, presumably, attention will be pushing up our effective confidence in prediction errors in a circumstance in which we would actually have low confidence (since the pavement is in our periphery). If the two concepts were conflated, would it be possible to attend and also have low-confidence about the attended stimulus?

Indeed, by requiring them to use, if you like, the same "channel", is there a sense to which the "attention as confidence" hypothesis loses valuable information by conflating the two, preventing them from being differentiable? For example, at a higher level of the sensory processing hierarchy, the system may register a very large precision-weighted prediction error, but it would not know what had caused that large amplitude – was it that there was high confidence in sensory inputs, but low attention, or was it that there was low confidence, but high attention? In fact, since the precision-weighted prediction error is a single number, it could actually also be that there was an extremely high prediction error, but low confidence and low attention.

Predicted: The third flavour of precision-modulation is perhaps the most serious with regard to unfalsifiability. In many experiments, attention is assumed to be engaged by presenting a stimulus frequently, i.e. making it expected, with cuing tasks a classic example. For example,

in the basic Posner task (Posner, Nissen & Ogden, 1978; Posner, 1980), the Valid cue is presented more frequently than the Invalid cue, see also, (Garner, Bowman & Raymond, 2021) and many others. Thus, in this context, we attend to what is *expected*, and it generates the largest evoked response[8], even if there is no direct instruction to attention (which might be argued to directly regulate precision). At the least, this position does not sit well with vanilla predictive coding, in which the largest evoked response is to the unexpected stimulus. Indeed, resolving this inconsistency may have been a motivator for the attention as precision association (Feldman & Friston, 2010).

Thus, might it be that it is not just confidence and attention that become conflated in precision-modulation, but it is also the expected. If high precision becomes high expectation, then prediction enters the model in two different ways – additively (subtraction) with top-down prediction and multiplicatively with precision, and in typical cases, these would work in opposition to each other.

This seems particularly problematic, since it implies that precision-modulated prediction errors really can be largest for unexpected stimuli in one setting and for expected stimuli in another. It is simply the amount that precision is titrated that determines whether the theory will generate predictive or contra-predictive evoked patterns.

Implications: the overloaded nature of precision should be an issue that experimentation can inform, for example, suggesting experiments that would compare the effect of manipulating sensory noise and attention on behaviour and/ or evoked responses. A number of potential experiments of this kind were highlighted in subsections “Sensory Noise vs Attention” and “Counter-intuitive Prediction” of the “Informal Predictions of Contra-predictive Pattern” section of the results. In addition to these, attention as precision-modulation of prediction error seems to suggest that low confidence-high attention cannot be distinguished from high confidence-low attention.

[8] As an illustration of this point, Hohwy 2012 says: “without attention, the better a stimulus is predicted the more attenuated its associated signal should be. Attention should reverse this attenuation because it strengthens the prediction error. However, attention depends on the predictability of the stimulus: there should be no strong expectation that an unpredicted stimulus is going to be precise. So there should be less attention-induced enhancement of the prediction error for unpredicted stimuli than for better predicted stimuli.”

As highlighted recently (Litwin & Miłkowski, 2020), precision is operationalised in many different ways across the prediction error literature, including being synonymous with attention, subjective feelings of confidence, and salience, thus allowing all predictive and contra-predictive results to be interpreted within the predictive coding framework. Consequently, the field will benefit from both clear computational models of the implications of precision-modulation, and from testable characterisations of the conditions under which precision will vary.

Is the Brain a Recognizer or a Predictor?

What though is the theory of brain function that predictive coding can be placed in opposition to? The fundamental debate is really between the Brain as a Recognizer and the Brain as a Predictor. The recognition system perspective might be considered the dominant theory of cognitivism (Haugeland, 1978; Lindsay & Norman, 2013; Mandler, 2002), which although still prominent, may, in some circles, be considered "on the back foot" because of the pervasiveness of predictive coding.

Notwithstanding the implications that we have discussed of precision-modulation, a central principle, as we have said, of (vanilla) predictive coding is that large evoked responses correspond to large prediction errors. This is in contrast to the recognition system perspective that the brain is trying to recognise the stimuli in the environment that fall onto sensory receptors, where the evoked response would reflect this recognition process. Importantly, recognition could simply be a feedforward process, i.e. whenever a stimulus is presented to the brain, the representation of the stimulus propagates forward along the sensory processing pathway in order to determine what the stimulus being viewed is. Furthermore, from this perspective, the evoked response tracks this forward propagation, and is generated *whether that stimulus is predicted or not*.

This recognition system perspective fits well with the mainstream of connectionism, neural networks and deep learning. For example, the brain's (putative) recognition system can be seen as solving a similar problem to deep learning systems that are categorising objects in images on the internet, e.g. (Ciresan, Meier, Masci, Gambardella & Schmidhuber, 2011). There are actually versions of the Brain as Recognizer perspective that go beyond pure feedforward models, with a key example being Adaptive Resonance Theory (ART)

(Carpenter & Grossberg, 2010). ART suggests that a large evoked response should be seen on a *match* (to a learnt pattern), which has some similarities to what we have called a contra-predictive pattern, while vanilla Predictive Coding suggests that a predictive evoked response pattern should be observed. Vanilla Predictive Coding resonates on mismatch, while Adaptive Resonance Theory resonates on match.

There has also been work suggesting a link between purely feedforward neural networks trained to perform recognition/ categorisation tasks and the brain's sensory processing pathways. For example, Khaligh-Razavi and Kriegeskorte (2014) provided evidence that a deep convolution neural network trained to perform recognition/ categorisation, constructs similar representations to those that can be observed in the ventral stream in the brain (although Kietzmann, Spoerer, Sörensen, Cichy, Hauk & Kriegeskorte (2019) argue that the addition of recurrent connections does improve the model fit). This work raises the possibility that the "good-old fashioned" recognition-based perspective may explain, at the least, a part of the computation performed by the visual processing pathway, or, in other words, predictive coding is not computationally the "only game in town".

From a broader theoretical perspective, the recognition versus prediction debate in many respects revisits the famous dispute in perception research between Gibson's direct perception (Gibson, 2002) and Gregory's constructivist perception (Gregory, 1970; Gregory, 1997); see also (Norman, 2002). Direct perception, as associated with ecological psychology, emphasised the need for the world to be veridically experienced, in order that it can be acted upon; thus, from this perspective, experience is not constructed, top down, it is specified bottom-up (Warren, 2021)[9]. In contrast, constructivist perception argued for top-down shaping of experience on the basis of expectation. Thus, there is a sense to which the Brain as Recognizer fits with the Gibsonian position, while the Brain as Predictor fits with the Gregorian position. Additionally, could the Gibsonian critique also be applied to the

[9] Indeed, the success of modern (purely feedforward) machine learning seems to sit well with Gibson's basic point that there is sufficient information in the world to support perception. Well, at the least, it suggests that there is sufficient information in 2d-images to classify without prior prediction, with the necessary information extracted through (bottom-up) statistical learning.

constructivist perspective that is also inherent to predictive coding, i.e. that it does not sit well with our capacity to act in the world?[10]

Furthermore, it is notable that the direct vs constructivist debate in perception ran and ran, without a definitive winner, suggesting that neither theory is in an absolute sense complete. Should we expect the same with regard to the Brain as Recognizer vs Brain as Predictor debate?

Perhaps the key point that this discussion of the Brain as Recognizer and as Predictor highlights is that from a philosophical/ theoretical perspective, there really are competing explanations of human perception. The existence of such alternatives does not sit well with the notion that one of these perspectives is unfalsifiable. In fact, there need to be a range of experimental claims that differentiate recognition from predictive theories and enable the relative contribution of the two to be assessed in empirical work.

Tractability and the Localist vs Distributed Debate

The long running debate in connectionist research concerning whether the brain uses localist or distributed codes (O'Reilly & Munakata, 2000) informs how predictive coding might be implemented; see also appendix 4 for a discussion of the implications of the choice of learning algorithm. In localist models, neurons are narrowly tuned to a unique concept (see Page (2000) for a more nuanced definition of localist representations). In contrast, with distributed representations, neurons are broadly tuned.

Predictive coding theories are often formulated within a localist neural network framework. In particular, predictions need to be directed to the relevant prediction error units, e.g. stimulus one in figure 1, and it is more difficult to do this fully generally with a distributed representation (although there are continuing developments in variational autoencoders and generative neural network models Doersch (2016)).

[10] Interestingly, Friston's more recent computational theory of brain function – Active Inference (Friston, Mattout & Kilner, 2011) – tackles this question "head on", by combining the capacity to act in the world (the Act in Active Inference, i.e. the Gibsonian perspective) with perception, indeed cognition, as inference (the Inference in Active Inference, i.e. the Helmholtz, Gregory perspective). In this sense, Active Inference may provide a key computational framework by which resolution of the constructivist vs direct perception debate can be explored.

There are strong arguments as to why localist representations would be found in the brain (Page, 2000; Bowers, 2009), however, localist representations do not scale as well as distributed representations. For example, with localist representations, a new neuron is required for every new concept being represented, while N neurons can represent many more than N concepts with distributed representations (Rolls & Treves, 1998). Additionally, modern deep learning is typically focused on distributed representations.

This raises the possibility of a trade-off between the predictor and recogniser hypotheses. Predictive coding/ Bayesian generative models enable perception as inference and are information-theoretically efficient, but the question is does the approach scale? However, the recognizer perspective, as instantiated in feed forward neural networks, including deep ones, does scale and demonstrably so. Is there an argument here for why the brain has both prediction and recognition?

This question of localist versus distributed representation and its relevance to the main predictive coding theories in cognitive neuroscience, could be explored experimentally by recording and analysing neural responses in non-human animals, e.g. Fusi, Miller & Rigotti (2016), as well as in humans, with implanted electrodes, e.g. Engel, Moll, Fried & Ojemann (2005). This work could focus on relevant laminar in brain areas where a link has been made to components of predictive coding, e.g. see Bastos, Usrey, Adams, Mangun, Fries & Friston (2012), and could explore how broadly tuned units are in those areas. As evidence that such a procedure is feasible, in different brain regions, experimentalists have found evidence for both distributed (mixed selectivity) units (Rigotti, Barak, Warden, Wang, Daw, Miller, & Fusi, 2013) and sparser more representation-invariant units (Quiroga, Reddy, Kreiman, Koch & Fried, 2005), suggestive of more localist representations.

Conscious Breakthrough

A number of the examples of contra-predictive evoked response patterns that we have identified have been conscious break-through effects, where stimuli are presented on the fringe of awareness (Bowman et al, 2013 & 2014; Bowman, Filetti, Wyble & Olivers, 2013a; Banellis, Sokoliuk, Wild, Bowman & Cruse, 2020). Could conscious break-through be a phenomenon that fits particularly badly with (certainly vanilla) predictive coding? That is, expecting that a salient stimulus will be presented, either as a result of instruction or the

contingencies of prior presentations, may be critical to enabling detection of that stimulus and evoked response generation. Importantly, detection and ensuing perception of salient stimuli is exceptionally challenging in this context, since the brain is trying to locate those stimuli from amongst a demanding background of high noise or attention-grabbing distractor onsets.

A salient stimulus being expected enables the perceptual system to set-up a template (broadly construed) to "look for" in the demanding presentation, and particularly focus on seeing matches for that template (see Meijs, Slagter, de Lange & van Gaal (2018) for behavioural evidence for this). This might be the optimal strategy in these demanding detection and identification environments. Indeed, as noted in sections "The P3 in Rapid Serial Visual Presentation (RSVP)" and "2nd-level Prediction Circuit: the Breakthrough-P3", in RSVP, the stimuli that are most unexpected are the distractor fillers, which typically occur very infrequently in the experiment. Thus, as we have discussed, from a predictive coding perspective, distractors should generate the largest prediction errors and would carry the most information. Distractors do contribute to the Steady State Visual Evoked Potential (SSVEP) (see, figure 8[C]), but do not elicit an evoked response beyond early visual processing areas of the brain, however, "expected" salient stimuli generate large (P3) responses, see figure 8[C].

Finally, our modelling in section **2nd-level Prediction Circuit: the Breakthrough-P3** can motivate experiments that look at the characteristics of how the Breakthrough-P3 changes its features (amplitude, latency and frequency) in response to manipulation of target stimulus salience and target predictability. Such experiments could specifically focus on whether precision-weighted (see figure 13[A]) or additive enhancement (see figure 13[B]) P3 patterns are obtained.

Looking forward

Consistent with the central argument of this paper, we need to know whether contra-predictive evoked response patterns, which are certainly present in the literature, involve a latency decrease and an increase in maximum frequency. This will tell us whether precision-modulation could generate the pattern.

Phase-resets: The PC-evoked model and the extant modelling in predictive coding is typically focussed on *amplitude-change* evoked responses. In such responses, a new set of neurons are driven to become active or, at least, to become more active, in response to a stimulus onset, i.e. where there would be a clear power increase associated with the evoked response. However, there is also considerable evidence that stimulus-driven transients can also arise from phase-reset patterns, e.g. Makeig et al (2002), i.e. simply because the phase of an on-going oscillation is reset by the stimulus onset, but without a power increase.

There are a number of neural models that generate phase-reset patterns in response to a stimulus onset, e.g. Parish et al (2021). However, further work is required to see how the PC-evoked model could be extended with phase-reset dynamics to obtain a classic reset data pattern for prediction errors, and thereby, to see whether phase reset dynamics can be reconciled with predictive coding.

Other Neuroscience Methods: Predictive coding comes with some quite strong claims concerning neurophysiological components of the theory, e.g. Kanai et al (2015). These could be used in intracranial recordings to inform some of the questions considered in this paper. For example, the central question of whether latencies shorten in situations in which evoked responses increase in amplitude, could be explored by recording from pyramidal cells. The latency of the response of superficial pyramidal cells, which are claimed to carry prediction errors, is of particular interest, while deep pyramidal cells should carry sustained predictions, indicating whether a stimulus is expected in the current context. More speculatively, considering the visual processing pathway, one may be able to obtain an indication of the current level of top-down precision from a structure such as the Pulvinar, with these top-down effects likely carried by neuromodulators (Kanai et al, 2015).

Additionally, elegant fMRI studies have been performed to search for predictive coding patterns in the BOLD response, as well as the possible presence of multiplicative gain, i.e. precision effects, e.g. Egner et al (2010). However, due to its low temporal resolution, it is difficult to see how fMRI could be used to identify the latency change predictions proposed in this paper.

New Deep Learning Approaches: There are now a number of deep learning approaches that endeavour to incorporate predictive coding, e.g. Choksi et al. (2021) and Han et al. (2018).

For example, Choksi et al. (2021) add predictive coding-like mechanisms to a deep convolution neural network (see appendix 1, *Further Justification of PC-Evoked model*, and Figure App 1 for more details), and provide evidence suggesting that the addition leads to a deep learning model that is more robust to noise. Very interestingly, these approaches do indeed combine Recognition and Prediction, with the former provided by the deep neural network and the latter by the augmentation with prediction mechanisms. Additionally, Heeger et al (2017) incorporates prediction mechanisms with a recognition neural network, with hyper-parameters regulating the extent to which these different functional influences dominates. These hybrid models highlight that such combined Recognizer-Predictors may be a key direction for future research in computational and cognitive neuroscience.

In the "Results" section, we highlighted a number of predictions that could assess the validity of predictive coding. This offers the possibility that one could constrain the theory from many experiments. If model fitting is used to do this, it is important not just to fit models separately to the results of each experiment, but to fit a single model across many experiments, essentially reducing "wiggle room" and constraining the parameter space for model fits.

Finally, the possibility that the brain is both a Recognizer and a Predictor needs to be embraced. This raises important theoretical questions about how these two theoretical frameworks could function together, i.e. how could a feedforward Brain as Recognizer be integrated with a generative Brain as Predictor, with emerging deep learning models providing initial steps in this direction?

Conclusion

Predictive coding is one of the most important and well attested theories in neuroscience, and there is no doubt that it is a substantial part of the story of brain processing, but the question is, is it the whole story? That is, our point is not that predictive coding is wrong, but rather, we raise the question of whether it is a *complete* explanation. However, to test this completeness, one requires properties that predictive coding does not imply. These are what we have sought to identify in this paper.

Acknowledgements

We would like to thank Karl Friston for many fruitful discussions concerning predictive coding. We would also like to thank two anonymous referees for very useful comments and suggestions, which have greatly improved this paper.

Appendices

Appendix 1: Further Justification of PC-Evoked model

Evoked Response as Prediction Error: In our first model, we are interested in the first evoked transient following the onset of a stimulus. Thus, the analogue in Rao and Ballard's model of our evoked transient is the first feed forward prediction error (which computes $I - f(U\boldsymbol{r})$); see figures App 1[B] and App 2 (annotation PErr). $f(U\boldsymbol{r})$ is the prediction and is not changing in this first response, since activation (i.e. prediction errors) need to propagate upwards first before (feedback) predictions can change. I is the sensory input, U a weight matrix and $\boldsymbol{r}$ are the causes/prediction. We see a similar configuration in Kanai et al (2015); see figure App 1[A], although the orientation of the circuits depiction has changed and a top-down precision pathway has been added.

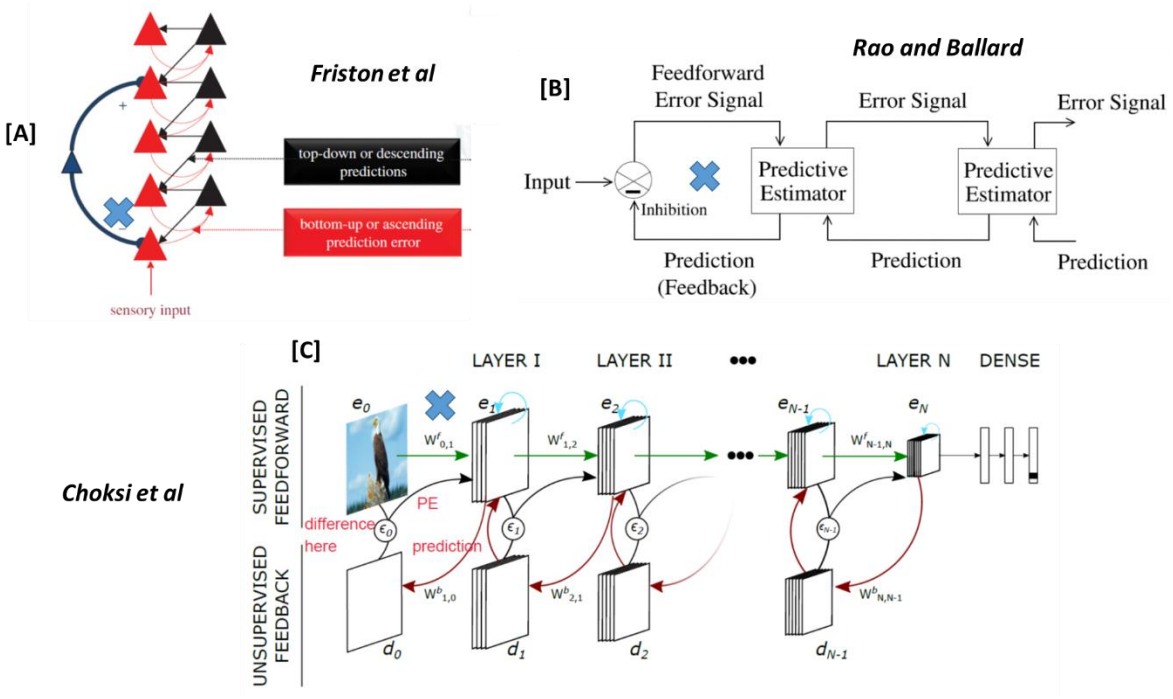

Figure App 1: Comparison of predictive coding models. [A] Friston et al predictive coding model, showing a part of figure 2 from Kanai et al (2015). [B] Rao and Ballard predictive coding model, showing a part of figure 1 from Rao and Ballard (1999). [C] Deep convolution predictive coding model, showing figure 1 from Choksi et al (2021). Importantly, [C] contains feedforward recognition links (green arrows), while [A & B] do not; see blue cross in each panel, showing where the recognition link is or would be, if it were present, for the first stage of each model. In [A & B], the only feedforward links are prediction errors.

We reproduce in figure App 1[C] a recent, deep learning model incorporating predictive coding by Choksi et al (2021); see inline heading *New Deep Learning Approaches* in the **Looking forward** section of the **Discussion**. In an excellent appendix, Choksi et al (2021) compare their equations directly to those of Rao and Ballard, and identify a difference between their model and Rao and Ballard's in terms of the feedforward sweep, stating,

"Equation 23 [of Choksi et al] also highlights the fact that our approach has an extra feedforward term that is not present in the original Rao and Ballard proposal. We believe that such a modification allows for rethinking the role of error-correction in network dynamics; where error-correction constituted the predominant mode of feed-forward communication in the Rao and Ballard implementation, it plays a more supporting role in our implementation, iteratively correcting the errors made by the feedforward convolutional layers."

This makes clear an important trend in the emerging deep learning literature on predictive coding: their models are combining feedforward recognition with feedforward prediction errors, to obtain what might be called hybrid approaches. However, as formulated, the PC-Evoked model only reflects the classic feedforward as prediction error formulation in the cognitive neuroscience literature.

More Detailed Relating to Rao and Ballard Model: Figure App 2 gives a more detailed relating of the PC-Evoked model to Rao and Ballard's predictive coding circuit. The PC-Evoked model (see figure 1, main body) provides an implementation that is conceptually related to the input end of Rao and Ballard's model. Thus, PC-Evoked's *Stim1* and *Stim2 prediction* are playing the role of Rao and Ballard's Prediction $[f(U\boldsymbol{r})]$, PC-Evoked's s_1 and

s_2 projections correspond to I here and the Evoked Response (in PC-Evoked) corresponds to PErr here.

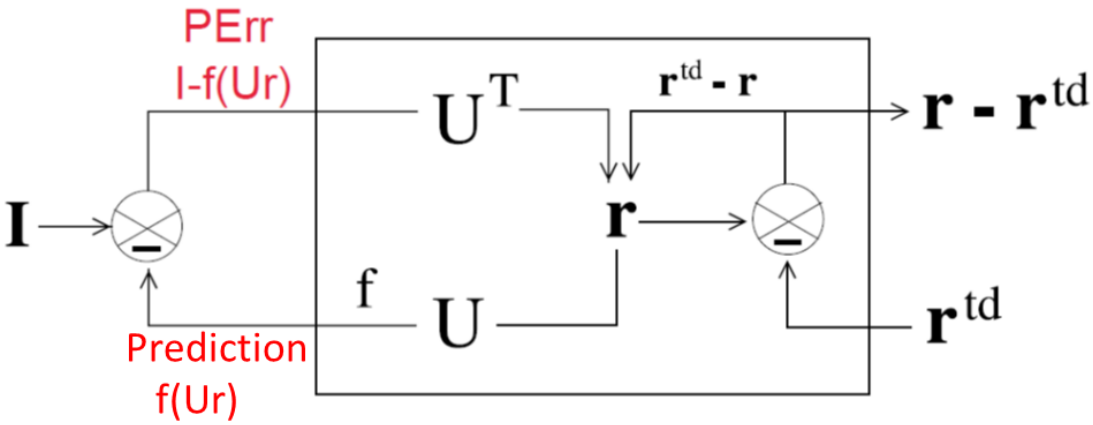

Figure App 2: Rao and Ballard predictive coding circuit, showing a part of figure 1 from Rao and Ballard (1999). The PC-Evoked model (see figure 1, main body of this paper) provides an implementation of the input end of Rao and Ballard's model. Thus, PC-Evoked's Stim1 and Stim2 predictions are playing the role of Rao and Ballard's Prediction [$f(U\boldsymbol{r})$], PC-Evoked's s_1 and s_2 projections correspond to I here and the link labelled Evoked Response (in PC-Evoked) corresponds to PErr here.

Note, the PC-Evoked model cannot be exactly the same as Rao and Ballard's, since it contains shunting dynamics and bio-physiologically more plausible activation dynamics, as required to generate evoked responses. However, we would argue that the PC-Evoked model has a broad correspondence to the Rao and Ballard model.

To understand PC-Evoked, it is important to see the analogue of the dynamics of Rao and Ballard's variable $\boldsymbol{r}$ (which corresponds to our prediction unit), which changes according to the following equation:

$$\frac{d\boldsymbol{r}}{dt} = -\frac{k_1}{2}\frac{\partial E}{\partial \boldsymbol{r}} = \frac{k_1}{\sigma^2}U^T\frac{\partial f^T}{\partial x}(\boldsymbol{I} - f(U\boldsymbol{r})) + \frac{k_1}{\sigma_{td}^2}(\boldsymbol{r}^{td} - \boldsymbol{r}) - \frac{k_1}{2}g'(\boldsymbol{r})$$

where E is the objective function, i.e. energy that is being minimised; k_1 is a constant update rate; σ^2 is the variance of the input; σ_{td}^2 is the variance of the top-level prediction; U is a weight matrix; f is an activation function; $\boldsymbol{r}^{td}$ is the top-level prediction and g is the negative logarithm of the prior over $\boldsymbol{r}$.

There are a number of simplifications that would apply to the context of the PC-Evoked model.

1. The activation function (f) is the identity (which has a derivative of 1) in the PC-Evoked model, i.e. it can be dropped.
2. Since the PC-Evoked model does not have cross-talk connections (e.g. from the Stimulus 1 circuit to the Stimulus 2 circuit) and the weights in one circuit are mirrored in the other, $U = w.\mathbb{I}$, where $\mathbb{I}$ denotes the identity matrix and w is a scalar.
3. $\boldsymbol{r}^{td}$ is assumed to be zero, suggesting that there are no higher-order expectations about the category or sequence of stimuli that are most probable. Also, it remains zero over the course of a simulation. This reflects the fact that higher-level prediction operates on a much longer time-scale. The Rao & Ballard circuit (see Figure App 2) is set-up so that with zero input (I) and zero $\boldsymbol{r}^{td}$, it will stabilse with $\boldsymbol{r}$ equal to zero. The leak in the PC-Evoked neurons will ensure the same.

Consequently, the update equation for $\boldsymbol{r}$,

$$\frac{d\boldsymbol{r}}{dt} = -\frac{k_1}{2}\frac{\partial E}{\partial \boldsymbol{r}} = \frac{k_1}{\sigma^2}U^T\frac{\partial f^T}{\partial x}(\boldsymbol{I} - f(U\boldsymbol{r})) + \frac{k_1}{\sigma_{td}^2}(\boldsymbol{r}^{td} - \boldsymbol{r}) - \frac{k_1}{2}g'(\boldsymbol{r})$$

can be simplified to,

$$\frac{1}{k_1}\frac{d\boldsymbol{r}}{dt} = \frac{1}{\sigma^2}w(\boldsymbol{I} - w.\boldsymbol{r}) - \frac{\boldsymbol{r}}{\sigma_{td}^2} - \frac{g'(\boldsymbol{r})}{2}$$

If we assume a Gaussian prior for $\boldsymbol{r}$, then since g is a negative logarithm of the prior, $g'(\boldsymbol{r}) = 2\alpha\boldsymbol{r}$, where α is a positive constant related to the variance of the Gaussian prior; see appendix of Rao and Ballard (1999). This allows us to rewrite our equation to,

$$\frac{1}{k_1}\frac{d\boldsymbol{r}}{dt} = \frac{1}{\sigma^2}w(\boldsymbol{I} - w.\boldsymbol{r}) - \frac{\boldsymbol{r}}{\sigma_{td}^2} - \frac{2\alpha\boldsymbol{r}}{2}$$

$$= \frac{1}{\sigma^2}w(\boldsymbol{I} - w.\boldsymbol{r}) - (\frac{1}{\sigma_{td}^2} + \alpha)\boldsymbol{r}$$

If we make the reasonable assumption that σ_{td}^2 is constant then we can see that the term $-(\frac{1}{\sigma_{td}^2} + \alpha)\boldsymbol{r}$ is just a decay term, which we can write as $-D\boldsymbol{r}$, with D a positive constant. Thus, we have,

$$\frac{1}{k_1}\frac{d\boldsymbol{r}}{dt} = \frac{1}{\sigma^2}w(\boldsymbol{I} - w.\boldsymbol{r}) - D\boldsymbol{r} = \frac{1}{\sigma^2}(wI - w.wr) - D\boldsymbol{r}$$

The remaining terms are as follows: $\frac{1}{\sigma^2}$ is the multiplicative precision; I the input vector, corresponding to *s₁* and *s₂*; and $\boldsymbol{r}$ corresponds to *Stim1 prediction* and *Stim2 prediction*. The decay is subsumed by the leak in the PC-Evoked equations, with D being reflected in the maximum conductance for the leak channel, G_l.

$(\boldsymbol{I} - w.\boldsymbol{r})$ here is the prediction error and it is computed in PC-Evoked's prediction error units, where $\boldsymbol{r}$ is reflected by PC-Evoked's prediction unit. $w(\boldsymbol{I} - w.\boldsymbol{r})$ corresponds to excitatory input into the prediction unit (see early circuit in figure 12). The implementation of prediction in the PC-Evoked model has a consistent form to this update equation for $\boldsymbol{r}$.

Learning: the weights in PC-Evoked are set by hand. Thus, there is no learning and Rao & Ballard's gradient descent adaptation of the weights in U is not relevant to our simulations.

Precision weighted prediction error: Although, Rao & Ballard's model did not associate precision with attention, see **Appendix 2: Precision, Gain and Attention** for that, they did have precision terms that weighted the prediction error. For example, in the following formula, $\frac{1}{\sigma^2}$ is a precision and $(\boldsymbol{I} - f(U\boldsymbol{r}))$ is a prediction error (the one PC-Evoked is focussed on), while $\frac{1}{\sigma_{td}^2}$ is a precision and $(\boldsymbol{r}^{td} - \boldsymbol{r})$ a (higher level) prediction error, in their update equation for $\boldsymbol{r}$:

$$\frac{d\boldsymbol{r}}{dt} = -\frac{k_1}{2}\frac{\partial E}{\partial \boldsymbol{r}} = \frac{k_1}{\sigma^2}U^T\frac{\partial f^T}{\partial x}(\boldsymbol{I} - f(U\boldsymbol{r})) + \frac{k_1}{\sigma_{td}^2}(\boldsymbol{r}^{td} - \boldsymbol{r}) - \frac{k_1}{2}g'(\boldsymbol{r})$$

Removal of Excitatory Reversal Term: In our efforts to obtain a formulation of the PC-Evoked model in which the evoked response is fully quenched with sustained prediction, we explored a version of the model in which the excitatory reversal term was removed from all units; see subsection Sustained Prediction of section **Simulations** of the **Results**. We justify

the statement that this reduced version of a unit is more consistent with the Rao and Ballard model here.

The relevant term in our equations is:

$$I_e(t) = g_e(t) \cdot Ge \cdot (Rev_e - V(t))$$

which is changed to:

$$I_e(t) = g_e(t) \cdot Ge$$

This change brings the model more into line with Rao & Ballard's equation:

$$\frac{1}{k_1}\frac{d\boldsymbol{r}}{dt} = \frac{1}{\sigma^2}w(\boldsymbol{I} - w.\boldsymbol{r}) - D\boldsymbol{r} = \frac{1}{\sigma^2}wI - \frac{1}{\sigma^2}w.wr - D\boldsymbol{r}$$

since the equivalent of $I_e(t)$ is $\frac{1}{\sigma^2}wI$, i.e. these are the excitatory contributions to the change in prediction, with $g_e(t)$ providing a weighting of the input in PC-Evoked.

Hierarchical Model: In section **2nd-level Prediction Circuit: the Breakthrough-P3** of the main-body of the paper, we introduce a hierarchical extension of our predictive coding model, called the *Hierarchical-PC-Evoked model*. We give more background on this extension here. Firstly, figure App 3 relates the model to Rao & Ballard's model, showing that there are structural correspondences between the two.

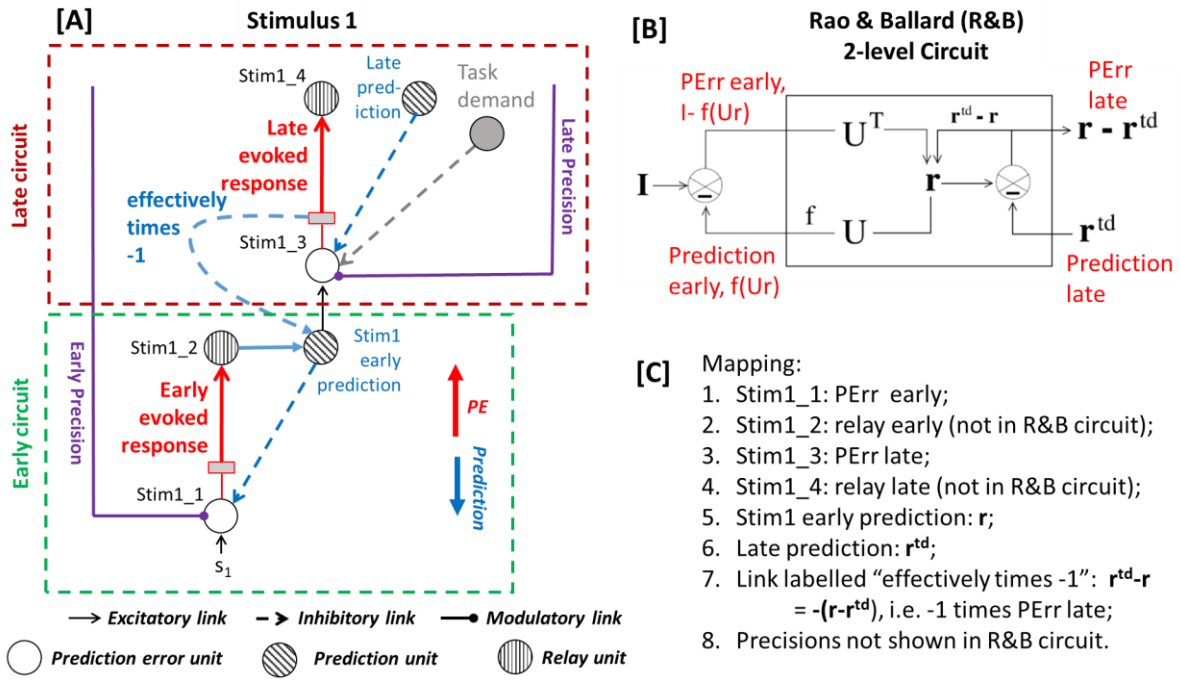

Figure App 3: Hierarchical PC-Evoked model compared to Rao and Ballard, where the Rao and Ballard model image (panel [B]) would have to be rotated to have input at the bottom to align with the Hierarchical PC-Evoked model (panel [A]). [A] Hierarchical PC-Evoked model in which a late circuit is added to the early circuit. We only depict the Stimulus 1 part of the full model. Also, we do not depict the blaster, which we add to the late circuit to simulate additive ensemble effects. [B] Rao and Ballard 2-level circuit, with annotations in red. [C] Mapping between PC-Evoked model and Rao and Ballard model.

Hierarchical model derivation

In the same way as we did in the derivation following inline heading "*More Detailed Relating to Rao and Ballard Model*" of Appendix 1, the update equation for $\boldsymbol{r}$,

$$\frac{d\boldsymbol{r}}{dt} = -\frac{k_1}{2}\frac{\partial E}{\partial \boldsymbol{r}} = \frac{k_1}{\sigma^2}U^T\frac{\partial f^T}{\partial x}(\boldsymbol{I} - f(U\boldsymbol{r})) + \frac{k_1}{\sigma_{td}^2}(\boldsymbol{r}^{td} - \boldsymbol{r}) - \frac{k_1}{2}g'(\boldsymbol{r})$$

can be simplified to,

$$\frac{1}{k_1}\frac{d\boldsymbol{r}}{dt} = \frac{1}{\sigma^2}w(\boldsymbol{I} - w.\boldsymbol{r}) + \frac{1}{\sigma_{td}^2}(\boldsymbol{r}^{td} - \boldsymbol{r}) - \frac{g'(\boldsymbol{r})}{2}$$

If we again assume a Gaussian prior for $\boldsymbol{r}$, then as previously, $g'(\boldsymbol{r}) = 2\alpha\boldsymbol{r}$. Now, we can rewrite our equation to,

$$\frac{1}{k_1}\frac{d\boldsymbol{r}}{dt} = \frac{1}{\sigma^2}w(\boldsymbol{I} - w.\boldsymbol{r}) + \frac{1}{\sigma_{td}^2}(\boldsymbol{r}^{td} - \boldsymbol{r}) - \frac{2\alpha\boldsymbol{r}}{2}$$

$$= \frac{1}{\sigma^2}w(\boldsymbol{I} - w.\boldsymbol{r}) - \frac{1}{\sigma_{td}^2}(\boldsymbol{r} - \boldsymbol{r}^{td}) - \alpha\boldsymbol{r}$$

since α is a positive constant, we can see that the term $-\alpha\boldsymbol{r}$ is a decay term, which we write, for notational consistency with earlier derivations, as $-D\boldsymbol{r}$, with D a positive constant. Thus, we have,

$$\frac{1}{k_1}\frac{d\boldsymbol{r}}{dt} = \frac{1}{\sigma^2}w(\boldsymbol{I} - w.\boldsymbol{r}) - \frac{1}{\sigma_{td}^2}(\boldsymbol{r} - \boldsymbol{r}^{td}) - D\boldsymbol{r}$$

The terms that remain in this equation can (broadly) be related to the Hierarchical PC-Evoked model (see figure App 3[A]) as follows:

- $\frac{1}{\sigma^2}$ is the early circuit multiplicative precision;
- $\frac{1}{\sigma_{td}^2}$ is the late circuit multiplicative precision;
- I is the input vector (s_1 and s_2);
- $\boldsymbol{r}$ corresponds to early prediction unit and $\boldsymbol{r}^{td}$ to late prediction unit;
- the decay is subsumed by the leak in the Hierarchical PC-Evoked equations, with D being reflected in the maximum conductance for the leak channel, Gl;
- $(\boldsymbol{I} - w.\boldsymbol{r})$ here corresponds to the early prediction error and it is computed in Hierarchical PC-Evoked's early prediction error unit;
- $w(\boldsymbol{I} - w.\boldsymbol{r})$ here corresponds to the (excitatory) bottom-up input to the early prediction unit (after passing through the relay unit in Hierarchical PC-Evoked);
- $\boldsymbol{r} - \boldsymbol{r}^{td}$ here corresponds to the late prediction error, computed in Hierarchical PC-Evoked's late prediction error unit; and
- $-(\boldsymbol{r} - \boldsymbol{r}^{td})$ here corresponds to the "effectively times -1" link in Hierarchical PC-Evoked.

Thus, we contend that broad correspondences can be made between the implementation of prediction in the Hierarchical PC-Evoked model and the update equation for $\boldsymbol{r}$ from Rao and Ballard. Note, we are definitely not arguing for a quantitative correspondence between the two models, many things prevent this, including the very different activation equations used.

Appendix 2: Precision, Gain and Attention

The linking of precision to gain and also then to attention, is justified by a large body of literature. For example, Feldman & Friston (2010) (a highly influential paper that, as of 10/5/2023, has 1310 citations) very explicitly makes this link, e.g. Feldman & Friston (2010) state that, "Inverse variance is called precision; therefore precision increases with certainty about states of the world. We will see that precision is encoded by the post-synaptic gain of sensory or prediction error-units. This means that state-dependent changes in precision may be modelled in the brain by activity-dependent modulation of the synaptic gain of principal cells originating forward connections. This is the optimization we associate with attention."

This perspective on precision and attention is also shown in Figure App 4, which is a re-presentation of Figure 2 from Kanai et al (2015). The purple lines transmit a setting of a modulatory gain, which manifests in the activation equations as a precision term. These modulatory links originate from the pulvinar, an area associated with top-down attentional control. In Kanai et al (2015), it is explicitly stated that, "The prediction errors are weighted by their expected precision— which we have associated with projections from the pulvinar."

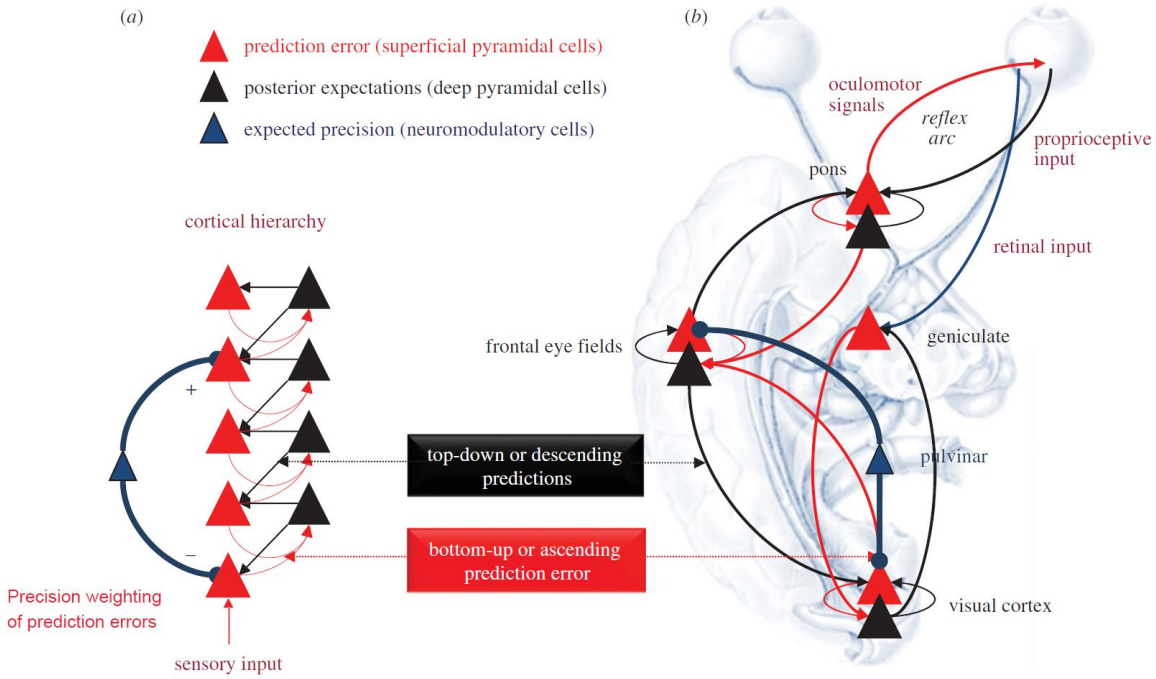

Figure App 4: Fristonian perspective on attention and precision: re-presentation of Figure 2 from Kanai et al (2015), see caption in that paper for full details. For our purposes, these images show precision as a modulator on prediction errors; see purple lines, with precision-setting neuron indicated with purple triangle. The image on the right is a more neurophysiologically detailed representation of the abstract representation on the left. The neuron setting precision is placed in the pulvinar, an area associated with top-down attentional control.

The prediction we make for an interaction between attention and sensory noise (see subsection *Shared channel saturation effect* in section Counter-intuitive Prediction of **Informal Predictions of Contra-predictive Pattern**) is formulated assuming a single shared channel by which precision weights prediction errors. It is assumed that this single channel is shared between precision's standard representation of the reciprocal of the sensory noise

level and by attentional influences. This idea of a shared channel is suggested by the theory developed in Feldman and Friston (2010). For example, Feldman and Friston (2010) explicitly state that, "Attention can be viewed as a selective sampling of sensory data that have high-precision (signal-to-noise) in relation to the model's predictions." and they also state that, " [we] consider generative models in which the states of the world (for example the presence of attentional cues) can change the precision of sensory data. A simple example of this would be the direction (state) in which we pointed a searchlight. This determines which part of the sensorium contains precise information; namely visual information reflected by surfaces that are illuminated."

Thus, for Feldman and Friston, it is the case that attentional mechanisms will reduce sensory noise (although, see Bowman et al (2013a)).

Appendix 3: Mathematical Definition of Responsiveness

Our definition of neural responsiveness (see equation following *Neural Responsiveness* inline heading of Neural Simulations section), gives us the relationship shown in Figure 2, and reproduced here.

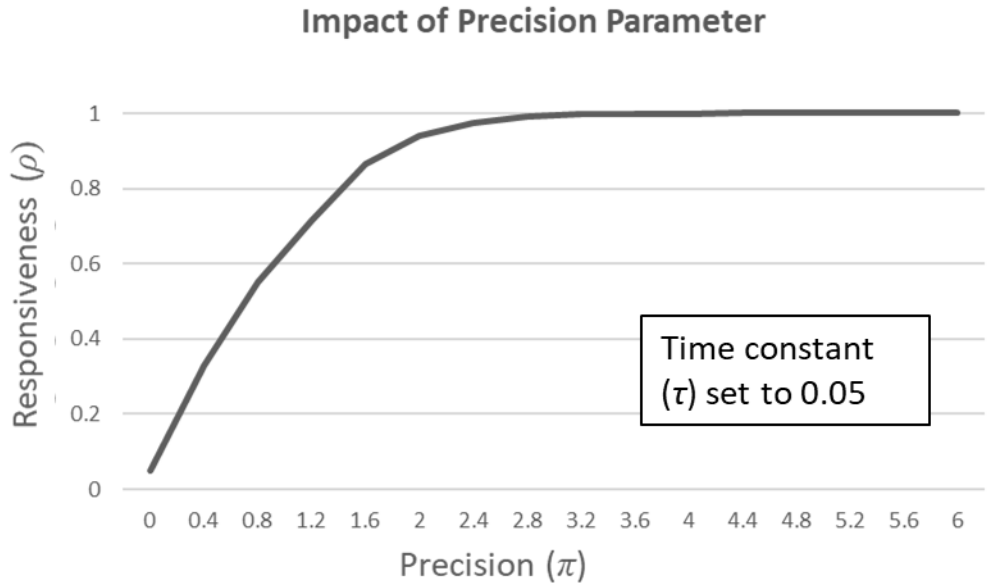

Figure 2: neural responsiveness by precision: precision (π) is shown on the x-axis and responsiveness (ρ) on the y-axis. The time constant (τ) is set to 0.05. As a result,

responsiveness is 0.05, when precision is zero. Responsiveness rises as precision increases, asymptotically approaching 1 for large precisions.

The properties that we wanted for our responsiveness variable were,

1) a fixed minimum responsiveness, so responsiveness cannot go to zero, which would have "flat-lined" the model;

2) a saturation level for precision, since there must be a maximum level for any parameter in the brain, due to fixed amounts of metabolic resource;

3) a responsiveness profile that followed a neurophysiologically plausible increase with precision; the "top" of a logistic function (which is what we see here) is exactly this - see prominence of logistic functions as a standard activation function for a neuron.

Figure 2 indicates that our implementation successfully realises these three properties.

Appendix 4: Local vs global learning

The Brain as Recognizer and the Brain as Predictor hypotheses bring with them associated learning algorithms, which inform the neurophysiological plausibility of these hypotheses. Firstly, as discussed previously, we take feedforward neural networks as the neuro-computational underpinnings of the Brain as Recognizer position. The learning algorithm typically employed in this context is back-propagation of error (O'Reilly and Munakata, 2000; Rumelhart, Hinton & Williams, 1986), in which, importantly, an error is determined at the output end of the neural network, and then propagated back through it. In this sense, back-propagation is a global learning rule – it is seeded at the output end and then passed backwards, to determine the contribution of earlier layers to that overall error. This leaves the question of how the error is transmitted backwards in the brain to neurons potentially many many synapses before the output layer.

A key contribution of predictive coding is to suggest how a hierarchical generative model enables errors to be generated at all hierarchical levels through local "message" exchange (e.g. Rao & Ballard, 1999). Thus, predictive coding provides a local learning rule, which in this respect, could much more plausibly be found in the brain, essentially because it does not require a long-range error signal, which would impact the entire configuration of the brain.

To clarify, some connectionists have, in fact, emphasized prediction-like mechanisms for some time – see for example, O'Reilly and Munakata (Implicit Expectation in Figure 5.12 in O'Reilly and Munakata, 2000) and McClelland (McClelland, 1994), who discuss how expectation can be used to avoid the need for an explicit teacher. However, these formulations were still based upon back-propagation or variants of it, such as, the Generalised Recirculation algorithm (O'Reilly and Munakata, 2000; Su, Gomez and Bowman, 2014).

Interestingly, there has been recent work identifying mappings between back-propagation and predictive coding, e.g. Song et al (2020) and Whittington & Bogacz (2017), with the objective of finding a more biologically plausible (local-learning) version of back-propagation. Could this line of research hold the key to reconciling the Brain as Recognizer with the Brain as Predictor?

Appendix 5: Details of PC-Evoked model

More mathematical details of the model are presented here. We present the full (hierarchical) model here. However, in the code, the early circuit, the simple (non-hierarchical) PC-Evoked model, can be obtained by just instantiating the early circuit of the model, with the late circuit units not present.

The model consists of two 'circuits': an early circuit and a late circuit. Subscripts indexing variables in the early circuit are in lowercase, whereas subscripts indexing variables in the late circuit are in uppercase.

Membrane potential equation for early prediction error units (denoted *Stim1_1* and *Stim2_1* in Figure 1):

$$\dot{V}_{\epsilon}^{j} = \rho_{\epsilon}(t) \cdot I_{net}(t)$$
$$I_{net}(t) = I_{e}(t) + I_{i}(t) + I_{l}(t)$$
$$I_{e}(t) = W_{s\epsilon} \cdot s_{j}(t) \cdot Ge_{\epsilon} \cdot \left(Rev_{e} - V_{\epsilon}^{j}(t)\right)$$
$$I_{i}(t) = W_{p\epsilon} \cdot V_{p}^{j}(t) \cdot Gi_{\epsilon} \cdot \left(Rev_{i} - V_{\epsilon}^{j}(t)\right)$$
$$I_{l}(t) = Gl_{\epsilon} \cdot \left(Rev_{l} - V_{\epsilon}^{j}(t)\right)$$

where the ϵ indicates an early circuit prediction error unit; $j \in \{1, \dots, k\}$ indexes the stimulus and k is the number of stimuli; $\rho_{\epsilon}(t)$ is precision weighting of early prediction

error, effectively modulating the time constant; $W_{s\epsilon}$ is the weight from the stimulus to the early prediction error units (this is the same for all stimuli); $s_j(t)$ is the stimulus input for stimulus j; $W_{p\epsilon}$ is the weight from the j^{th} early prediction unit to the early prediction error unit (this is the same for all stimuli); $V_p^j(t)$ is the membrane potential of the j^{th} early prediction unit at time t (to be defined shortly). Additionally, since our output activation functions are the identity, $V_p^j(t)$ is the output activation of the prediction unit.

Membrane potential equation for early relay units (denoted *Stim1_2* and *Stim2_2* in Figure 1):

$$\dot{V}_r^j = \tau_r \cdot I_{net}(t)$$
$$I_{net}(t) = I_e(t) + I_l(t)$$
$$I_e(t) = W_{\epsilon r} \cdot V_\epsilon^j(t - lag_r) \cdot Ge_r \cdot \left(Rev_e - V_r^j(t)\right)$$
$$I_l(t) = Gl_r \cdot \left(Rev_l - V_r^j(t)\right)$$

where r indicates early relay unit; $W_{\epsilon r}$ is the weight from the early prediction error unit to the early relay unit (the same for all stimuli); and V_ϵ^j is the membrane potential of the j^{th} early prediction error unit (as just defined); and τ_r is a time constant. For simplicity, we reference V_ϵ^j directly, without an activation equation. The time-lag between the early prediction error units and early relay units is handled by lag_r.

Membrane potential equation for early prediction units (denoted *Stim1 prediction* and *Stim2 prediction* in Figure 1):

$$\dot{V}_p^j = \tau_p \cdot I_{net}(t)$$
$$I_{net}(t) = I_e(t) + I_i(t) + I_l(t)$$
$$I_e(t) = W_{rp} \cdot V_r^j\left(t - lag_p\right) \cdot Ge_p \cdot \left(Rev_e - V_p^j(t)\right)$$
$$I_i(t) = W_{Ep} \cdot V_E^j(t) \cdot Gi_p \cdot \left(Rev_i - V_p^j(t)\right)$$
$$I_l(t) = Gl_p \cdot \left(Rev_l - V_p^j(t)\right)$$

where p indicates early prediction unit; W_{rp} is the weight from the early relay unit to the early prediction unit; V_r^j is the membrane potential of the early relay unit; W_{Ep} is the weight from the late prediction error unit to the early prediction unit (marked "effectively times -1" in Figure App 3[A]); V_E^j is the membrane potential of the late prediction error unit; and τ_p is

a time constant for the prediction unit. The time-lag between the early relay units and early prediction units is handled by lag_p.

Version 1 of late circuit (gain control): Membrane potential equation for late prediction error units (*Stim1_3* in Figure 12 and note difference between lower and upper case p's):

$$\dot{V}_E^j = \rho_E(t) \cdot I_{net}(t)$$
$$I_{net}(t) = I_e(t) + I_i(t) + I_l(t)$$
$$I_e(t) = W_{pE} \cdot V_p^j(t) \cdot Ge_E \cdot \left(Rev_e - V_E^j(t)\right)$$
$$I_i(t) = W_{PE} \cdot \left(V_P^j(t) + T\right) \cdot Gi_E \cdot \left(Rev_i - V_E^j(t)\right)$$
$$I_l(t) = Gl_E \cdot \left(Rev_l - V_E^j(t)\right)$$

where E indicates late prediction error unit; W_{pE} is the weight from the early prediction unit to the late prediction error unit; $V_p^j(t)$ is the membrane potential of the early prediction unit (defined above); W_{PE} is the weight from the late prediction unit to the late prediction error unit; and $V_P^j(t)$ is the membrane potential of the late prediction unit. T is the tonically fixed task set (see Task demand in figure 12[A] and App 3[A]), which is set to 0.13 in the late prediction error unit membrane potential equation for distractor stimuli and zero for targets.

Membrane potential equation for late relay units (*Stim1_4* in Figure 12):

$$\dot{V}_R^j = \tau_R \cdot I_{net}(t)$$
$$I_{net}(t) = I_e(t) + I_l(t)$$
$$I_e(t) = W_{ER} \cdot V_E^j(t - lag_R) \cdot Ge_R \cdot \left(Rev_e - V_R^j(t)\right)$$
$$I_l(t) = Gl_R \cdot \left(Rev_l - V_R^j(t)\right)$$

where R indicates late relay unit; W_{ER} is the weight from the late prediction error unit to the late relay unit; $V_E^j(t)$ is the membrane potential of the late prediction error unit; τ_R is a time constant for the late relay unit and the time-lag between the late prediction error and late relay units is handled by lag_R.

Membrane potential equation for late prediction units (*Late prediction* in Figure 12):

$$\dot{V}_P^j = \tau_P \cdot I_{net}(t)$$
$$I_{net}(t) = I_l(t)$$

$$I_l(t) = Gl_P \cdot \left(Rev_l - V_P^j(t)\right)$$

where P indicates late prediction unit; τ_P is a time constant for the late prediction unit.

Version 2 of late circuit (additive enhancement): For this second version of the late circuit, late relay, prediction and task demand are unchanged from version 1 (see figure 12[B]). However, late precision is removed, changing the input to the late prediction error unit and a blaster circuit is added. We outline the new equations here.

Membrane potential equation for the *blaster relay unit* (see figure 12[B]):

$$\begin{aligned}
\dot{V}_b &= \tau_b \cdot I_{net}(t) \\
I_{net}(t) &= I_e(t) + I_l(t) \\
I_e(t) &= W_{sb} \cdot s_1(t) \cdot Ge_b \cdot \left(Rev_e - V_b(t)\right) \\
I_l(t) &= Gl_b \cdot \left(Rev_l - V_b(t)\right)
\end{aligned}$$

where $s_1(t)$ is stimulus 1 input and W_{sb} is the weight from stimulus 1 to the blaster relay unit. This projection is only from stimulus 1 because it is assumed to be the only salient stimulus in the model.

Membrane potential equation for the blaster unit[11]:

$$\begin{aligned}
\dot{V}_B &= \tau_B \cdot I_{net}(t) \\
I_{net}(t) &= I_e(t) + I_l(t) \\
I_e(t) &= W_{bB} \cdot V_b(t)(t - lag_B) \cdot Ge_B \cdot \left(Rev_e - V_B(t)\right) \\
I_l(t) &= Gl_B \cdot \left(Rev_l - V_B(t)\right)
\end{aligned}$$

where $V_B(t)$ is the membrane potential of the blaster unit; W_{bB} is the weight from the blaster relay unit to the blaster unit; $V_b(t)$ is the membrane potential of the blaster relay unit (defined above); and τ_B is a time constant for the blaster unit.

As previously stated, membrane potential equations for the late prediction error unit change between Version 1 and 2 and become the following:

$$\begin{aligned}
\dot{V}_E^j &= \rho_E(t) \cdot I_{net}(t) \\
I_{net}(t) &= I_e(t) + I_i(t) + I_l(t)
\end{aligned}$$

[11] The blaster (Bowman & Wyble; 2007) is a pure temporal spotlight, generating an item non-specific enhancement when a salient stimulus is detected. Here, this is implemented as an enhancement of the target. We could add blaster projections to distractors, but since these are (in any case) strongly suppressed at the second level, due to task-demand, the blaster would not be able to drive them to a meaningful level of activation.

$$I_e(t) = W_{pE} \cdot \left(V_p^j(t) + V_B(t)\right) \cdot Ge_E \cdot \left(Rev_e - V_E^j(t)\right)$$
$$I_i(t) = W_{PE} \cdot \left(V_P^j(t) + T\right) \cdot Gi_E \cdot \left(Rev_i - V_E^j(t)\right)$$
$$I_l(t) = Gl_E \cdot \left(Rev_l - V_E^j(t)\right)$$

where the blaster membrane potential now enters as $V_B(t)$ and pairs of excitatory and of inhibitory inputs share the same weight, for simplicity. Also, the precision in $\rho_E(t)$ is set to zero.

The following table gives the parameter settings of the model.

Parameter	Value	Comment
$W_{s\epsilon}$	0.1	The weight from the stimulus to the early prediction error units, early circuit (*Stim1_1* and *Stim2_1*)
$W_{\epsilon r}$	$W_{p\epsilon}$	The weight from the prediction error units to the relay units, early circuit
W_{rp}	0.1	The weight from the relay units to the prediction units, early circuit
$W_{p\epsilon}$	14.5	The weight from the prediction units to the prediction error units, early circuit
W_{pE}	0.1	The weight from the early prediction unit to the late prediction error unit
W_{Ep}	0.1	The weight from the late prediction error unit to the early prediction unit
W_{ER}	W_{PE}	The weight from the prediction error unit to the relay unit, late circuit
W_{PE}	14.5	The weight from the prediction unit to the prediction error unit, late circuit
W_{sb}	1	The weight from the stimulus to blaster relay unit
Ge_ϵ	1	The max. excitatory conductance for prediction error units, early circuit
Gi_ϵ	1	The max. inhibitory conductance for prediction error units, early circuit
Gl_ϵ	0.9	The max. leak conductance for prediction error units, early circuit
Ge_r	$1/W_{\epsilon r}$	The max. excitatory conductance for relay units, early circuit
Gl_r	0.4	The max. leak conductance for relay units, early circuit
Ge_p	1	The max. excitatory conductance for prediction units, early circuit
Gi_p	1	The max. inhibitory conductance for prediction units, early circuit
Gl_p	0.1	The max. leak conductance for prediction units, early circuit

Ge_E	1	The max. excitatory conductance for prediction error units, late circuit
Gi_E	1	The max. inhibitory conductance for prediction error units, late circuit
Gl_E	0.9	The max. leak conductance for prediction error units, late circuit
Ge_R	$1/W_{\epsilon r}$	The max. excitatory conductance for relay units, late circuit
Gl_R	0.4	The max. leak conductance for relay units, late circuit
Ge_B	1	The max. excitatory conductance for blaster units
Gl_B	0.9	The max. leak conductance for blaster units
Gl_P	0.1	The max. leak conductance for prediction units, late circuit
Gl_b	0.9	The max. leak conductance for blaster relay unit
Ge_b	1	The max. excitatory conductance for blaster relay unit
Rev_e	1	The excitatory reversal potential
Rev_i	0	The inhibitory reversal potential
Rev_l	0	The leak reversal potential
τ_ϵ	0.05	Early prediction error unit time constant
τ_r	0.2	Early relay unit time constant
τ_p	0.04	Early prediction unit time constant
τ_E	0.01	Late prediction error time constant
τ_R	0.01	Late relay unit time constant
τ_P	0.04	Late prediction unit time constant
τ_b	0.01	Time constant for blaster relay unit
τ_B	0.01	Blaster unit time constant
π_e	0.54	Early prediction error unit precision (See Neural simulations section of **Methods**)
π_E	0.54	Late prediction error unit precision
C_{er}	-10	Early prediction error and relay unit presentation constant (See Neural simulations section of **Methods**)
C_p	-8	Early prediction unit presentation constant
C_{ERB}	10	Late prediction error, relay, and blaster unit presentation constant
C_P	8	Late prediction unit presentation constant
lag_r	$100ms$	The time-lag between prediction error units and relay units, early circuit
lag_p	$70ms$	The time-lag between relay units and prediction units, early circuit
lag_R	$150ms$	The time-lag between prediction error units and relay units, late circuit
lag_P	$70ms$	The time-lag between relay units and prediction units, late circuit
lag_B	$300ms$	The time-lag between blaster relay and blaster unit

The details of each simulation are as follows:

Simulation 1 – No gain

No changes to standard parameter settings.

Simulation 2 – Gain on

Gain turned on, i.e. with $\rho(t)$ as per Eqn Responsiveness, as the time-constant and τ_ϵ=0.05, no changes to other parameters.

Simulation 3 – Titrating the gain and additive ensemble scale modulation

Gain run for multiple precision values:

π_e – Varied from 0 to 0.54 in steps of 0.02, with τ_ϵ=0.05.

Simulation 4 – RSVP with one stimulus repeated

Number of stimuli – 45.

Length of simulation – 3000ms (This means stimuli are presented at 15Hz).

Repeated stimuli with no changes to parameters, then with the following:

τ_p – 0.005,

W_{pe} – 100.

Simulation 5 – RSVP vanishing current

Number of stimuli – 40 (changed from 45, since removal of reversal term means that there can be instability with more repetitions).

Length of simulation – 3000ms.

With and without the excitatory reversal term in the membrane potential equation:

τ_p – 0.005,

W_{pe} – 100.

Simulation 6 – RSVP with targets and distractors, and late circuit active, precision modulation

Number of stimuli – 15,

Number of streams – 2 (for SSVEP simulation) and 1 (for P3 late circuit),

Parameters changed to the following:

τ_p – 0.005,

W_{pe} – 80,

π_e – 0.

And the following parameter is modulated for P3 simulation:

π_E – from 0 to 0.54.

Simulation 7 – Blaster Simulation

W_{bB} – Varied from 0 (for non-salient stimuli) to 0.02 (for salient stimuli).

Removal of Reversal term: the early prediction unit has the following dynamics when the reversal term is removed in Simulation 5:

$$\dot{V}_p^j = \tau_p \cdot I_{net}(t)$$
$$I_{net}(t) = I_e(t) + I_i(t) + I_l(t)$$
$$I_e(t) = W_{rp} \cdot V_r^j \cdot Ge_p$$
$$I_i(t) = W_{Ep} \cdot V_E^j \cdot Gi_p \cdot (Rev_i - V_p^j(t))$$
$$I_l(t) = Gl_p \cdot (Rev_l - V_p^j(t))$$

Appendix 6: Frequency Domain Features for Additive Ensemble Effects

We present, here, the, rather straightforward, frequency domain features of the Additive Ensemble Effect, i.e. scaling the evoked response. These are presented in figure App 5.

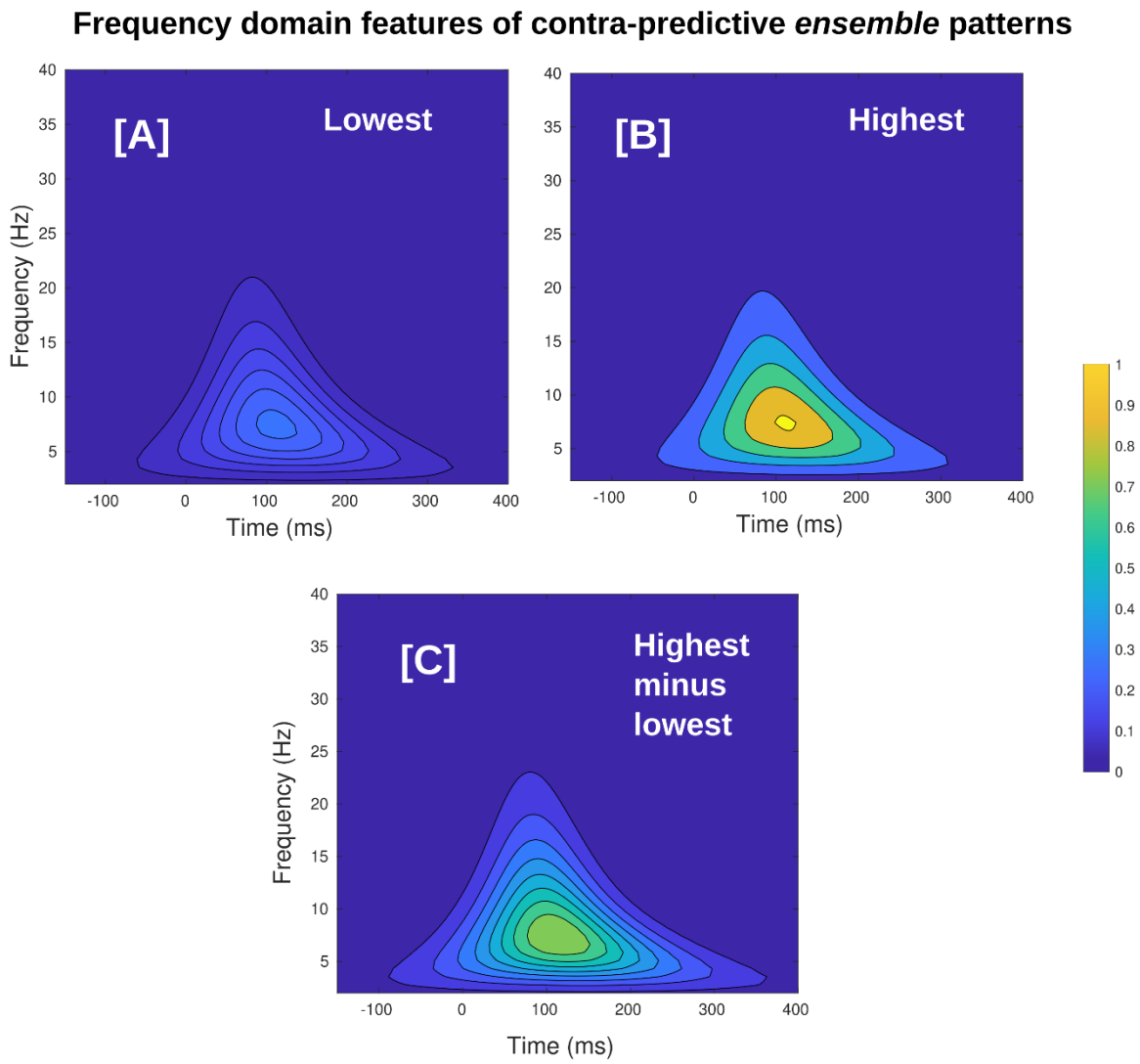

Figure ~~7~~ App 5: frequency domain features of contra-predictive pattern obtained from scale-modulation reflecting additive ensemble effects (see Simulation 3 of Appendix 5). [A] time-frequency feature obtained when I_c = 1. [B] time-frequency feature obtained when I_c = 2. [C] panel B minus panel A. Unlike Figure 5, no normalizations have been performed here, thus amplitude differences are observable.

References

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. Elife, 5, e18103.

Alsufyani, A., Hajilou, O., Zoumpoulaki, A., Filetti, M., Alsufyani, H., Solomon, C. J., ... & Bowman, H. (2019). Breakthrough percepts of famous faces. Psychophysiology, 56(1), e13279.

Alsufyani, A., Harris, K., Zoumpoulaki, A., Filetti, M., & Bowman, H. (2021). Breakthrough percepts of famous names. Cortex, 139, 267-281.

Aviles, A., Anderson, O., Orun, E., Gibson, S., Solomon, C., Via, F., & Bowman, H. (2023). Glimpse perception in RSVP can detect weak similarity. in preparation.

Avilés, A., Bowman, H., & Wyble, B. (2020). On the limits of evidence accumulation of the preconscious percept. Cognition, 195, 104080.

Banellis, L., Sokoliuk, R., Wild, C. J., Bowman, H., & Cruse, D. (2020). Event-related potentials reflect prediction errors and pop-out during comprehension of degraded speech. Neuroscience of consciousness, 2020(1), niaa022.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. Neuron, 76(4), 695-711.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. Proceedings of the National Academy of Sciences, 106(5), 1672-1677.

Boldt, A., De Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. Journal of experimental psychology: human perception and performance, 43(8), 1520.

Boring, E. G. (2008). History of experimental psychology. Genesis Publishing Pvt Ltd.

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. Psychological review, 116(1), 220.

Bowman, H., & Avilés, A. (2021). Fragile Memories for Fleeting Percepts. psyArxiv.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. Psychological review, 114(1), 38.

Bowman, H., Filetti, M., Alsufyani, A., Janssen, D., & Su, L. (2014). Countering countermeasures: detecting identity lies by detecting conscious breakthrough. PloS one, 9(3), e90595.

Bowman, H., Filetti, M., Janssen, D., Su, L., Alsufyani, A., & Wyble, B. (2013). Subliminal salience search illustrated: EEG identity and deception detection on the fringe of awareness. PLoS One, 8(1), e54258.

Bowman, H., Filetti, M., Wyble, B., & Olivers, C. (2013a). Attention is more than prediction precision [Commentary on target article]. Behavioral and Brain Sciences, 36(3), 206-208.

Bowman, H., Wyble, B., Chennu, S., & Craston, P. (2008). A reciprocal relationship between bottom-up trace strength and the attentional blink bottleneck: Relating the LC–NE and ST2 models. Brain Research, 1202, 25-42.

Brodski-Guerniero, A., Paasch, G. F., Wollstadt, P., Özdemir, I., Lizier, J. T., & Wibral, M. (2017). Information-theoretic evidence for predictive coding in the face-processing system. Journal of Neuroscience, 37(34), 8273-8283.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. Psychological review, 112(2), 291.

Carpenter, G. A., & Grossberg, S. (2010). Adaptive resonance theory.

Cave, K. R. (1999). The FeatureGate model of visual selection. Psychological research, 62(2), 182-194.

Cengel, Y. A., Boles, M. A., & Kanoğlu, M. (2011). Thermodynamics: an engineering approach (Vol. 5, p. 445). New York: McGraw-hill.

S. Chennu, P. Craston, B. Wyble & H. Bowman (2009) “Attention Increases the Temporal Precision of Conscious Perception: Verifying the Neural ST2 Model.“ PLoS Comp Biology. 5(11), Nov 2009.

Choksi, B., Mozafari, M., Biggs O'May, C., Ador, B., Alamia, A., & VanRullen, R. (2021). Predify: Augmenting deep neural networks with brain-inspired predictive coding dynamics. Advances in Neural Information Processing Systems, 34, 14069-14083.

Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011, June). Flexible, high performance convolutional neural networks for image classification. In Twenty-second international joint conference on artificial intelligence.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and brain sciences, 36(3), 181-204.

Clark A. 2015. Embodied Prediction. In Open MIND. Frankfurt am Main: MIND Group.

Coles, P. (2001). Einstein, Eddington and the 1919 eclipse. arXiv preprint astro-ph/0102462.

P. Craston, B. Wyble, S. Chennu, & H. Bowman (2009) “The attentional blink reveals serial working memory encoding: Evidence from virtual & human event-related potentials.” Journal of Cognitive Neuroscience, 21(3):550-566, March 2009.da Silva, F. L. (2004). Functional localization of brain sources using EEG and/or MEG data: volume conductor and source models. Magnetic resonance imaging, 22(10), 1533-1538.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. Journal of Experimental Psychology: General, 134(2), 222.

Dayan, P., & Yu, A. J. (2002). Expected and unexpected uncertainty: ACh and NE in the neocortex. Advances in neural information processing systems, 15.

Dayan, P., & Yu, A. J. (2005). Norepinephrine and neural interrupts. Advances in neural information processing systems, 18.

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. Proceedings of the national Academy of Sciences, 95(24), 14529-14534.

Den Ouden, Hanneke EM, Peter Kok, and Floris P. De Lange. "How prediction errors shape perception, attention, and motivation." Frontiers in psychology 3 (2012): 548.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. Frontiers in psychology, 5, 781.

Doersch, C. (2016). Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating. Behavioral and brain sciences, 11(3), 357-427.

Dosher, B. A., & Lu, Z. L. (2000). Noise exclusion in spatial attention. Psychological Science, 11(2), 139-146.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. Journal of Neuroscience, 30(49), 16601-16608.

Ellias, S. A., & Grossberg, S. (1975). Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. Biological Cybernetics, 20(2), 69-98.

Engel, A. K., Moll, C. K., Fried, I., & Ojemann, G. A. (2005). Invasive recordings from the human brain: clinical insights and beyond. Nature Reviews Neuroscience, 6(1), 35-47.

Ermentrout, G. B., Terman, D. H., Ermentrout, G. B., & Terman, D. H. (2010). The hodgkin–huxley equations. Mathematical foundations of neuroscience, 1-28.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. Frontiers in human neuroscience, 4, 215.

Fell, J., Dietl, T., Grunwald, T., Kurthen, M., Klaver, P., Trautner, P., ... & Fernández, G. (2004). Neural bases of cognitive ERPs: more than phase reset. Journal of cognitive neuroscience, 16(9), 1595-1604.

Friston, K. (2018). Does predictive coding have a future? Nature neuroscience, 21(8), 1019-1021.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. Biological cybernetics, 104(1), 137-160.

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: dopamine and decision-making. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1655), 20130481.

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., ... & Bestmann, S. (2012). Dopamine, affordance and active inference. PLoS computational biology, 8(1), e1002327.

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. Current opinion in neurobiology, 37, 66-74.

Garcia-Molina, G., & Milanowski, P. (2011). Dynamics of the alpha peak frequency during flicker stimulation. In 2011 19th European Signal Processing Conference (pp. 1549-1553). IEEE.

Garner, K. G., Bowman, H., & Raymond, J. E. (2021). Incentive value and spatial certainty combine additively to determine visual priorities. Attention, Perception, & Psychophysics, 83(1), 173-186.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. Clinical neurophysiology, 120(3), 453-463.

Gibson, J. J. (2002). A theory of direct visual perception. Vision and Mind: selected readings in the philosophy of perception, 77-90.

Gregory, R. L. (1970). The intelligent eye, McGraw-Hill.

Gregory, R. L. (1997). Knowledge in perception and illusion. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 352(1358), 1121-1127.

Grossberg, S. (2013). Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. Neural networks, 37, 1-47.

Han, K., Wen, H., Zhang, Y., Fu, D., Culurciello, E., & Liu, Z. (2018). Deep predictive coding network with local recurrent processing for object recognition. Advances in neural information processing systems, 31.

Harris, K., Miller, C., Jose, B., Beech, A., Woodhams, J., & Bowman, H. (2021). Breakthrough percepts of online identity: Detecting recognition of email addresses on the fringe of awareness. European Journal of Neuroscience, 53(3), 895-901.

Haugeland, J. (1978). The nature and plausibility of cognitivism. Behavioral and Brain Sciences, 1(2), 215-226.

Heeger, D. J. (2017). Theory of cortical function. Proceedings of the National Academy of Sciences, 114(8), 1773-1782.

Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex?. Neuroscience, 389, 54-73.

Herweg, N. A., & Bunzeck, N. (2015). Differential effects of white noise in cognitive and perceptual tasks. Frontiers in psychology, 6, 1639.

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. Frontiers in psychology, 3, 96.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. Vision research, 40(10-12), 1489-1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence, 20(11), 1254-1259.

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1668), 20140169.

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS computational biology, 10(11), e1003915.

Kiebel, S. J., Garrido, M. I., Moran, R. J., & Friston, K. J. (2008). Dynamic causal modelling for EEG and MEG. Cognitive neurodynamics, 2(2), 121.

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. Proceedings of the National Academy of Sciences, 116(43), 21854-21863.

Knill, D. C., & Richards, W. (Eds.). (1996). Perception as Bayesian inference. Cambridge University Press.

Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. Cerebral cortex, 22(9), 2197-2206.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annual review of psychology, 62, 621-647.

Lindsay, P. H., & Norman, D. A. (2013). Human information processing: An introduction to psychology. Academic press.

Litwin, P., & Miłkowski, M. (2020). Unification by fiat: arrested development of predictive processing. Cognitive Science, 44(7), e12867.

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2002). Dynamic brain sources of visual evoked responses. Science, 295(5555), 690-694.

Mandler, G. (2002). Origins of the cognitive (r) evolution. Journal of the History of the Behavioral Sciences, 38(4), 339-353.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. Journal of Experimental Psychology: Human perception and performance, 17(4), 1057.

McClelland, J. L. (1994). The interaction of nature and nurture in development: A parallel distributed processing perspective. In P. Bertelson, P. Eelen, & G. D'Ydewalle (Eds.), Current advances in psychological science: Ongoing research (pp. 57–88). Hillsdale, NJ: Erlbaum.

Meijs, E. L., Slagter, H. A., de Lange, F. P., & van Gaal, S. (2018). Dynamic Interactions between top–down expectations and conscious awareness. Journal of Neuroscience, 38(9), 2318-2327.

Min, B. K., Busch, N. A., Debener, S., Kranczioch, C., Hanslmayr, S., Engel, A. K., & Herrmann, C. S. (2007). The best of both worlds: phase-reset of human EEG alpha activity and additive power contribute to ERP generation. International Journal of Psychophysiology, 65(1), 58-68.

Moss, F., Ward, L. M., & Sannita, W. G. (2004). Stochastic resonance and sensory information processing: a tutorial and review of application. Clinical neurophysiology, 115(2), 267-281.

Mozer, M., & Baldwin, D. (2007, January). Experience-Guided Search: A Theory of Attentional Control. In NIPS (pp. 1033-1040).

Murakami, S., & Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. The Journal of physiology, 575(3), 925-936.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. Ear and hearing, 16(1), 6-18.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. Behavioral and brain sciences, 25(1), 73.

O'Reilly, R. C., & Munakata, Y. (2000). Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain. MIT press.

Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. Behavioral and Brain Sciences, 23(4), 443-467.

Parish, G., Michelmann, S., Hanslmayr, S., & Bowman, H. (2021). The Sync-Fire/deSync Model: Modelling the reactivation of dynamic memories from cortical alpha oscillations. Neuropsychologia, 107867.

Penrose, R. (2005). The road to reality: A complete guide to the laws of the universe. Random house.

Pincham, H. L., Bowman, H., & Szucs, D. (2016). The experiential blink: Mapping the cost of working memory encoding onto conscious perception in the attentional blink. *Cortex*, *81*, 35-49.

Polich, J. (1986). Attention, probability, and task demands as determinants of P300 latency from auditory stimuli. Electroencephalography and clinical neurophysiology, 63(3), 251-259.

Posner, M. I. (1980). Orienting of attention. Quarterly journal of experimental psychology, 32(1), 3-25.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. Modes of perceiving and processing information, 137(158), 2.

Potter, M. C., & Levy, E. I. "Recognition memory for a rapid sequence of pictures." Journal of experimental psychology 81.1 (1969): 10.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature, 435(7045), 1102-1107.

Ransom, M., & Fazelpour, S. (2015). Three problems for the predictive coding theory of attention. Midas Online.

Ransom, M., & Fazelpour, S. (2020). The Many Faces of Attention: Why Precision Optimization Is Not Attention. The Philosophy and Science of Predictive Processing, 119.

Ransom, M., Fazelpour, S., & Mole, C. (2017). Attention in the predictive mind. Consciousness and cognition, 47, 99-112.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience, 2(1), 79-87.

Rauss, K., & Pourtois, G. (2013). What is bottom-up and what is top-down in predictive coding?. Frontiers in psychology, 4, 276.

Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature, 497(7451), 585-590.

Rimmele, J. M., Golumbic, E. Z., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex, 68, 144-154.

Rolls, E. T. & Treves, A., (1998). Neural networks and brain function (Vol. 572). Oxford: Oxford university press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088), 533-536.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell system technical journal, 27(3), 379-423.

Shirazibeheshti, A., Cooke, J., Chennu, S., Adapa, R., Menon, D. K., Hojjatoleslami, S. A., ... & Bowman, H. (2018). Placing meta-stable states of consciousness within the predictive coding hierarchy: the deceleration of the accelerated prediction error. Consciousness and cognition, 63, 123-142.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. perception, 28(9), 1059-1074.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. Journal of Neuroscience, 32(25), 8443-8453.

Song, Y., Lukasiewicz, T., Xu, Z., & Bogacz, R. (2020). Can the brain do backpropagation?---exact implementation of backpropagation in predictive coding networks. Advances in neural information processing systems, 33, 22566-22579.

Spence, M. L., Dux, P. E., & Arnold, D. H. (2016). Computations underlying confidence in visual perception. Journal of Experimental Psychology: Human Perception and Performance, 42(5), 671.

Su, L., Gomez, R., & Bowman, H. (2014). Analysing neurobiological models using communicating automata. Formal Aspects of Computing, 26(6), 1169-1204.

Vidal-Gran, C., Sokoliuk, R., Bowman, H., & Cruse, D. (2020). Strategic and non-strategic semantic expectations hierarchically modulate neural processing. Eneuro, 7(5).

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: selection is probabilistic; responses are all-or-none samples. Journal of Experimental Psychology: General, 138(4), 546.

Warren, W. H. (2021). Information Is Where You Find It: Perception as an Ecologically Well-Posed Problem. i-Perception, 12(2), 20416695211000366.

Whittington, J. C., & Bogacz, R. (2017). An approximation of the error backpropagation algorithm in a predictive coding network with local hebbian synaptic plasticity. Neural computation, 29(5), 1229-1262.

Wierda, S. M., Taatgen, N. A., van Rijn, H., & Martens, S. (2013). Word frequency and the attentional blink: the effects of target difficulty on retrieval and consolidation processes. PLoS One, 8(9), e73415.

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. Journal of Neuroscience, 32(40), 14010-14021.

Witon, A., Shirazibehehsti, A., Cooke, J., Aviles, A., Adapa, R., Menon, D. K., ... & Bowman, H. (2020). Sedation Modulates Frontotemporal Predictive Coding Circuits and the Double Surprise Acceleration Effect. Cerebral Cortex, 30(10), 5204-5217.

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The attentional blink provides episodic distinctiveness: sparing at a cost. Journal of experimental psychology: Human perception and performance, 35(3), 787.

Wyble, B., Callahan-Flintoft, C., Chen, H., Marinov, T., Sarkar, A., & Bowman, H. (2020). Understanding visual attention with RAGNAROC: A reflexive attention gradient through neural AttRactOr competition. Psychological Review, 127(6), 1163.

Yon, D., & Frith, C. D. (2021). Precision and the Bayesian brain. Current Biology, 31(17), R1026-R1032.