# Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex?

**Micha Heilbron** [a,b*] **and Maria Chait** [c]

[a] *Département de Biologie, École Normale Supérieure, Paris 75005, France*

[b] *Université Pierre et Marie Curie P6, Paris 75005, France*

[c] *Ear Institute, University College London, London WC1X 8EE, United Kingdom*

**Abstract**—Predictive coding is possibly one of the most influential, comprehensive, and controversial theories of neural function. While proponents praise its explanatory potential, critics object that key tenets of the theory are untested or even untestable. The present article critically examines existing evidence for predictive coding in the auditory modality. Specifically, we identify five key assumptions of the theory and evaluate each in the light of animal, human and modeling studies of auditory pattern processing. For the first two assumptions – that neural responses are shaped by expectations and that these expectations are hierarchically organized – animal and human studies provide compelling evidence. The anticipatory, predictive nature of these expectations also enjoys empirical support, especially from studies on unexpected stimulus omission. However, for the existence of separate error and prediction neurons, a key assumption of the theory, evidence is lacking. More work exists on the proposed oscillatory signatures of predictive coding, and on the relation between attention and precision. However, results on these latter two assumptions are mixed or contradictory. Looking to the future, more collaboration between human and animal studies, aided by model-based analyses will be needed to test specific assumptions and implementations of predictive coding – and, as such, help determine whether this popular grand theory can fulfill its expectations.

*This article is part of a Special Issue entitled: Sensory Sequence Processing in the Brain.* © 2017 The Author(s). Published by Elsevier Ltd on behalf of IBRO. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Key words: predictive coding, auditory, MMN, SSA, bayesian brain.

## INTRODUCTION

How does the brain make sense of the world? A popular theory addressing this question is predictive coding (PC). Simply put, PC states that the brain infers what is 'out there' by constantly *predicting* what is out there, and then improving those predictions. More technically, PC proposes that the brain constructs a hierarchical, *generative* model of the world – a model capable of *generating* patterns of activity 'from the top-down' that external stimuli would elicit 'from the bottom-up'. The perceiving brain continuously tries to 'fit' such models by predicting the incoming sensory input. Bad fits signal *prediction errors* that leverage increasingly accurate estimates (recognition); and, over time, a modified model (perceptual learning).

As a biological basis for Bayesian theories of perception and cognition, PC offers compelling explanations for phenomena from psychology (Knill and Pouget, 2004) neuroanatomy (Friston, 2005) and electrophysiology (Rao and Ballard, 1999). Hailed by some as providing a 'grand unified theory of the brain' (Friston, 2010) the framework has drawn a considerable amount of attention (Hohwy, 2013; Clark, 2013, 2016). But predictive coding faces many challenges. By ascribing a central role to top-down *expectations* of bottom-up inputs, PC advocates a radical break with traditional feed-forward accounts of perception. A break, some worry, too radical since core tenets of the theory are, at best, untested (Egner and Summerfield, 2013) or, at worst, untestable (Kogo and Trengove, 2015).

Initially, PC was conceptualized in the context of visual processing (Rao and Ballard, 1999; Lee and Mumford, 2003). However, the auditory system quickly became a popular test bed, with many studies capitalizing on the auditory Mismatch Negativity (MMN;

Näätänen et al., 1978, 2007), perhaps the most well-studied neural signature of surprise or error processing. The present review critically evaluates the evidence for PC in auditory cortex. In keeping with this Special Issue, we will limit ourselves to relatively low-level auditory patterns (as opposed to e.g. speech and language; but see Arnal et al., 2011; Sohoglu et al., 2012; Gagnepain et al., 2012). There exist several recent reviews of predictive representation in audition (Winkler and Schröger, 2015; Schröger et al., 2014, 2015; Winkler and Czigler, 2012; Winkler et al., 2009). In contrast, the present analysis specifically attempts to delineate key assumptions shared by different PC models (cf. Rao and Ballard, 1999; Rao, 2005; Friston, 2005, 2010; Bastos et al., 2012; Spratling, 2008a,b, 2010; see Spratling, 2015 for review) and assess whether these assumptions are supported by empirical evidence in the auditory modality.

In the next section we will briefly recap these basic assumptions and their empirical ramifications, before evaluating them in more detail in the light of recent evidence.

## PREDICTIVE CODING IN CORTEX – FOUNDATIONS AND ASSUMPTIONS

Sensory cortex is organized hierarchically. At each processing level, neurons integrate information from multiple neurons at the level below, thus encoding increasingly abstract information over ever larger temporal and spatial scales. But cortex is reciprocally connected, so neurons also receive input from the level above (Felleman and van Essen, 1991).

Traditionally, higher levels were assumed only to *modulate* lower levels, e.g. by prioritizing the processing of certain inputs over others. But in PC, following the proposal by Mumford (1992), the abstract information at higher levels *informs* and potentially *drives* neurons at lower levels by signaling a (prior) 'best guess' of their activity. At the lower level, the difference between the predicted and actual activity elicits a *prediction error* that is propagated back to the level above, where it is used to generate a new and improved (posterior) estimate. This routine is repeated, simultaneously throughout the hierarchy, until the most likely estimate is reached and the stimulus is perceived.

In this scheme – arguably the standard version of PC (Rao and Ballard, 1999; Friston, 2005; Bastos et al., 2012) – a strict cortical asymmetry exists between *back-ward* connections (carrying predictions) and forward connections (carrying prediction errors). Since forward connections originate in superficial (II/III) pyramidal neurons, and backward connections originate in deep (V/VI) pyramidal neurons (Felleman and van Essen, 1991) this asymmetry has a straightforward anatomical consequence: prediction neurons reside in *deep layers*, and error neurons in *superficial layers* (Fig. 1).

Note that this 'standard model' is not the only implementation of PC. Other models propose different arrangements, some dispensing with the functional asymmetry between forward and backward connections, and locating prediction and error neurons differently
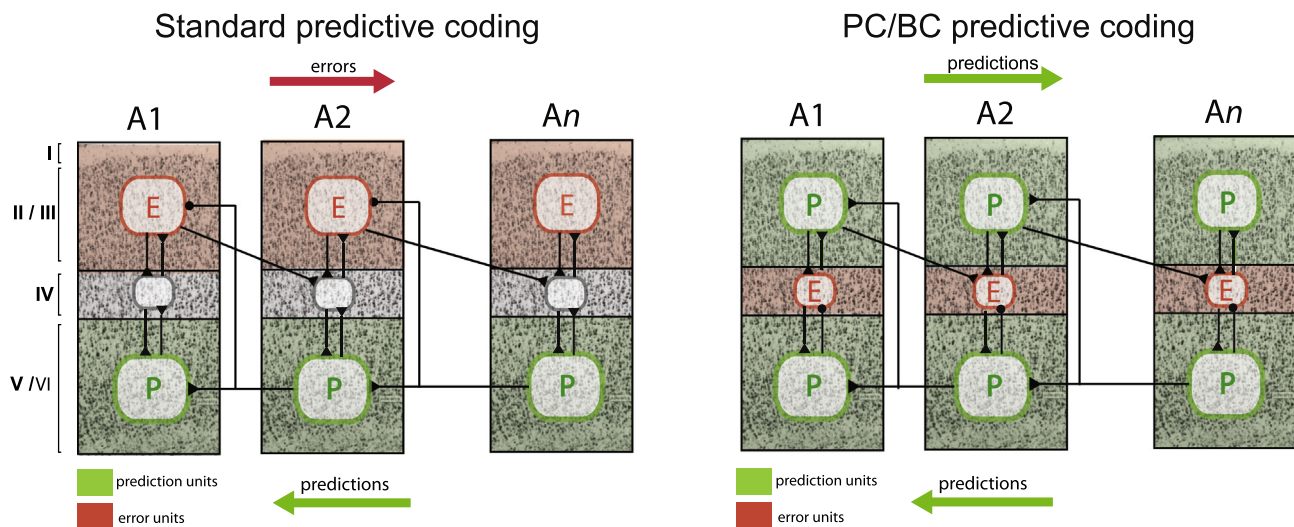


Fig. 1. Different arrangements of error and expectation neurons in auditory cortex implied by different formulations of Predictive Coding (PC). Columns denote hierarchically arranged cortical columns corresponding to primary (A1), secondary (A2) and higher order (An) auditory areas. In standard PC (left), errors flow upward and predictions downward; error units are therefore identified with superficial layers (II/III) and expectation units with deep layers (V/VI). Prediction units at higher levels can suppress error units at lower levels via (poly-synaptic) top-down inhibitory connections (black circles). In Biased Competition models of PC (Spratling, 2008a,b; right), expectations flow upward and downward, error is computed at input layer IV, prediction units suppress error units only via intracolumnar inhibition, and top-down connections are fully excitatory (black arrows). Please note that this schematic is intended to illustrate differences in laminar profiles only. For simplicity, various details have been omitted, such as the distinction between excitatory and inhibitory populations, and between hidden causes and hidden states. For a more detailed exposition of the a models, and possible physiological mappings, see Shipp (2016), Bastos et al. (2012), and Spratling (2015). Laminar image of auditory cortex was adapted from Winer (1985).

(Spratling, 2008a,b, 2010; See Fig. 1). However, all formulations assume that predictions and errors are computed by *separate neurons* in *different cortical layers* – as such, prediction and error responses are assumed to have *distinct laminar profiles*.

In PC, attention is formalized as a process that infers the level of predictability of sensory inputs. Again, models differ in details (Feldman and Friston, 2010; Rao, 2005; Spratling, 2008a,b, 2010) but all conceptualize attention as the weighting of sensory signals by their *precision* (inverse variance). The brain thus not only generates (first-order) predictions about the *content* of a signal, but also generates (second-order) estimates about its *reliability*. When this is low, deviations are *down-weighted* and may go unnoticed; when it is high, deviations are amplified and prioritized for further processing. Physiologically, this is thought to be implemented by the post-synaptic gain on superficial (error or prediction) neurons tuned to the attended dimension (e.g. feature-based or spatial attention).

Finally, different PC-variables are sometimes associated with different cortical rhythms: error cells are thought to propagate their messages forward via the gamma-band ( > 30 Hz), while deep layers send downward predictions via beta-band (12–30 Hz) or lower frequencies (Arnal and Giraud, 2012; Bastos et al., 2012). Since this assumption is based on known oscillatory differences between forward and backward signals (e.g. van Kerkoerle et al., 2014) it only applies to standard PC, which postulates a strict functional asymmetry between backward connections (carrying predictions) and forward connections (carrying errors).

In sum, PC makes a number of key assumptions with clear empirical consequences:

---

(1) **Sensory cortex implements a hierarchical, *generative* model of the world: neurons at higher processing stages *generate predictions* that *bias* processing at lower levels.**
(2) **Population responses (i.e. gross activity measured with MEG, EEG or BOLD) reflect (at least in part) 'transient expressions of prediction error' (Friston, 2005, p.829) – therefore, neural responses should be shaped by (hierarchically nested) *expectations*.**
(3) **Prediction-generation and error-detection are implemented by *separate neural subpopulations* that reside in different cortical layers – as a consequence, prediction and error computations should have *distinct laminar profiles*.**
(4) **Attention is the weighting of sensory input by its reliability – accordingly, the gain on upward projections should reflect (estimated) *sensory precision*.**
(5) **In standard PC, top-down predictions and bottom-up errors have distinct oscillatory profiles: predictions are conveyed via lower frequencies (beta and below) and (precision-weighted) prediction errors via higher frequencies (gamma).**

---

In the next sections, we will evaluate each assumption in the light of recent evidence.

## ANIMAL STUDIES

### Prediction in auditory cortical neurons

Most animal research on auditory prediction and surprise focusses on Stimulus Specific Adaptation (SSA). SSA refers to the selective attenuation of responses to repeated (common) stimuli and can be seen as a single-cell analog of MMN (Ulanovsky et al., 2003). Although their exact relation remains debated, SSA is probably not a direct substrate of MMN, since the phenomena differ in latencies, NMDA-dependence, and sensitivity to certain regularities (Khouri and Nelken, 2015). There is a large literature on SSA, most of which is beyond the scope of this review as it does not address key features of PC such as *prediction* (but see Khouri and Nelken, 2015 for review). Interestingly, it is unclear whether SSA, despite what the name implies, is caused by simple adaptation. Ulanovsky et al. (2004) showed that SSA – here defined as the difference in responses to the same sound presented with different probabilities – depended not just on local context but also on a longer stimulus history, beyond the order of seconds at which habituation processes like synaptic depression are thought to occur. Moreover, SSA is observable for tones with frequency differences smaller than typical tuning curves, which also cannot be explained by models of synaptic habituation (Taaseh et al., 2011; Yaron et al., 2012).

Recently, Rubin et al. (2016) re-analyzed the data from Ulanovsky et al. (2004), in a first attempt to quantify the longer-term dependencies. Anesthetized cats were exposed to 'Bernouli sequences' with two tones occurring independently with a fixed probability. The authors reasoned that some representation of (long-term) stimulus history influenced responses; moreover, this representation was not a one-to-one copy but a *reduced representation*. Assuming that only *stimulus probability* was represented, rather than *transitional probability* (but see Meyniel et al., 2016; Mittag et al., 2016) the authors computed the predictive power of representations reduced to a different degree. The key assumption here was that responses reflected *prediction error*, expressed as negative log probability. The prediction error account offered good fits, explaining up to 50% of observed variability. Interestingly, representations incorporating less than 10 preceding stimuli (7.3 s) were almost never in the top 10% with the most power. The authors concluded that neurons in A1 signal *prediction errors,* based on reduced representations incorporating long-term stimulus history *'to generate predictions about the future'* (2016, p.2). Although the authors are agnostic about the underlying mechanism – which may or may not resemble schemes envisioned by PC – the interpretation forms a departure from earlier accounts of SSA, which (as the name suggests) tend to focus on stimulus-driven explanations such as synaptic depression.

More fundamental insights are presented by Gill et al. (2008) who explored surprise as a model for auditory receptive fields. At several levels in the Zebra Finch auditory hierarchy, the authors compared three receptive field models: first, a traditional approach modeling neurons as responding to specific spectrotemporal patterns of intensi-

ties (STRF); secondly, a derivative approach, modeling *changes* in intensities; finally, a model describing neurons as responding to *surprise,* quantified as the inverse conditional probability of a range of frequencies, given the preceding frequencies, based on naturalistic Zebra Finch song. This 'surprise model' substantially outperformed traditional models. Interestingly, its advantage depended on hierarchical level: in area MLD (homolog of inferior colliculus) models did not differ significantly. In field L (homolog of thalamorecipient neurons in A1) surprise was 20% better than STRF models on average. And in CLM (homolog of higher-order auditory cortex) the surprise model performed a striking 67% better on average. The authors concluded that expectations are increasingly important at higher levels, leading to an ever sparser neural code that eventually only propagates input not expected on the basis of preceding input. The authors stress that these 'expectations' were based on very short preceding time windows (3–7 ms), and that the effect is thus not a direct substrate of high-level (conscious) 'surprise'. The paper is also silent on how the expectations are implemented. But it does show that, at a fundamental level, '*expectations . . . form a key part of the* [auditory] *neural code'* (2008 p., 218).

### The laminar profile of prediction

Laminar differences between prediction and error signals (*Assumption 3*) are a key tenet of predictive coding theories. This important notion remains woefully understudied; we have only found two papers relevant to this assumption. The first is by Szymanski et al. (2009), who studied the laminar profile of SSA in rat auditory cortex using an oddball paradigm. To the best of our knowledge, this is the first auditory study comparing laminar profiles for expected and unexpected stimuli. Remarkably, their results seemed opposite to what PC predicts: the authors found no clear differences between layers. In all layers, neurons responded more strongly to deviants than to the same stimulus when it was a standard (SSA). If anything, this difference seemed to be *stronger* for deeper layers. If SSA is an expression of prediction error, this effect seems to contradict PC schemes that identify deep layers as 'prediction layers'. It is worth noting, however, that the rats where anesthetized with ketamine, an NMDA-antagonist that, probably by blocking NMDA-dependent plasticity, impairs MMN (Umbricht et al., 2002) and abolishes global mismatch responses (Uhrig et al., 2016) leading to a pattern of effects that has been interpreted (Strauss et al., 2015) as a 'disruption of predictive coding' (see Section Expectation and surprise along the auditory hierarchy).

A recent study in awake animals did find strong laminar-specific effects. The authors (Rummell et al., 2016) trained mice to press a lever that generated noise bursts. Every 5–10 s the bursts were also generated randomly, allowing the researchers to compare responses to the same sound when it was self-generated and when it occurred randomly (Fig. 2A). Clear attenuation of responses to self-generated sounds was observed in primary auditory cortex, replicating non-invasive work in humans (Baess et al., 2009). The predictive, stimulus-specific nature of this effect was confirmed when the

authors replaced the noise bursts by pure tones that had one frequency in 75% of trials and another in 25%. Responses to common ('expected') self-generated sounds were attenuated more strongly – suggesting that auditory cortex embodied a stimulus-specific expectation of the generated sound. Multi-site recordings from auditory thalamus, auditory cortex and hippocampus revealed increasing attenuation, often resulting in near-silencing in hippocampus (see Fig. 3). This again points to a familiar picture of the auditory hierarchy in which predictable components are progressively 'filtered out'.

Intriguingly however, the authors found that attenuation was much *stronger* in deep layers. This again seems to contradict the idea that deep layers encode predictions, since it should be the *error* that is suppressed. Conclusions should be drawn with care, however, since predicting the sensory consequences of motor commands may be very different from sensory prediction in general, so that evidence for the one (see Eliades and Wang, 2008; Keller and Hahnloser, 2009 for earlier evidence for auditory efference copy) is not necessarily evidence for the other.

Finally, Jaramillo and Zador (2011) studied expectation in rat auditory cortex. Rats were presented a train of short pure tones containing a frequency-modulated target. The target, which appeared either 'early' (450 ms) or 'late' (1500 ms), signaled if the correct (rewarded) response was right or left. Expectation was manipulated over blocks in which the target appeared early in 85% of trials and late in 15%, or *vice versa*. Behaviorally, rats responded faster and more accurately to targets appearing at the expected time-window. Expectation also modulated single-unit and LFP responses, and this modulation correlated with performance. For both the preceding stimulus and the target itself, expectation *increased* rather than attenuated the neural response, which apparently contradicts the notion that neurons signal surprise. However, the stimulus of which expectancy was manipulated, was also the *target*. As such, *prediction* (what is likely?) and *attention* (what is relevant?) are confounded. This confound characterizes many common paradigms, including the classic Posner task (Posner, 1980), where attention is controlled by manipulating probability (Fig. 2B). In such situations, PC makes similar predictions as conventional accounts of attention: enhanced gain on the relevant (informative) feature, which is prioritized for processing. Hence, to distinguish assumptions of PC, attention and prediction must be manipulated independently (see Section Attention as precision).

### Discussion

In sum, animal-model studies relevant to the assumptions of predictive coding are scarce and show mixed results. None of the discussed studies explicitly tested PC, which may contribute to the inconclusiveness of the results. Nevertheless, they report some remarkable findings. Firstly, in support of *Assumption 2*, expectation appears to shape neural responses in auditory cortex. Surprise – both task-based at timescales of several seconds (Rubin et al., 2016), and species-based at timescales of milliseconds (Gill et al., 2008) – offers a good model for neural
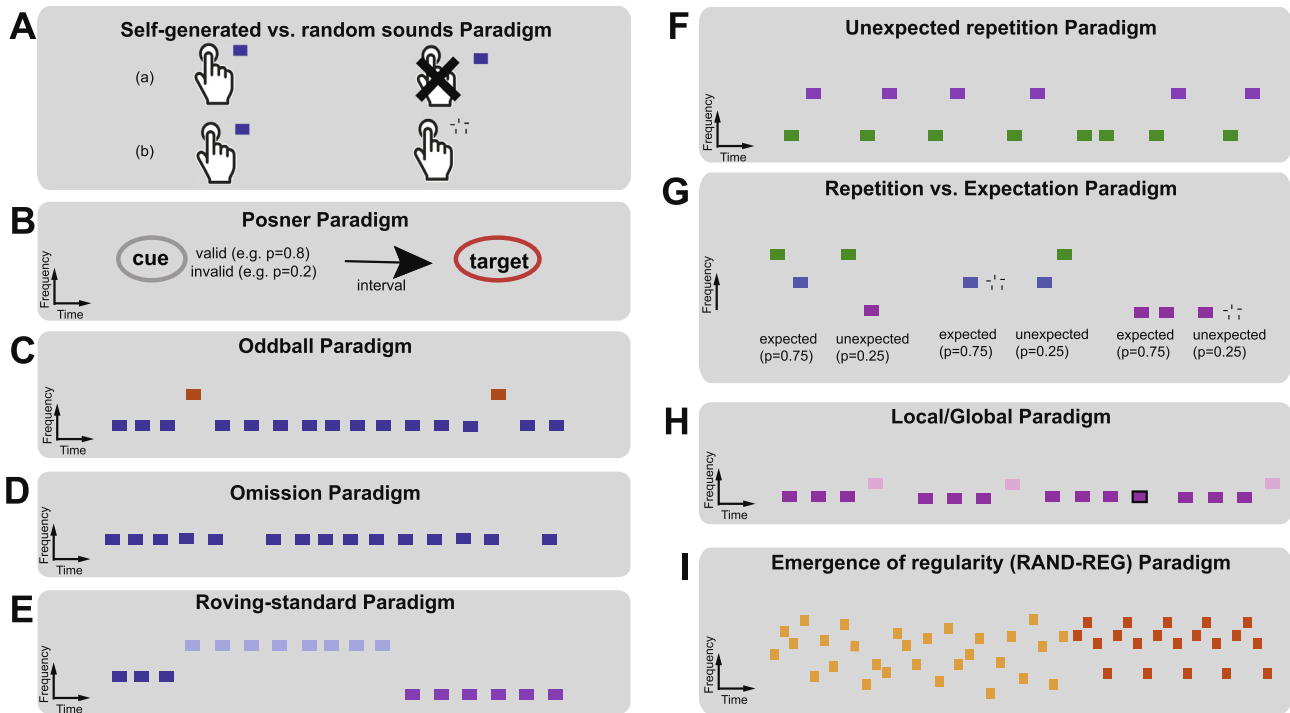
**Fig. 2.** Paradigms often used in the literature to study the effects of context and predictability on behavioral and brain responses. In the schematic representations colored squares represent sounds. (A) The 'Self-Generated vs. Random sounds' paradigm compares responses to sounds when they are self-generated (triggered by a button press; and therefore predictable) or randomly generated by a computer (and therefore unpredictable). Another version of the paradigm (b) compares self-generated sounds (triggered by a button press) to omissions (when the participant pressed the button but no sound was presented). (B) The 'Posner paradigm' is a class of experimental designs where a 'cue', which can be implemented as specific stimulus or a context which is induced during the experimental session, predicts the target with a certain probability. The paradigm therefore allows to measure responses to the target as a function of its predictability. (C) The standard MMN Oddball paradigm involves the presentation of a repeating standard tone, occasionally replaced by a deviant tone. (D) The standard MMN omission paradigm is similar to the Oddball paradigm except the deviant tone is replaced by silence. (E) The Roving standard paradigm is a variation of the oddball paradigm that replaces the deviant stimulus with a variable standard. After a number of repetitions, the standard changes, creating a 'deviant' that *becomes* a 'standard' – while remaining physically identical. (F) The unexpected repetition paradigm consists of pairs of sounds that are infrequently replaced by a repetition. The schematic here shows a simple version of the paradigm where the tone pairs consist of the same sounds, but instances where different pairs are presented are also used. (G) The 'Repetition vs. Expectation' paradigm is used to dissociate the effects of prediction from simple effects of repetition. The paradigm depicted here was used in Todorovic and de Lange (2012). The stimulus set consisted of 3 different tones (illustrated here by the use of different colors) arranged in pairs but such that the first tone in a pair was predictive of the second one. For example tone1 (green) was predictive of tone2 (blue) in 75% of the trials but was occasionally (in 25% of the trials) followed by tone2 (purple). Tone2 (blue) was predictive of an omission but which was replaced in 25% of the trials by tone 3 (green), etc. (H) The Local/Global paradigm is designed to dissociate responses to local deviants from responses to global deviants. In the example depicted here the stimulus consists of 'standard' (commonly occurring) and 'oddball' (rarely occurring) sequences. The last tone in each 'standard' sequence is a local deviant; In contrast, 'global deviance' is manifested here by the *absence* of change. A similar approach with expected and unexpected tone omissions is also commonly used. (I) The 'Emergence of regularity' (RAND-to-REG) paradigm introduced by Barascud et al. (2016) is based on rapid tone-pip sequences which contain transitions from a random (RAND) frequency pattern (in yellow) to a regularly repeating (REG; predictable) frequency pattern (in orange). In this example the REG pattern consists of a cycled sequence of 4 different tones.

responsiveness. The methodological differences between these studies, and the fact that both did not address the mechanisms of prediction, unfortunately limit their conclusiveness with respect to PC. However, both studies make the conceptual shift from characterizing neurons as encoding bottom-up data features, to encoding hypotheses or predictions, and propagating only the divergence from these predictions. In support of *Assumption 1*, there is also evidence for the idea that the effects of expectations are hierarchical, in the sense that expected components seem to be progressively filtered out (Rummell et al., 2016; Gill et al., 2008). Attention, as shown by Jaramillo and Zador (2011), can influence processing in A1 in an anticipatory way – however, it remains unclear whether this form of modulation is in line with attentional modulation as described by PC. Finally, the two studies that investigated laminar differences in processing of

expected versus unexpected stimuli – a signature characteristic of PC – found (under ketamine anesthesia) no distinct laminar profiles and (using self-generated sounds) strong expectation suppression in the deep rather than superficial layers of cortex. Although methodological issues prevent strong conclusions from being drawn, the animal-model literature contains fascinating results that call for more experiments in awake animals, since only studies of this type can ultimately confirm or falsify key assumptions of predictive coding.

## HUMAN IMAGING AND ELECTROPHYSIOLOGY

### Predictive coding and MMN

Human auditory studies on predictive coding often use some variation of the Mismatch Negativity or 'MMN'

paradigm. MMN is measured using a method in which a sequence of stimuli (typically a repeated tone) establishes a regularity that is violated by a 'deviant' stimulus ('oddball paradigm'; Fig. 2C). MMN is the negative component of a *difference wave,* obtained by subtracting the 'standard' from the 'deviant' response, and is found at 100–250 ms.

Traditionally, two main hypotheses on MMN exist. According to the memory-based hypothesis (Näätänen et al., 1978; Winkler and Czigler, 1998), MMN is generated by a system comparing auditory inputs with a memory template. When a difference is detected, the system signals an error, and adjusts the template. According to the adaptation hypothesis (May et al., 1999; Jääskeläinen et al., 2004; May and Tiitinen, 2010) cells tuned to repeated 'standard' tones simply adapt – due to passive processes such as synaptic depression – while neighboring cells tuned to 'deviants' remain unadapted and elicit stronger responses. By implication, the oddball-evoked MMN is not a separate evoked potential but rather a delayed and attenuated N1, that *appears* separate only in the difference wave.

In this dispute, PC takes a middle ground position. Like all memory-based accounts, PC interprets MMN as a mismatch signal – a mismatch, however, between the input and a *prospective* prediction, rather than a *retrospective* template. But like the adaptation hypothesis, PC considers MMN not as a separate evoked response, but simply as an amplified contrast between an expected (standard) response and a surprising (novel) response. In the adaptation hypothesis, however, both the response and its suppression are stimulus-driven: there is no error signal. Under PC, *every* response is an expression of error and can be larger or smaller depending on predictions. This last point – the dependence on predictions – is also what makes PC considerably less parsimonious than the adaptation hypothesis. Applying Ockham's razor, we can only consider evidence in favor of PC if it cannot be explained by a simpler process – which, in audition, is often simple adaptation. Beyond the standard oddball paradigm, the MMN literature has shown that listeners are sensitive to the violation of potentially very complex patterns (see Paavilainen et al., 2007, 2013 for review), which is usually interpreted as evidence for the exquisite sensitivity of auditory cortex to patterns in sound. Unfortunately, most studies with an explicit focus on predictive coding in the auditory modality (see reviewed below) have used the standard oddball paradigm or its variations in which predictability, or regularity, is manipulated by repetition, making adaptation all the more difficult to exclude.

**Repetition suppression – Adaptation or expectation?**

The neuroimaging analog of the physiological phenomenon of adaptation is repetition suppression (RS). As reviewed by Grill-Spector et al. (2006), multiple mechanisms for RS have been proposed. We can distinguish between mechanisms that explain RS via passive adaptation effects, sometimes called 'neural fatigue',

and accounts that interpret it as a signature of increased processing *efficiency*. PC belongs to the second type: it ascribes the suppression not only to the repetition itself, but also to the expectations it induces. Interesting support for this account comes from Costa-Faidella et al. (2011) who recorded EEG responses in a roving standard paradigm (Fig. 2E). This is a variation of the oddball paradigm that replaces the deviant stimulus with a variable standard. After a number of repetitions, the standard changes, creating a 'deviant' that *becomes* a 'standard' – while remaining physically identical. The authors used two conditions, with predictable and unpredictable timing. In the predictable condition, Inter Stimulus Intervals (ISI) were fixed. In the unpredictable condition, ISIs varied randomly. The suppressive effect of repetition was reduced in the condition with unpredictable timing. Because the average ISI and number of stimulations were identical between conditions, this suggests that repetition suppression is modulated by predictability.
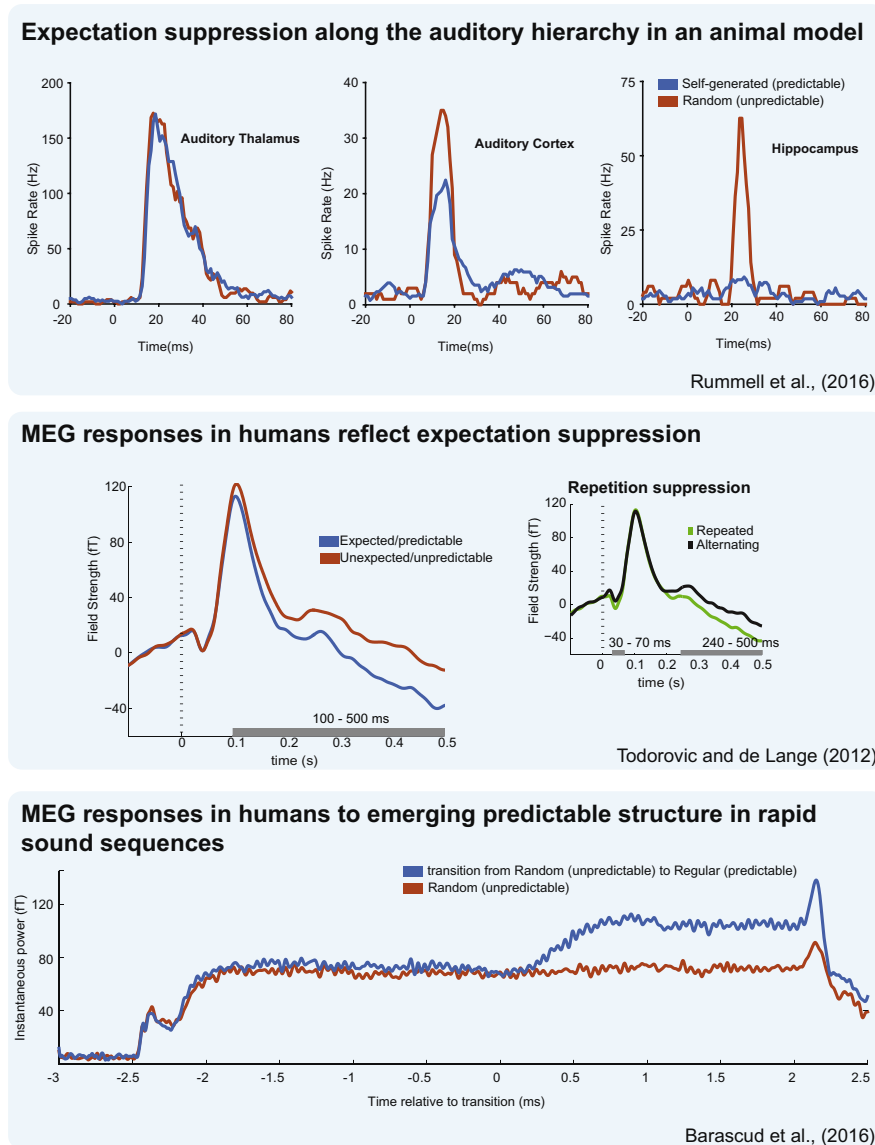
Also in a roving paradigm, Lieder et al. (2013) used computational modeling to compare prediction and adaptation. For each stimulus presentation they calculated the 'MMN amplitude', by subtracting the final ('standard') presentation from the earlier ('deviant') presentations. The authors then compared different models to explain trial-by-trial fluctuations in this MMN amplitude. The first model was an adaptation model. This model was 'phenomenological' in the sense that it made no assumptions on the *mechanism* behind adaptation, but simply embodied fluctuating responsiveness of populations tuned to different frequencies. This 'phenomenological' approach was contrasted to a computational approach in which MMN amplitudes were regressed on several parameters from a hidden Markov model which tracked transition probabilities by means of prediction error minimization. Overall, prediction error and model update, as calculated by the Markov model, explained the fluctuations better than adaptation. Together, the authors write, this suggests that attenuation observed in a roving paradigm is best explained as a form of *learning*, rather than as adaptation.

More modeling results are found in Wacongne et al. (2012) who present a PC model of A1. Contrary to Lieder et al. (2013) and the DCM studies (see Section Effective connectivity – clues from DCM) Wacongne et al. (2012) specified their model at the level of individual spiking neurons, thus committing to a much more detailed implementation of PC. The model comprised two cortical columns, each selectively responsive to a different tone (A or B). Crucially (and unlike standard PC) error units are located in the thalamorecipient granular layer. In that same error layer, GABA-ergic neurons receive excitatory input from predictive units in layer II/III, effectively subtracting the prediction from the incoming input, resulting in an error term. This error term is sent to the predictive layers, where it forms a memory trace used to adapt the internal model via spike-timing dependent plasticity at NMDA-weights. Using the sum of postsynaptic currents in each layer as a proxy for the ERP, Wacongne et al. (2012) show that this set-up – intentionally lacking synaptic habituation mechanisms – can account for an array of phenomena from the MMN literature, such as the para-

metric modulation of MMN amplitude by stimulus probability (e.g. Sams et al., 1983); MMN to unexpected *repetition* (Fig. 2F; e.g. Saarinen et al., 1992; Horváth and Winkler, 2004); MMN to omission (Fig. 2D; e.g. Yabe et al., 1997; Raij et al., 1997); and blindness to context (Wacongne et al., 2011, see below).

### Expectation suppression along the auditory hierarchy in an animal model



Rummell et al., (2016)

### MEG responses in humans reflect expectation suppression



Todorovic and de Lange (2012)

### MEG responses in humans to emerging predictable structure in rapid sound sequences



Barascud et al., (2016)

**Fig. 3.** Brain responses to predictable and unpredictable sounds. Top: Progressive attenuation of responses to self-generated sounds at different cortical regions of a mouse model (Rummell et al., 2016). The progressive weakening often resulted in near-silencing in hippocampus, and suggests the existence of an increasingly sparse code, in which eventually only non-predicted components are propagated for further processing. Middle: Repetition suppression and expectation dissociated in time (Todorovic and de Lange, 2012): Repetition (but not expectation) attenuated the early MEG responses (40–60 ms). Expectation (but not repetition) attenuated the response at an intermediate latency (100–200 ms). Both repetition and expectation affected the late response (200–500 ms). Gray horizontal bars under the figure indicate the time intervals with a significant difference between conditions. Bottom: In contrast to the oft reported attenuation of brain responses to predictable sounds, Barascud et al. (2016) found the opposite effect: Brain responses to rapid tone sequences that transitioned from a random to repeating pattern are manifest as a substantial increase in the MEG response. This finding demonstrates that the brain appears to encode the state (RAND vs REG) rather than the transition (as in e.g. MMN). As discussed in Barascud et al. (2016) the DC shift appears to vary consistently with the predictability (negentropy) of the ongoing stimulus pattern.

To further dissociate adaptation and PC based accounts, Wacongne et al. (2012) performed a MEG experiment on the 'repetition MMN'. Participants listened to tone-pairs that were overwhelmingly alternations (AB), and only rarely repetitions (AA; Fig. 2F). PC predicts that an unexpected repetition (AA) should evoke a stronger response – an *inversion* of the standard MMN. The adaptation-based explanation (May and Tiitinen, 2010) suggests this reflects adaptation at higher order neurons, tuned to the AB tone-*pair*. To exclude this possibility, Wacongne et al. (2012) inserted an interval of 10 s between each pair – much longer than the recovery time of synaptic depression. In every individual participant, AA indeed elicited an MMN, while no difference between BA and BB was observed. Although this result seems highly suggestive, a replication with a larger number of participants is needed, since Wacongne and colleagues tested only 5.

Using a similar paradigm, Todorovic et al. (2011) measured RS for expected and unexpected repetitions. Expectancy was manipulated in blocks where either 75% of stimuli were tone-pairs and, 25% single tones (repetition expected) or *vice versa* (repetition unexpected). Clear RS was observed in the 100–500 ms range, that was strongly reduced in the unexpected condition, suggesting that RS itself might comprise an expectancy effect. However, since the blockwise manipulation affected the overall occurrence of the tones – and the authors used an inter-trial interval of 4–6 s – the effect could, theoretically at least, be explained by passive adaptation.

### Expectation and surprise along the auditory hierarchy

In a follow-up study, Todorovic and de Lange (2012) addressed this issue by adding an extra hierarchical level of expectations, that allowed them to manipulate repetition and expectation orthogonally (Fig. 2G). Each trial consisted of either an identical or non-identical tone-pair, or a single tone. Orthogonally to this, the frequency of the first tone predicted that of the second tone with a high validity. Using MEG, the authors observed a dissociation: repetition (but not expectation) attenuated the early response (40–60 ms) and expectation (but not repe-

tition) attenuated the intermediate response (100–200 ms). By contrast, both repetition and expectation affected the late response (200–500 ms; Fig. 3). This suggests that RS might be non-unitary, consisting of a late stage, which reflects the effects of expectation, and an early stage, which does not (cf. Grotheer and Kovács, 2015). However, the results may also be compatible with PC, if one casts repetition as a special, low-level form of expectation (cf. Auksztulewicz and Friston, 2015b).

Similar findings were reported by Wacongne et al. (2011), who recorded MEG and EEG responses to violations of local and global regularities (Fig. 2H). Participants passively listened to stimuli consisting of five tones, of which the first four were always identical and the last one varied. Within each block, a particular variation (e.g. 'xxxxY') was dominant, occurring 75% of trials. In the remaining 25% of trials, the last tone was replaced by either a deviant (e.g. 'xxxxX' within 'xxxxY' blocks) or an omission ('xxxx_'). The authors found that local deviants (i.e. 'xxxxY' even when it was the common stimulus) were always accompanied by a measurable MMN (at 80–150 ms), but that in xxxxX blocks (where the *local* deviant 'xxxxY' was also *globally* unexpected) this deflection was larger. By contrast, global deviant responses were found at later latencies (150–600 ms), although no interaction was reported for this dissociation. Note that, for xxxxY blocks, a global deviant amounts to an unexpected repetition eliciting a *stronger* response than an expected alternation – an inversion of the MMN.

Recently, Dürschmid et al. (2016) provided more evidence for a hierarchical organization of mismatch signals, dissociating not time-scales but brain regions, using high-gamma ( > 60 Hz) activity as an index of local spiking. Dürschmid et al. (2016) were able to measure high-gamma components using ECoG recordings from patients with frontal and temporal electrodes, who listened to predictable and unpredictable deviants embedded in an uninterrupted train of tones. In the predictable condition, the deviant tone (550 Hz) always occurred after four consecutive standard tones (500 Hz), rendering the deviant 'globally' predictable. In the unpredictable condition, the deviant tone occurred randomly after at least three presentations of the standard tone, rendering the deviant fully unpredictable. The authors found no main effect of block type, but they did find an interaction: high gamma was found for unpredictable, but not predictable deviants at frontal electrodes, while at temporal electrodes both deviant-types elicited high-gamma responses. The authors interpreted this as demonstrating that frontal cortex monitors 'the bigger picture'. This interpretation is compatible with the source reconstruction results by Wacongne et al. (2011), who also found that global (but not local) deviants activated a broad fronto-parietal network. However, this claim could have been stronger had Dürschmid et al. (2016) manipulated local and global regularities independently, instead of comparing repeating versus random patterns.

Strauss et al. (2015) did present such an independent manipulation. In the same MEG-EEG paradigm as used by Wacongne et al. (2011), the authors showed that late-latency responses to global deviants disappeared categorically in all stages of sleep, concluding that predictive coding was 'disrupted'. This is remarkable, because the MMN persists during sleep (Sculthorpe et al., 2009) and even coma (Fischer et al., 2000; but see Dykstra and Gutschalk, 2015). However, Strauss et al. (2015) demonstrate that the persisting 'sleep MMN' is strongly reduced and lacks sustained fronto-parietal activity. Over and above these differences in degree, sleep-MMN also seemed to be *qualitatively* different. Strauss et al. (2015) showed this by training a classifier to distinguish local standards and deviants. When trained and tested on responses recorded during wakefulness, the algorithm reliably distinguished signals from early (76 ms) to late (620 ms) latencies. However, when the classifier – trained on wakefulness data – was tested on sleep data, it only generalized to early (76–100 ms) and late (212–588 ms) signals. For signals from the MMN latency (100–200 ms) it did not generalize at all, and failed to perform better than chance. The authors interpret this result as new evidence for the idea that MMN might be a consequence of several independent processes: an automatic process arising from passive adaptation (May and Tiitinen, 2010) and therefore persistent during sleep, as well as an active, predictive process which is abolished during sleep.

Interestingly, the effects of sleep were corroborated in the realm of anesthesia. Uhrig et al. (2014) had earlier reported the first neural signature hierarchical novelty responses (potentially an index of PC) in non-human animals, using primate fMRI in macaque. They found that only globally deviant sequences recruited a large fronto-parietal network known in humans as the neuronal workspace (Dehaene et al., 1998). Recently, Uhrig et al. (2016) repeated the experiment under varying degrees of anesthesia. Both anesthetics (propofol and ketamine) weakened local and distorted global mismatch responses. Ketamine was especially powerful, effectively abolishing the global mismatch effect. Since both plasticity (Collingridge and Bliss, 1987) and intra-regional feedback (Self et al., 2012) are thought to be NMDA-dependent, and ketamine impairs MMN even at light dosages (Umbricht et al., 2002), this is perhaps unsurprising. However, ketamine is a popular anesthetic, used by three of the five animal studies here reviewed (Szymanski et al., 2009; Jaramillo and Zador, 2011; Rubin et al., 2016). Since it abolishes global mismatch responses, and the persisting (local) mismatch responses may be qualitatively different (Strauss et al., 2015), these findings underline that future studies of PC should avoid the use of ketamine – and, ideally, of anesthesia altogether.

Finally, Lecaignard et al. (2015) manipulated (global) predictability of auditory deviants, but found no hierarchical effects. Deviant predictability affected ERP amplitudes at early ( < 70 ms), MMN (100–250 ms) and late ( > 300 ms) latencies. Puzzlingly, however, the biggest effect of global predictability was found at the earliest time-window ( < 70 ms), where the MMN was completely abolished *only* in the *globally* predictable condition; an effect which stands in contrast to other studies on hierar-

chical PC and hierarchical deviance detection more generally (Grimm and Escera, 2012; Cornella et al., 2012; Escera and Malmierca, 2014).

Altogether, hierarchy is central to PC and converging evidence now demonstrates that effects of prediction depend on hierarchical processing level. Nevertheless, some ambiguities remain. A first issue is whether hierarchically 'high' and 'low' effects reflect a single mechanism. Some human electrophysiology studies claim they do not (Todorovic and de Lange, 2012; Strauss et al., 2015; López-Caballero et al., 2016). These studies associate effects on early processing with passive adaptation (e.g. the early, sleep-persistent MMN in Strauss et al., 2015) and effects on later, 'higher' processing with prediction (see also Grotheer and Kovács, 2015). However, this hard dichotomy seems at odds with results from animal electrophysiology which report prediction effects already at A1 (Rubin et al., 2016; Rummell et al., 2016; Gill et al., 2008; Ulanovsky et al., 2004). What adds to the ambiguity is that most studies used repetitions or Bernoulli sequences to manipulate prediction, causing expectation and adaptation to be confounded. A second, more subtle ambiguity is whether the discussed interactions between hierarchy and prediction constitute evidence for *hierarchical prediction*, in the sense of hierarchical Bayesian inference. Theoretically, interactions as those in the animal literature – showing that prediction effects become stronger at higher hierarchical levels (Rummell et al., 2016; Gill et al., 2008) – do not necessarily support the notion of hierarchically nested predictions, which would require a task which manipulates multiple, nested (or hierarchically dependent) regularities – as only few studies have done so far. Accordingly, while there is clear evidence for the effect of expectations on responses (Assumption 2) and suggestive clues for hierarchical organization of expectations (Assumption 1) progress will now depend on studies that use stimuli with multiple nested regularities, and which manipulate expectation in a way not confounded by adaptation.

## Hearing silences: Omission as a window into prediction

When omitting a highly expected sound such as a tone in a beat, listeners can 'hear' the absence. In such circumstances, neural responses time-locked to the omitted sound have been observed (Yabe et al., 1997; Raij et al., 1997; Fig. 2D). These 'omission responses' offer an appealing vantage point to study top-down prediction decoupled from bottom-up input, and have become a popular method for studies on predictive coding.

Theoretically, detecting silences could happen either retrospectively (by comparing perceptual input and memory template *after* the input is processed) or prospectively (by directly matching *predictions* to input, as proposed by PC). Bendixen et al. (2009) attempted to dissociate these possibilities. Participants listened to isofrequent tone-pairs of which either the first or the second tone was occasionally omitted. If the second tone was omitted, it could nonetheless be predicted by the first tone ('predictable' condition). But if the first tone was omit-

ted, its identity could only be 'restored' after hearing the second tone (restorable condition). The authors compared evoked responses to a control condition in which the tones were neither predictable nor restorable. When comparing the amplitudes of the early component (up to 50 ms post tone/omission onset) the authors found omission responses in the predictable condition but not in the restorable condition. This was interpreted as pre-activation of the sensory representations of the predicted tones. The authors concluded that auditory expectation works prospectively *and not retrospectively*. However, since they looked for main effects at very short latencies (< 50 ms post onset, identical to the duration of the tone), and focused exclusively on evoked (as opposed to not time locked) responses the analysis may have been biased to finding prospective pre-activations, and not retrospective memory effects.

Hughes et al. (2001) took a similar approach to test whether change-detection involves prediction. Patients undergoing intracranial recordings from temporal cortex performed an oddball paradigm with tones or tone-pairs as standards and silences as oddballs. Strikingly, in all patients, channels firing to tones *also* fired to omissions, often more strongly. Furthermore, 5 of 10 patients exhibited 'omission selective' channels that *only* responded to unexpected omissions, and to other unexpected stimuli like bird-chirps. Finally, and contrary to other demonstrations of omission responses (Raij et al., 1997; Chennu et al., 2016) the effects seemed wholly independent of attention. The omission-selective channels may have been the first recordings of error-units. Unfortunately, Hughes et al. (2001) did not report the exact location or depth of their electrodes, other than being associative (non-primary) auditory cortex, which makes the striking findings somewhat difficult to interpret.

A different approach is described in SanMiguel et al. (2013a,b), who used self-generated sounds to elicit omission responses (Fig. 2A). Participants were asked to press a button every 600–1200 ms, after which a sound was generated in 88%, 50% or 0% of trials. To control for motor activity, the response after button presses that were never followed by a sound (the 0% block) was subtracted from the omission AEP evoked by the unexpected 'silence'. After subtraction, significant omission responses were present in the 88% block, but not in the 50% (random) block. In a follow-up experiment, SanMiguel et al. (2013b) showed that omission responses to self-generated sounds were only elicited if a button press was predictive of both the identity and timing of the elicited sound, rather than just the timing, which suggests that timing alone is not enough to form an accurate prediction of a stimulus.

Chennu et al. (2016) compared omission responses recorded with EEG and MEG. Using a local–global paradigm (Fig. 2H), the fifth tone was a global standard in 74% of trials, and a global deviant or omission in 13% of trials. To confirm that omission responses reflected *expectation* effects and not passive carry-over effects such as oscillatory entrainment (May and Tiitinen, 2010), unexpected omissions of a fifth tone (occurring 14% of trials) were compared to 'expected omissions' from sequences in

which the fifth tone was always omitted. In the EEG recordings, this revealed clear omission responses that were modulated by attention. Surprisingly, in the MEG data the omission response was absent. This divergence between MEG and EEG is interesting but difficult to interpret, and most likely arises from the orientation of the neural sources or measurement noise. However, it might also be consistent with a specific interpretation of omission responses as reflecting prediction units only, which may reside in deeper layers and should therefore be more difficult to detect with MEG.

Fujioka et al. (2009), who also used MEG, induced (and violated) expectations by using a regular musical beat, from which tones were occasionally omitted. Each tone elicited a short gamma ( > 40 Hz) burst, as is typical for stimuli. However, the authors also observed a slow, oscillatory modulation of the beta band that was phase locked to the occurrence of the tone. This slow power-modulation steadily decreased after each beat, reaching its peak just *before* the occurrence of a new tone, potentially indicating an internal rhythmic anticipation signal (see also Fujioka et al., 2012). Intriguingly, when a tone was unexpectedly omitted from the beat, the decrease in beta power was not observed, but a (stimulus-like) sudden peak in gamma was observed. This observation not only supports sensory prediction during beat perception, but also, indirectly, the notion that beta ('prediction') and gamma ('error') bands signal different computational variables (see Section The rhythms of prediction).

Finally, a number of earlier discussed studies also reported omission responses of a varying extent. Todorovic et al. (2011) and Todorovic and de Lange (2012) reported higher field strengths after unexpected than expected silences. However, their effects were rather small and in Todorovic and de Lange (2012) limited to late latencies (200–500 ms). More akin to 'real' evoked responses are the omission responses in Wacongne et al. (2011), who also compared expected with unexpected omissions and found (contra May and Tiitinen, 2010) significant responses only for unexpected omissions, in both MEG and EEG.

Altogether, evidence from EEG (Bendixen et al., 2009; SanMiguel et al., 2013a,b; Chennu et al., 2016), MEG (Wacongne et al., 2011; Todorovic et al., 2011; Todorovic and de Lange, 2012; Chennu et al., 2016) and ECoG (Hughes et al., 2001) shows that omissions can evoke responses that are time-locked to the omitted stimulus and appear to be generated in auditory cortex and superior temporal gyrus. Crucially, omission responses seem to occur only after *unexpected* omissions (Wacongne et al., 2011; Chennu et al., 2016) – challenging the suggestion that they could reflect passive carry-over effects – and only if the omitted sounds are prospectively *predictable* (Bendixen et al., 2009) – suggesting a *predictive* mechanism (cf. Assumption 1). However, the literature also shows some remarkable variability. For instance, using MEG, Todorovic et al. (2011) and Todorovic and de Lange (2012) find small and late deflections, unlike 'real' auditory-evoked fields, and Chennu et al. (2016) find no omission responses at all. Using EEG, Chennu et al. (2016) and Bendixen et al. (2009) find clearer omission responses. However, they are still quite different from 'real' AEPs, or from the striking responses in SanMiguel et al. (2013a,b) or in Hughes et al. (2001). Moreover, while the MEG/EEG omission responses in Raij et al. (1997) and Chennu et al. (2016) are strongly affected by attention, attention had no effect on the ECoG omission responses in Hughes et al. (2001).

Beyond the empirical variability, there is some theoretical variability in how omission responses are interpreted. For some authors (e.g. SanMiguel et al., 2013a,b; Schröger et al., 2015) they are simply expressions of prediction error. This would render omission responses as perhaps *the* signature finding of PC, by showing that evoked responses fundamentally reflect *surprise* – and are thus even observable in the absence of sensory input. However, as Wacongne et al. (2012) point out, this interpretation critically depends on how prediction error is calculated. If one uses subtraction, implemented e.g. by a focussed inhibitory pulse that 'subtracts' predictions from sensory input, it is difficult to see how omissions could elicit prediction error without allowing negative firing rates. In that case, omission responses are perhaps better interpreted as reflecting purely prediction signals, which speaks to their relative weakness and variability. Due to these ambiguities, it is difficult to directly interpret the implications of omission responses to (specific formulations of) predictive coding. Nevertheless, collectively, these studies present highly suggestive, converging evidence of anticipatory mechanisms, operating without conscious expectation, in auditory cortex.

## Predictability and precision

Results with the MMN paradigm demonstrate that listeners are sensitive to the violation of a variety of sound patterns, including very complex regularities. This has been interpreted as (indirect) evidence for the brain's remarkable sensitivity to acoustic patterning. However, a crucial missing link is an understanding of the process by which the brain acquires an internal model of regularities in the environment.

Recently, Barascud et al. (2016; see below for replication by Southwell et al., 2017) presented direct evidence of the discovery and representation of acoustic patterns, using rapid, statistically structured sequences of tone-pips that transitioned from random to regular, and *vice versa* (Fig. 2I). Methodologically, this paradigm constitutes a departure from previous paradigms in two ways: firstly, the use of very rapid sequences precludes conscious discovery of regularity, instead mostly tapping bottom-up-driven processes. Secondly, regularity was manipulated independently from repetition, thus decoupling the effects of predictability from low-level adaptation.

Behaviorally, Barascud et al. (2016) first observed that listeners were extremely quick at detecting the emergence of regular patterns, performing on par with an ideal observer model. Brain responses measured from naïve listeners were equally rapid. Remarkably, the onset of regularity manifested as a large-scale increase in sustained amplitude (Fig. 3). Offsets of regularity (transitions toward randomness), by contrast, were associated with a

large *drop* in sustained power. Source reconstruction identified a network of sources in auditory cortex (AC), inferior frontal gyrus (IFG) and the hippocampus. AC and IFG sources are commonly reported in the context of the MMN and interpreted as supporting the re-entrant error-minimizing process underlying it (Opitz et al., 2002; Garrido et al., 2009b; see Section Dynamic Causal Modeling of MMN).

The finding that emergence (and disappearance) of regularity in unfolding sound sequences is associated with large-scale sustained responses is interesting for various reasons: firstly, it suggests the brain encodes the *state* ('regular' vs 'random') rather than just the transition (as in e.g. MMN). Secondly, the amplitude pattern [(regular) > (random)] is not easily interpretable in terms of simple physical attributes of the signal – adaptation, for example, would result in the opposite pattern. Finally, the neural signature of complex regularity detection (i.e. enhanced responses) is *opposite* to that of simpler regularity detection (i.e. attenuated responses) observed in many previous experiments, for example using the roving standard paradigm.

Interestingly, the effect is also opposite to all PC effects we have been considering so far, in which predictability is associated with *weaker* responses. Barascud et al. (2016) suggested precision-weighting could underlie this inversion: if regularity is an index of reliability or precision, PC predicts that regular signals are up-weighted and prioritized for further processing (see Section Attention as precision). As many biological stimuli unfold as regularities over time, it also seems bio-logically useful to prioritize such signals, for instance for subsequent auditory object formation or scene analysis.

Evidence in line with this interpretation was subsequently presented by Sohoglu and Chait (2016b) who used artificial 'scenes' consisting of concurrent tone-pip streams (modeling acoustic sources) which were temporally regular or random. Participants were quicker and more accurate to detect an object appearing in a tem-porally regular scene, and enjoyed an additional slight benefit if the object itself was regular. MEG responses in both passive listeners and listeners actively engaged in detecting the occasional appearance of a new source within the scene revealed increased sustained activity in scenes comprised of regular sources. Over and above this 'scene effect', new source appearance in regular sce-nes was also associated with increased responses rela-tive to random scenes – an effect interpreted as evidence for a mechanism that infers the precision of sen-sory input and uses this information to up-regulate neural processing toward more reliable sensory signals.

More clues on the amplifying effect of regularity are found in Hsu et al. (2015). Subjects listened to sequences of tones with ascending frequencies in which the final tone varied. In 75% of trials, the tone complied with the local regularity (predicted condition). In 12.5% of trials, the last tone was unexpectedly lower than the first tone, violating the expectation induced by the ascending sequence ('mispredicted' condition). Finally, in 12.5% of trials, the sequence was jumbled altogether. The authors found that

while predicted tones elicited a weaker N1 deflection than mispredicted ones (a well-documented expectation effect), wholly *unpredicted* tones elicited an even weaker N1 still. According to Hsu et al. (2015), this is because predicted and mispredicted responses express both a prediction and a (small or large) prediction error, but unpredicted responses reflect *only prediction error* and are therefore weakest. However, as remarked by Ross and Hansen (2016), it seems at odds with the probabilistic nature of PC to assume predictions are *absent* in the unpredicted condition: rather, what distinguishes the unpredicted condition is the low predictability of the signal. The attenuated N1 to wholly unpredictable stimuli might be understood as *inversion* of the enhanced response to predictable stimuli in Barascud et al. (2016) and Sohoglu and Chait (2016b): the brain might consider the jumbled tone ladder as noisy and uninformative, hence down-weighting the response.

In sum, accumulating evidence suggests that, at least under certain conditions, predictability may enhance, rather than suppress, neural responses. This result fits into the PC framework if one considers effects of precision: sequences of random stimuli may be deemed uninteresting noise (low precision) and hence down-weighted, while streams containing a regularity are considered informative and are hence up-weighted. However, since precision can explain effects that are opposite to 'traditional' PC effects, invoking it begs the question when, exactly, predictability is supposed to suppress neural responses and when it should enhance them. As we will see, this need for a 'principled account' will be a recurring theme in studies that examine the main manifestation of precision-weighting — i.e. attention.

## Attention as precision

Because the world is variable and the brain noisy, a degree of prediction error is inevitable. Distinguishing such 'residue error' (related to noise) from relevant error (related to incorrect beliefs or changes in the world) requires that not all prediction error is treated equally. A Bayes-optimal approach, successfully applied in engineering (Kalman, 1960) as well as neuroscience (Yu, 2014) is to weight errors by their *reliability*, typically quantified as the uncertainty of predictions relative to that of observations, a coefficient known as *Kalman gain* (Kalman, 1960; Anderson and Moore, 1979). When the gain (precision) is high, inputs are up-weighted and will dominate inference; when it is low, inputs are down-weighted and predictions dominate inference. Several authors in the predictive coding field (Rao, 2005; Spratling, 2008a,b, 2010; Feldman and Friston, 2010; see also Dayan and Zemel, 1999; Dayan and Yu, 2003; Yu and Dayan, 2005a,b) have used such optimal handling of uncertainty as a framework for attention, since it offers normative principles that can explain selective processing by motivating *why* some signals are computationally more relevant than others.

Uncertainty-weighting affects inference and learning differently; here, we will focus on perceptual *inference* (but see Yu, 2014, for a treatment of Bayesian

approaches to attention which also covers learning). During inference, reliable inputs are weighted more strongly, and PC proposes that attending to a feature amounts to *expecting* that signals with this feature will be *reliable* or *informative,* and should thus be prioritized for processing (Feldman and Friston, 2010). Response strength should therefore always be a function of both the size of the error and its precision. In other words, every brain response should be sensitive to attentional modulation. This proposal implies a departure from accounts of MMN that describe MMN as pre-attentive (Garrido et al., 2009b; Jääskeläinen et al., 2004; Winkler and Czigler, 1998).

Preliminary support is found in Chennu et al. (2013). The authors recorded EEG while presenting blocks of tone sequences to one ear, occasionally replacing the fifth tone by either a different tone in the same ear (monaural deviant) or by the same tone in the opposite ear (interaural deviant). Additionally, participants counted deviant tones (attend tones) or deviant sequences (attend sequences) or performed a visual task (distraction). Focussing here on local deviants at MMN latencies, both monaural and interaural MMN were reduced during distraction compared to attending sequences. Attending tones, however, instead of amplifying the MMN (by increasing gain of error-neurons), attenuated it. The authors suggest their manipulation may have been confounded: counting deviant tones did not only focus attention on tones (just as counting sequences did) but might have also increased conscious expectation of unexpected tones, thus decreasing overall surprise.

To circumvent this confound, other studies manipulated attention and prediction orthogonally. Auksztulewicz and Friston (2015a) used a roving standard paradigm in which participants attended to one of two time-windows (early or late), after which the roving standard was presented at each window with an independent probability of 50%. Participants reported if there was *no* stimulus at the attended latency. Only trials where the tone was presented at *both* latencies were included, thus rejecting all possible motor artifacts. A significant interaction effect was found; specifically, MMN was observed in attended, but not in unattended time-windows. This attentional enhancement of MMN is compatible with precision weighting. Note, however, that the non-significance of MMN outside the scope of attention seems to contradict earlier findings that MMN is not dependent on attention; note, too, that the effects reported by Auksztulewicz and Friston (2015a) are relatively late and relatively short – for instance, the MMN only reached significance between 190 and 210 ms and the deviance-attention interaction only between 193 and 197 ms. Both may have been related to a lack of power after rejecting so many trials.

Another independent manipulation was reported by Hsu et al. (2014). The authors presented two streams of tone pairs: in one stream, the frequency of the second tone in a pair was always two natural keys higher than the first; in the other stream, the relationship between the first and second tone was random. Attention was manipulated by asking participants to report occasional tones with decreased loudness in one of the streams. The authors found an interaction of attention and prediction on N1 amplitudes. Specifically, attended/predictable tones elicited a stronger response than all other tones, between which differences were non-significant. This includes attended *versus* unattended unpredictable tones, hence the authors concluded that attentional enhancement of N1 depends on prediction. Note, however, that this interacting effect between attention and prediction (attention reversing the effect of prediction) is at odds with Auksztulewicz and Friston (2015a), who found the opposite (attention enhancing prediction effects) 100 ms later.

A recent EEG study by Garrido et al. (2017) compared the two accounts explicitly. Participants were presented Gaussian white noise to both ears and had to detect silent gaps in one or both ears. Embedded in the noise, task-irrelevant oddball sequences were presented. The authors formulated two models of the interplay between attention and prediction: in the first, attention could reverse the effect of prediction (Hsu et al., 2014; see also Kok et al., 2012). In the second model, attention enhanced responses, predicted and unpredicted alike. The authors observed a MMN, and found that attention enhanced it, but contrary to Hsu et al. (2014) they found no interaction. In line with this observation, Bayesian model comparison favored the opposition model. Contrary to Auksztulewicz and Friston (2015a), but in line with the MMN literature, MMN was also found in the absence of attention.

Rather than deliberately directed, attention is sometimes automatically attracted to a stimulus. Stimuli with this capacity are called *salient* (Itti et al., 1998). Predictive coding accounts for salience by appealing to the *intrinsic precision* of stimuli. Intense stimuli, for instance, can be seen as having a high signal-to-noise ratio due to sheer signal strength; inversely, regular stimuli would enjoy high precision by virtue of *low variance*. Indeed, this latter effect was proposed by Barascud et al. (2016) to explain large increases in MEG signals induced by auditory regularities (Fig. 3). Such up-weighting of regular sounds seems ethologically sensible, as regular patterns often carry stable, behaviorally relevant information about the world. The account also has a straightforward empirical consequence – regular stimuli should attract attention. In vision, a recent study indeed demonstrates this effect (Zhao et al., 2013).

Southwell et al. (2017) tested this idea in the auditory domain. Using EEG, the authors first replicated the MEG-effects observed by Barascud et al. (2016): task-irrelevant regular sequences (as used by Barascud and colleagues) induced large increases in sustained EEG amplitude. Next, the authors tested behaviorally whether the same regular patterns would capture attention more strongly, measured as the interference with concurrent tasks. Remarkably, their results suggested that regularity was not more distracting (if task-irrelevant) or more salient (if task-relevant) than random patterns. The fact that *neurally*, regularity induces marked sustained amplitude increases, but *behaviorally* the same patterns are not more salient, contradicts the attentional gain explanation proposed by Barascud et al. (2016). Southwell et al. (2017) suggest that this leaves us with three alternative

hypotheses: Either the gain in amplitude reflects an upsurge of (poly-synaptic) *inhibition* or explaining away by higher regions, which is not dissociable from excitation using M/EEG. Alternatively, it may reflect a number of quite distinct processes. Or finally, it could reflect some form of precision-weighting which does not manifest as high-level attentional capture. This last possibility, however, would imply that under PC attention is an expression of precision-weighting, but precision-weighting does not (always) express as attention. While logically possible, this creates an awkward disconnect between neural responses and cognition, and calls for a more principled approach to decide when precision weighting is attentional or not.

Altogether, the depiction of attention as the weighting of sensory signals by their (expected) *precision* (Feldman and Friston, 2010; see also Rao, 2005; Spratling, 2008a, b) elegantly integrates many known attentional effects into the realm of prediction. However, the increased opportunities this creates for post-hoc explanations are – at least in the auditory domain – not yet met by a proportional increase in rigorous confirmatory results. For endogenous attention, studies explicitly testing the account report small and sometimes conflicting effects (Chennu et al., 2013; Hsu et al., 2014; Auksztulewicz and Friston, 2015a; Garrido et al., 2017). For exogenous attention, precision-weighting offers a compelling explanation for the enhancing effect of regularity (Barascud et al., 2016; Sohoglu and Chait, 2016b; Hsu et al., 2015; Southwell et al., 2017); however, the direct consequence of this claim (that regularity should be salient) does not seem to hold (Southwell et al., 2017). More research is needed to test and potentially revise the notion of auditory precision-weighting, and to explore differences with vision where it may apply more readily (e.g. Kok et al., 2012; Zhao et al., 2013).

**The rhythms of prediction**

In systems neuroscience, distinct oscillatory signatures for feedforward processing (operating mainly via the gamma band) and feedback processing (using alpha and beta bands) have been demonstrated in considerable detail (van Kerkoerle et al., 2014; Buschman and Miller, 2007). In standard PC, this *oscillatory* asymmetry is hypothesized to be linked to the *functional* asymmetry between (upward) errors and (backward) predictions. In other words, predictions and errors should have distinct *oscillatory signatures* (Arnal and Giraud, 2012; Bastos et al., 2012). However, evidence for this claim has remained indirect (see Arnal et al., 2011 for a demonstration in speech perception; van Pelt et al., 2016 in causal cognition).

Recently, Sedley et al. (2016) provided more direct evidence, using a simple parametric task to generate auditory stimuli while recording local field potentials using ECoG. Three human subjects listened to short (300-ms) sequences of harmonic complexes of which only the fundamental frequency varied. In any given trial there was a 7/8 chance that f0 would be sampled from the same Gaussian population, and a 1/8 chance that it would be sampled from a new one. Assuming that subjects uncon-

sciously tracked the statistics, the authors used a Bayes-optimal inversion of their generative algorithm to calculate trial-by-trial estimates of four key inferential variables: *prediction error*, *surprise*, *prediction change* and *prediction precision* (where surprise is the precision-weighted variant of prediction error). The authors then regressed these variables against a time-frequency decomposition of the LFP trace. As expected, the authors found that gamma was predicted by surprise (more so than by prediction error). Moreover, beta-band modulations were significantly predicted by *prediction change*. Finally, and not explicitly predicted by PC, the authors found that alpha band modulations were significantly predicted by the *precision of predictions*, although this effect was less pronounced than that in the beta and gamma band.

Among the earlier discussed studies, only Fujioka et al. (2009) reported effects similarly compatible with PC. There, an oscillatory stimulus (a beat) induced an oscillatory modulation of the beta band that was time-locked to the beat. When a tone was omitted, the immediate decrease in beta-power was not observed, suggesting that the rhythmic beta-power modulations may have reflected an oscillatory *expectation*. Moreover, omissions did induce short gamma bursts, characteristic of stimuli (or surprise). Other studies, however, did not report clear oscillatory dissociations. Signatures of prediction in the beta-band, for instance, were absent in Dürschmid et al. (2016) who reported ECoG recordings to predictable and unpredictable deviants. The authors made sure they compared electrodes with similar sensitivity for different frequency bands, and nevertheless only found effects seemed in the high-gamma band ( > 60 Hz) and at low frequencies related to evoked potentials, but hardly in between.

El Karoui et al. (2015) presented ECoG recordings of patients performing a local–global paradigm and found a *decrease* in sustained beta power after global mismatches (which would arguably involve *more* prediction-change). However, the global deviants were also the behavioral target, confounding attention and prediction, and making interpretation difficult. Finally, Todorovic et al. (2015) found effects of attention and expectation *only* in the beta-band, which decreased in power after unexpected tones, but only if attention was directed to another, earlier time window.

To summarize, evidence for distinct oscillatory signatures of prediction and error processing is limited, indirect and mixed: only two of six studies revealed spectral patterns compatible with the predictions of PC. Methodological differences make it difficult to draw an unequivocal conclusion on the existence of oscillatory differences between prediction and error processing. Given the increasing evidence for laminar differences between alpha/beta and gamma band dominance (e.g. Scheeringa et al., 2016), oscillatory differences are a potential tool to test the standard implementation of PC, and future studies using parametric methods like Sedley et al. (2016) may offer much needed confirmatory evidence. However, simply interpreting different bands as reflecting different variables without employing a parametric approach to calculate the relevant variables on a trial-

by-trial basis seems empirically unwarranted given the highly variable results of studies without such a model-based approach.

# EFFECTIVE CONNECTIVITY – CLUES FROM DCM

Measurement in neuroscience typically allows for high spatial or high temporal resolution. Accordingly, many studies probe the 'when' or 'where' of neural responses. However, this provides little insight in *how* responses emerge. Causal modeling techniques attempt to overcome this by estimating changes in causal influences between sources underlying effects of interest. One of these techniques – Dynamic Causal Modelling (Friston et al., 2003) – has been extensively used to test predictive coding, especially in relation to the MMN. Since DCM is a theory-driven method which makes several enabling assumptions, we will briefly recapitulate the ideas behind DCM before discussing the studies that used it.

## DCM for MEG and EEG

DCM is a hypothesis testing framework, which works by predicting neural responses based on several hypotheses, and then comparing these predictions to the data. Predictions are generated by combining a *neuronal* and an *observational model*. In DCM for M/EEG (Kiebel et al., 2006, 2008), the observational model is a *lead field* as used in source reconstruction, which maps hidden dipoles in the skull to observable deflections at the scalp. DCM goes beyond this 'common' reconstruction method by using a *neuronal model* to explicitly model intracranial current flow. Neuronal models in DCM for M/EEG (see Moran et al., 2013 for review) are mostly *mass models*, which do not capture the complex dynamics between large numbers of individual neurons (as found in the skull) but rather the simpler dynamics between massively synchronized *populations* of neurons (as measurable at the scalp). Typically, a region is described with three or four sub-populations of inhibitory and excitatory neurons (each modeled using an ordinary second-order differential equation) that operate as a dampened linear oscillator (David and Friston, 2003; David et al., 2006).

In DCM, hypotheses are embodied as *architectures*: cortical sources connected in a specific, directional way. Responses can be generated by injecting a Gaussian impulse into one source (e.g. A1), after which the current flow ensuing from the network is passed through the lead field to generate observational patterns for the modality in use (EEG or MEG). Between-trial effects are modeled as changes in *extrinsic* or *intrinsic* connectivity. *Extrinsic connectivity* refers to coupling strength between regions, is modeled by directional coupling parameters, and can be thought of as inter-regional synaptic modulation (c.f. learning). *Intrinsic connectivity* refers to the strength with which a signal is propagated *within* a region. It is adjusted by changing the maximum firing rate of excitatory populations, and can be thought of as changing the *excitability* of a region (c.f. adaptation). Ultimately, the architecture that can most

readily explain the effect – yielding the best fit with the least complexity – is deemed *most likely.*

By virtue of these assumptions, DCM aims to provide an *in silico* environment for testing hypotheses about both the neural architecture underlying experimental data, and the changes *within* this architecture that best explain between-trial effects of interest.

## Dynamic Causal modeling of MMN

The first application of DCM to MEG and EEG is described in Garrido et al. (2007a,b) who modeled the difference between standard and deviant ERPs from an oddball paradigm. Garrido et al. (2007a,b) found that the difference between standard and deviant responses was best explained by bidirectional connectivity changes between Heschl's Gyrus (A1), superior temporal gyrus (STG) and right inferior frontal gyrus (rIFG). Garrido et al. (2007b) replicated this basic result at the group-level and verified that backward modulations were especially important for explaining ERP differences at later latencies (200–400 ms).

Having established these foundational results, Garrido et al. (2008) used DCM to compare theoretical accounts of MMN. In the study, the authors modeled a series of responses from the roving standard paradigm (see Fig. 2), from deviant (first tone) to standard (last tone). They then compared which MMN–hypothesis could explain the associated ERP differences – and thus the differential MMN – best. Each MMN hypothesis was embodied as a different variation of the frontotemporal architecture outlined above (see Fig. 4). The adaptation hypothesis was modeled as a network in which only the excitability of A1 varied over trials. The model-adjustment hypothesis (which explains the MMN as a fronto-temporal memory-adjustment; cf. Näätänen et al., 1978, 2007) was modeled as a network in which only the between-region connectivity varied between trials. Finally, predictive coding was embodied in a model in which both the excitability of A1 and inter-regional connectivity varied. The idea was that PC incorporates both adaptation and model adjustment (see also Section Predictive coding and MMN) – in this view, changes in excitability of A1 and fronto-temporal coupling are expressions of belief-updating at different hierarchical levels (intra-regional microcircuitry versus inter-regional network connectivity). Model comparison showed that the hybrid PC model explained the ERP differences best. The superiority of hybrid model was later replicated in a study using the 'classic' frequency oddball (Garrido et al., 2009a).

## Temporal deviants and top-down predictions

Within the same model space, Phillips et al. (2015) replicated this result using MEG and stimuli that deviated across various dimensions, such as frequency, intensity, or duration. To study all these deviant dimensions, the authors used an optimized oddball paradigm (Näätänen et al., 2004), in which each block starts with several standard tones, after which standards start alternating with *different* deviants – e.g. standard, frequency-deviant, standard, duration-deviant, standard, *etc*. First, within
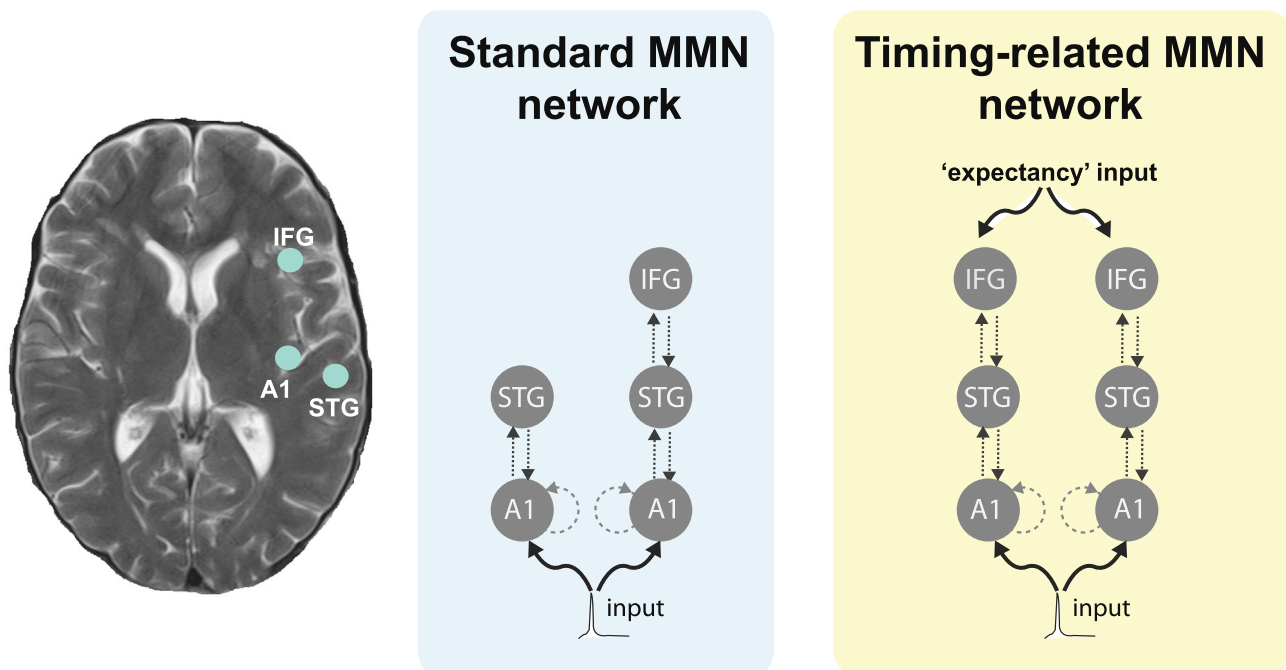
**Fig. 4.** Graphical specification of connectivity models underlying the MMN as suggested by DCM. Left: connectivity modulations in an asymmetric frontotemporal network, combined with neuronal excitability modulations in A1, was shown to best explain the MMN across a variety of paradigms and modalities (Garrido et al., 2008, 2009a; Phillips et al., 2015, 2016; Chennu et al., 2016; Barascud et al., 2016). Right: connectivity model including left IFG and frontal 'expectancy inputs' which was found to best explain MMN responses to temporal irregularities (duration and silent gap) or omissions (Phillips et al., 2015, 2016; Chennu et al., 2016).

the model space of Garrido et al. (2008, 2009a), the model with forward, backward and intrinsic modulations in A1 was confirmed to 'win' for all dimensions. Subsequently, the model space was extended to include architectures with left IFG and models with an additional, frontal input. Usually, Gaussian impulse functions are models of sensory inputs, and are only 'injected' at primary sensory regions. By contrast, Phillips et al. (2015) located a second input at IFG ('expectancy inputs'; see Fig. 4). Interestingly, models that included a prefrontal 'expectation' input only provided a better fit for temporal deviants – that is, either tones containing a short silent gap in the center, or tones that deviated in duration. Models that included bilateral IFG were more likely across all stimulus dimensions.

In a follow-up study, Phillips et al. (2016) first replicated these findings by performing the same analysis on a new MEG recording of 50 subjects. They then extended the analyses to ECoG data. As explained above, in DCM for MEG and EEG current flow ensuing from the network is passed through a *lead field* to generate observational patterns specific to M/EEG. As this additional model may introduce uncertainty, it is important to verify whether inverting a DCM *without* observation model (i.e. using signals directly from cortex) yields similar results. The authors recruited two patients: one with electrodes over *right* IFG and STG, and one with electrodes over left IFG and STG. The ECoG DCM results matched earlier DCM results with respect to the relative importance of forward/backward interactions. However, the frontal expectancy input 'won' only in the patient with left-lateralized

electrodes. Strikingly, this asymmetry was also found in the MEG results: temporal deviants were best explained by models with a *left,* rather than bilateral, IFG input. This apparent lateralization is remarkable and calls for a replication, since earlier studies did not consider left IFG a 'main MMN generator' (Opitz et al., 2002; Garrido et al., 2008, 2009a,b; Chennu et al., 2016). Alternatively, the effect may be related to differences in electrode locations of left versus right IFG. This artifact would be propagated to the MEG results because the coordinates from the ECoG electrodes were used as source coordinates in the observation model.

Finally, Chennu et al. (2016) performed a DCM analysis on MEG and EEG data from a local–global paradigm that included omissions. In two conditions, participants either counted uncommon sequences (attend-auditory) or performed an unrelated visual task (attend-visual). For deviant tones, the 'classic' architecture used by (Garrido et al., 2007a,b, 2008, 2009a) best explained the data both in the attended and unattended condition. For the omission responses, by contrast, an architecture that included bilateral IFG and a frontal expectancy input (which replaced the thalamic sensory input) best explained the data, which is compatible with the idea that omission responses reflect top-down prediction (rather than prediction error).

**Discussion**

To summarize, DCM studies show that models which modulate both A1 excitability and fronto-temporal

connectivity explain deviant responses in oddball paradigms (Garrido et al., 2007a,b, 2009a) and variations thereof (Garrido et al., 2008; Phillips et al., 2015, 2016; Chennu et al., 2016) better than models that modulate only A1 excitability or fronto-temporal connectivity. Moreover, responses to tones that deviate temporally, or are omitted altogether, are best explained by models with frontal 'expectation inputs' which replace (Chennu et al., 2016) or augment (Phillips et al., 2015, 2016) the thalamic sensory input.

These patterns of effects are in line with PC by describing MMN not only via A1 adaptation or long-range connectivity, but via a mechanism that combines both. Moreover, an interesting analogy might be drawn between the need for frontal inputs to explain temporal deviants in DCM (Phillips et al., 2015, 2016) and the fact that temporal deviants constitute a key difference between network-level MMN, which is sensitive to temporal deviants, and neuron-level SSA, which is not (Khouri and Nelken, 2015). Although this post-hoc analogy would require further investigation, the fact that only intracellular recordings and DCM appear to consistently distinguish temporal deviants from other deviants illustrates the potential of the technique to extend beyond traditional analysis of non-invasive data.

However, the DCM studies have several limitations. The first issue is that DCM relies on assumptions and simplifications which are not fully validated. The neural mass models used in most DCM for M/EEG studies are even abstracted to such degree that some parameters don't have obvious physiological substrates. One response to this is to develop more complex models with more biologically meaningful parameters (Moran et al., 2013); an approach that is showing promising results (Gilbert et al., 2016). However, this does not yet address the issue of validation. Although initial studies have established the *face validity* of DCM for M/EEG (Garrido et al., 2007a,b, 2009a,b) and the extensively replicated MMN results demonstrate *predictive validity* (Phillips et al., 2015, 2016; Chennu et al., 2016) much needs to be done before DCM can be said to have *construct validity*. Combining different techniques, such as in Phillips et al. (2016), will be critical in this process. Note however that Phillips et al. (2016) only partially validated the *observation model*, which was arguably the least controversial.

A second issue is to what extent these results support predictive coding. Even if we fully accept the network modulations suggested by DCM, this doesn't mean that these changes necessarily reflect predictive coding, or even a single underlying mechanism. Indeed, it is difficult to see why changes in A1 excitability and STG-IFG connectivity should be uniquely characteristic of predictive coding. This problem is reinforced by the fact that the discussed studies have mostly used designs in which expectation and adaptation are confounded, which makes arbitrating between predictive and non-predictive interpretations difficult. As such, while the discussed studies constitute exciting methodological developments in the analysis of non-invasive electrophysiological data, their strength as empirical support for predictive coding theory seems limited.

## CONCLUSION

In this review we aimed to provide a comprehensive empirical evaluation of five key assumptions of predictive coding theory in the context of auditory pattern processing. Findings from animal, human and computational neuroscience provide converging evidence for the fundamental influence of expectations on neural responses and specifically the notion of *prediction error* as a model of neural responsiveness (Assumption 2). Studies on unexpectedly omitted stimuli provide support for the anticipatory, *predictive* nature of these expectancy effects (Assumption 1). Moreover, the dissociation of expectancy effects at different hierarchical levels in both animal and human literature seems suggestive of the hierarchical nesting of predictions, as postulated by predictive coding theory and implied by Dynamic Causal Modeling results (Assumption 1), although more experiments are needed that explicitly manipulate multiple, nested regularities. As to the remaining three assumptions, the picture is less clear. Critically, for the existence of separate prediction and error neurons residing in distinct cortical layers (Assumption 3), there is currently no evidence in the auditory domain in line with this idea (but see Bell et al., 2016; Kok et al., 2016, for recent studies in vision). The recent development to conceptualize attention as the weighting of sensory input by sensory precision (Assumption 4) has provided elegant post-hoc explanations for a broad range of phenomena, but in the auditory modality these have not yet been supported by rigorous confirmatory results. Finally, the dissociation between different frequency bands and computational variables in PC (Assumption 5) has been demonstrated by one study which explicitly estimated the variables on a trial-by-trial basis; studies that did not use such a model-based approach however mostly failed to find similar associations. Looking to the future, progress in the field will critically depend on investigating these assumptions in order to test and revise or falsify specific implementations of PC. Doing so will require closer collaboration between sub-disciplines, in particular between animal and human research, where methodological and conceptual differences currently create interpretational difficulties. Finally, to test crucial theoretical distinctions (e.g. prediction error versus precision-weighted prediction error) there is an ongoing need for computationally explicit analyses in both human and animal neuroscience.

In short, over the past decade a broad range of findings in auditory neuroscience have pointed to a fundamental role of expectations and prediction errors in sensory processing. Going from these findings to the alternative, overarching framework envisioned by PC, however, requires a number of theoretical steps between which the empirical links are currently missing. Uncovering, revising or potentially refuting these

'missing links' is difficult but feasible, and provides an exciting neuroscientific challenge for the years to come.

## ACKNOWLEDGMENTS

## REFERENCES

Anderson BDO, Moore JB (1979) Optimal filtering. Eaglewood Cliffs, NJ: Prentice-Hall.

Arnal LH, Wyart V, Giraud A-L (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Neurosci 14:797–801. https://doi.org/10.1038/nn.2810.

Arnal LH, Giraud AL (2012) Cortical oscillations and sensory predictions. Trends Cogn Sci. https://doi.org/10.1016/j.tics.2012.05.003.

Auksztulewicz R, Friston K (2015a) Attentional enhancement of auditory mismatch responses: a DCM/MEG study. Cereb Cortex 1–11. https://doi.org/10.1093/cercor/bhu323.

Auksztulewicz R, Friston K (2015b) Repetition suppression and its contextual determinants in predictive coding. Cortex. https://doi.org/10.1016/j.cortex.2015.11.024.

Baess P, Widmann A, Roye A, Schröger E, Jacobsen T (2009) Attenuated human auditory middle latency response and evoked 40-Hz response to self-initiated sounds. Eur J Neurosci 29:1514–1521. https://doi.org/10.1111/j.1460-9568.2009.06683.

Barascud N, Pearce MT, Griffiths TD, Friston KJ, Chait M (2016) Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proc Natl Acad Sci USA 113: E616–E625. https://doi.org/10.1073/pnas.1508523113.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76:695–711. https://doi.org/10.1016/j.neuron.2012.10.038.

Bell AH, Summerfield C, Morin EL, Malecek NJ, Ungerleider LG (2016) Encoding of stimulus probability in macaque inferior temporal cortex. Curr Biol. https://doi.org/10.1016/j.cub.2016.07.007.

Bendixen A, Schröger E, Winkler I (2009) I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. J Neurosci 29:8447–8451. https://doi.org/10.1523/JNEUROSCI.1493-09.2009.

Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315:1860–1862. https://doi.org/10.1126/science.1138071.

Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibáñez A, Owen AM, Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction. J Neurosci 33:11194–11205. https://doi.org/10.1523/JNEUROSCI.0114-13.2013.

Chennu S, Noreika V, Gueorguiev D, Shtyrov Y, Bekinschtein TA, Henson R (2016) Silent expectations: dynamic causal modeling of cortical prediction and attention to sounds that weren't. J Neurosci 36:8305–8316. https://doi.org/10.1523/JNEUROSCI.1125-16.2016.

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204. https://doi.org/10.1017/S0140525X12000477.

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

Collingridge GL, Bliss TVP (1987) NMDA receptors – Their role in long-term potentiation. Trends Neurosci 10:288–293. https://doi.org/10.1016/0166-2236(87)90175-5.

Cornella M, Leung S, Grimm S, Escera C (2012) Detection of simple and pattern regularity violations occurs at different levels of the auditory hierarchy. PLoS One 7:e43604. https://doi.org/10.1371/journal.pone.0043604.

Costa-Faidella J, Baldeweg T, Grimm S, Escera C (2011) Interactions between "what" and "when" in the auditory system: temporal predictability enhances repetition suppression. J Neurosci 31:18590–18597. https://doi.org/10.1523/JNEUROSCI.2599-11.2011.

David O, Friston KJ (2003) A neural mass model for MEG/EEG: Coupling and neuronal dynamics. Neuroimage 20:1743–1755. https://doi.org/10.1016/j.neuroimage.2003.07.015.

David O, Kilner JM, Friston KJ (2006) Mechanisms of evoked and induced responses in MEG/EEG. Neuroimage 31:1580–1591. https://doi.org/10.1016/j.neuroimage.2006.02.034.

Dayan, P., Zemel, R.S.R., 1999. Statistical models and sensory attention. 9th Int. Conf. Artif. Neural Networks ICANN '99 1999, 1017–1022, doi: 10.1049/cp:19991246.

Dayan P, Yu AJ (2003) Uncertainty and learning. IETE J. Res. 49:171–181. https://doi.org/10.1080/03772063.2003.11416335.

Dehaene S, Kerszberg M, Changeux JP (1998) A neuronal model of a global workspace in effortful cognitive tasks. Proc Natl Acad Sci USA 95:14529–14534. https://doi.org/10.1073/pnas.95.24.14529.

Dürschmid S, Edwards E, Reichert C, Dewar C, Hinrichs H, Heinze H-J, Kirsch HE, Dalal SS, Deouell LY, Knight RT (2016) Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. Proc Natl Acad Sci USA 113:6755–6760. https://doi.org/10.1073/pnas.1525030113.

Dykstra AR, Gutschalk A (2015) Does the mismatch negativity operate on a consciously accessible memory trace? Sci Adv 1: e1500677. https://doi.org/10.1126/sciadv.1500677.

Egner T, Summerfield C (2013) Grounding predictive coding models in empirical neuroscience research. Behav Brain Sci 36:210–211. https://doi.org/10.1017/S0140525X1200218X.

El Karoui I, King JR, Sitt J, Meyniel F, Van Gaal S, Hasboun D, Adam C, Navarro V, Baulac M, Dehaene S, Cohen L, Naccache L (2015) Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: An intracranial study in humans. Cereb Cortex 25:4203–4212. https://doi.org/10.1093/cercor/bhu143.

Eliades SJ, Wang X (2008) Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453:1102–1106. https://doi.org/10.1038/nature06910.

Escera C, Malmierca MS (2014) The auditory novelty system: an attempt to integrate human and animal research. Psychophysiology 51:111–123. https://doi.org/10.1111/psyp.12156.

Feldman H, Friston KJ (2010) Attention, uncertainty, and free-energy. Front Hum Neurosci 4:215. https://doi.org/10.3389/fnhum.2010.00215.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47. https://doi.org/10.1093/cercor/1.1.1.

Fischer C, Morlet D, Giard M (2000) Mismatch negativity and N100 in comatose patients. Audiol Neurootol 5:192–197. https://doi.org/10.1159/000013880.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19:1273–1302. https://doi.org/10.1016/B978-012264841-0/50054-8.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836. https://doi.org/10.1098/rstb.2005.1622.

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138. https://doi.org/10.1038/nrn2787.

Fujioka T, Trainor LJ, Large EW, Ross B (2012) Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. J Neurosci 32:1791–1802. https://doi.org/10.1523/JNEUROSCI.4107-11.2012.

Fujioka T, Trainor LJ, Large EW, Ross B (2009) Beta and gamma rhythms in human auditory cortex during musical beat processing. Ann N Y Acad Sci:89–92. https://doi.org/10.1111/j.1749-6632.2009.04779.x.

Gagnepain P, Henson RN, Davis MH (2012) Temporal predictive codes for spoken words in auditory cortex. Curr Biol 22:615–621. https://doi.org/10.1016/j.cub.2012.02.015.

Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2007a) Evoked brain responses are generated by feedback loops. Proc Natl Acad Sci USA 104:20961–20966. https://doi.org/10.1073/pnas.0706274105.

Garrido MI, Kilner JM, Kiebel SJ, Stephan KE, Friston KJ (2007b) Dynamic causal modelling of evoked potentials: A reproducibility study. Neuroimage 36:571–580. https://doi.org/10.1016/j.neuroimage.2007.03.014.

Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM (2008) The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage 42:936–944. https://doi.org/10.1016/j.neuroimage.2008.05.018.

Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2009a) Dynamic causal modeling of the response to frequency deviants. J Neurophysiol 101:2620–2631. https://doi.org/10.1152/jn.90291.2008.

Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009b) The mismatch negativity: A review of underlying mechanisms. Clin Neurophysiol. https://doi.org/10.1016/j.clinph.2008.11.029.

Garrido Marta I, Rowe Elise G, Veronika Halász JBM (2017) Bayesian mapping reveals that attention boosts neural responses to predicted and unpredicted stimulino title. Cereb Cortex 1–12. https://doi.org/10.1093/cercor/bhx087.

Gilbert JR, Symmonds M, Hanna MG, Dolan RJ, Friston KJ, Moran RJ (2016) Profiling neuronal ion channelopathies with non-invasive brain imaging and dynamic causal models: Case studies of single gene mutations. Neuroimage 124:43–53. https://doi.org/10.1016/j.neuroimage.2015.08.057.

Gill P, Woolley SMN, Fremouw T, Theunissen FE (2008) What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. J Neurophysiol 99:2809–2820. https://doi.org/10.1152/jn.01270.2007.

Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10:14–23. https://doi.org/10.1016/j.tics.2005.11.006.

Grimm S, Escera C (2012) Auditory deviance detection revisited: evidence for a hierarchical novelty system. Int J Psychophysiol 85:88–92. https://doi.org/10.1016/j.ijpsycho.2011.05.012.

Grotheer M, Kovács G (2015) Can predictive coding explain repetition suppression? Cortex. https://doi.org/10.1016/j.cortex.2015.11.027.

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

Horváth J, Winkler I (2004) How the human auditory system treats repetition amongst change. Neurosci Lett 368:157–161. https://doi.org/10.1016/j.neulet.2004.07.004.

Hsu Y, Le Bars S, Ha JA (2015) Distinctive representation of mispredicted and unpredicted prediction errors in human electroencephalography. J Neurosci 35:14653–14660. https://doi.org/10.1523/JNEUROSCI.2204-15.2015.

Hsu Y-F, Hamalainen J, Waszak F (2014) Both attention and prediction are necessary for adaptive neuronal tuning in sensory processing. Front Hum Neurosci 8:152.

Hughes HC, Darcey TM, Barkan HI, Williamson PD, Roberts DW, Aslin CH (2001) Responses of human auditory association cortex to the omission of an expected acoustic event. Neuroimage 13:1073–1089. https://doi.org/10.1006/nimg.2001.0766.

Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259. https://doi.org/10.1109/34.730558.

Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin F-H, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW (2004) Human posterior auditory cortex gates novel sounds to consciousness. Proc Natl Acad Sci USA 101:6809–6814. https://doi.org/10.1073/pnas.0303760101.

Jaramillo S, Zador AM (2011) The auditory cortex mediates the perceptual effects of acoustic temporal expectation. Nat Neurosci 14:246–251. https://doi.org/10.1038/nn.2688.

Keller GB, Hahnloser RHR (2009) Neural processing of auditory feedback during vocal practice in a songbird. Nature 457:187–190. https://doi.org/10.1038/nature07467.

Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82:35. https://doi.org/10.1115/1.3662552.

Khouri L, Nelken I (2015) Detecting the unexpected. Curr Opin Neurobiol. https://doi.org/10.1016/j.conb.2015.08.003.

Kiebel SJ, David O, Friston KJ (2006) Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. Neuroimage 30:1273–1284. https://doi.org/10.1016/j.neuroimage.2005.12.055.

Kiebel SJ, Garrido MI, Moran RJ, Friston KJ (2008) Dynamic causal modelling for EEG and MEG. Cogn Neurodyn 2:121–136. https://doi.org/10.1007/s11571-008-9038-0.

Knill DC, Pouget A (2004) The Bayesian brain: The role of uncertainty in neural coding and computation. Trends Neurosci 27:712–719. https://doi.org/10.1016/j.tins.2004.10.007.

Kogo N, Trengove C (2015) Is predictive coding theory articulated enough to be testable? Front Comput Neurosci 9:111. https://doi.org/10.3389/fncom.2015.00111.

Kok P, Rahnev D, Jehee JFM, Lau HC, De Lange FP (2012) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22:2197–2206. https://doi.org/10.1093/cercor/bhr310.

Kok P, Bains LJ, Van Mourik T, Norris DG, De Lange FP (2016) Selective activation of the deep layers of the human primary visual cortex by top-down feedback. Curr Biol 26:371–376. https://doi.org/10.1016/j.cub.2015.12.038.

Lecaignard F, Bertrand O, Gimenez G, Mattout J, Caclin A (2015) Implicit learning of predictable sound sequences modulates human brain responses at different levels of the auditory hierarchy. Front Hum Neurosci 9. https://doi.org/10.3389/fnhum.2015.00505.

Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am A Opt Image Sci Vis 20:1434–1448. https://doi.org/10.1364/JOSAA.20.001434.

Lieder F, Daunizeau J, Garrido MI, Friston KJ, Stephan KE (2013) Modelling trial-by-trial changes in the mismatch negativity. PLoS Comput Biol 9. https://doi.org/10.1371/journal.pcbi.1002911.

López-Caballero F, Zarnowiec K, Escera C (2016) Differential deviant probability effects on two hierarchical levels of the auditory novelty system. Biol Psychol 120:1–9. https://doi.org/10.1016/j.biopsycho.2016.08.001.

May P, Tiitinen H, Ilmoniemi RJ, Nyman G, Taylor JG, Näätänen R (1999) Frequency change detection in human auditory cortex. J Comput Neurosci 6:99–120. https://doi.org/10.1023/A:1008896417606.

May P, Tiitinen H (2010) Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology 47:66–122. https://doi.org/10.1111/j.1469-8986.2009.00856.x.

Meyniel F, Maheu M, Dehaene S (2016) Human inferences about sequences: A minimal transition probability model. PLoS Comput Biol:66–122. https://doi.org/10.1371/journal.pcbi.1005260.

Mittag M, Takegata R, Winkler I (2016) Transitional probabilities are prioritized over stimulus/pattern probabilities in auditory deviance detection: memory basis for predictive sound processing. J Neurosci 36:9572–9579. https://doi.org/10.1523/JNEUROSCI.1041-16.2016.

Moran R, Pinotsis DA, Friston K (2013) Neural masses and fields in dynamic causal modeling. Front Comput Neurosci 7:57. https://doi.org/10.3389/fncom.2013.00057.

Mumford D (1992) On the computational architecture of the neocortex – II The role of cortico-cortical loops. Biol Cybern 66:241–251. https://doi.org/10.1007/BF00198477.

Näätänen R, Gaillard AW, Mäntysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. Acta Psychol 42:313–329. https://doi.org/10.1016/0001-6918(78)90006-9.

Näätänen R, Pakarinen S, Rinne T, Takegata R (2004) The mismatch negativity (MMN): Towards the optimal paradigm. Clin Neurophysiol 115:140–144. https://doi.org/10.1016/j.clinph.2003.04.001.

Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol 118:2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026.

Opitz B, Rinne T, Mecklinger A, von Cramon DY, Schröger E (2002) Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. Neuroimage 15:167–174. https://doi.org/10.1006/nimg.2001.0970.

Paavilainen P, Arajärvi P, Takegata R (2007) Preattentive detection of nonsalient contingencies between auditory features. Neuroreport 18:159–163. https://doi.org/10.1097/WNR.0b013e328010e2ac.

Paavilainen P (2013) The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. Int J Psychophysiol. https://doi.org/10.1016/j.ijpsycho.2013.03.015.

Phillips HN, Blenkmann A, Hughes LE, Bekinschtein TA, Rowe JB (2015) Hierarchical organization of frontotemporal networks for the prediction of stimuli across multiple dimensions. J Neurosci 35:9255–9264. https://doi.org/10.1523/JNEUROSCI.5095-14.2015.

Phillips HN, Blenkmann A, Hughes LE, Kochen S, Bekinschtein TA, Cam CAN, Rowe JB (2016) Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. Cortex 82:192–205. https://doi.org/10.1016/j.cortex.2016.05.001.

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32:3–25. https://doi.org/10.1080/00335558008248231.

Raij T, McEvoy L, Mäkelä JP, Hari R (1997) Human auditory cortex is activated by omissions of auditory stimuli. Brain Res 745:134–143. https://doi.org/10.1016/S0167-8760(97)85548-1.

Rao RPN (2005) Bayesian inference and attentional modulation in the visual cortex. Neuroreport 16:1843–1848. https://doi.org/10.1097/01.wnr.0000183900.92901.fc.

Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87. https://doi.org/10.1038/4580.

Ross S, Hansen NC (2016) Dissociating prediction failure: considerations from music perception. J Neurosci 36:3103–3105. https://doi.org/10.1523/JNEUROSCI.0053-16.2016.

Rubin J, Ulanovsky N, Nelken I, Tishby N (2016) The representation of prediction error in auditory cortex. PLoS Comput Biol 12:e1005058. https://doi.org/10.1371/journal.pcbi.1005058.

Rummell BP, Klee JL, Sigurdsson XT (2016) Attenuation of responses to self-generated sounds in auditory cortical neurons. J Neurosci 36:12010–12026. https://doi.org/10.1523/JNEUROSCI.1564-16.2016.

Saarinen J, Paavilainen P, Schöger E, Tervaniemi M, Näätänen R (1992) Representation of abstract attributes of auditory-stimuli in the human brain. Neuroreport 3:1149–1151.

Sams M, Alho K, Näätänen R (1983) Sequential effects on the ERP in discriminating two stimuli. Biol Psychol 17:41–58. https://doi.org/10.1016/0301-0511(83)90065-0.

SanMiguel I, Saupe K, Schröger E (2013b) I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when". Front Hum Neurosci 7. https://doi.org/10.3389/fnhum.2013.00407.

SanMiguel I, Widmann A, Bendixen A, Trujillo-Barreto N, Schröger E (2013a) Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. J Neurosci 33:8633–8639. https://doi.org/10.1523/JNEUROSCI.5821-12.2013.

Scheeringa R, Koopmans PJ, van Mourik T, Jensen O, Norris DG (2016) The relationship between oscillatory EEG activity and the laminar-specific BOLD signal. Proc Natl Acad Sci USA 113:6761–6766. https://doi.org/10.1073/pnas.1522577113.

Schröger E, Bendixen A, Denham SL, Mill RW, Bohm TM, Winkler I (2014) Predictive regularity representations in violation detection and auditory stream segregation: From conceptual to computational models. Brain Topogr. https://doi.org/10.1007/s10548-013-0334-6.

Schröger E, Marzecová A, SanMiguel I (2015) Attention and prediction in human audition: a lesson from cognitive psychophysiology. Eur J Neurosci 41:641–664. https://doi.org/10.1111/ejn.12816.

Sculthorpe LD, Ouellet DR, Campbell KB (2009) MMN elicitation during natural sleep to violations of an auditory pattern. Brain Res 1290:52–62. https://doi.org/10.1016/j.brainres.2009.06.013.

Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Howard MA, Griffiths TD (2016) Neural signatures of perceptual inference. Elife 5. https://doi.org/10.7554/eLife.11476.

Self MW, Kooijmans RN, Supèr H, Lamme VA, Roelfsema PR (2012) Different glutamate receptors convey feedforward and recurrent processing in macaque V1. Proc Natl Acad Sci USA 109:11031–11036. https://doi.org/10.1073/pnas.1119527109.

Shipp S (2016) Neural elements for predictive coding. Front Psychol 7:1792. https://doi.org/10.3389/FPSYG.2016.01792.

Sohoglu E, Chait M (2016b) Detecting and representing predictable structure during auditory scene analysis. Elife 5:1–17. https://doi.org/10.7554/eLife.19113.

Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. J Neurosci 32:8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.201.

Southwell R, Baumann A, Gal C, Barascud N, Friston K, Chait M (2017) Is predictability salient? A study of attentional capture by auditory patterns. Philos Trans R Soc Lond B Biol Sci.

Spratling MW (2008a) Predictive coding as a model of biased competition in visual attention. Vision Res 48:1391–1408. https://doi.org/10.1016/j.visres.2008.03.009.

Spratling MW (2008b) Reconciling predictive coding and biased competition models of cortical function. Front Comput Neurosci 2:4. https://doi.org/10.3389/neuro.10.004.2008.

Spratling MW (2010) Predictive coding as a model of response properties in cortical area V1. J Neurosci 30:3531–3543. https://doi.org/10.1523/JNEUROSCI.4911-09.2010.

Spratling MW (2015) A review of predictive coding algorithms. Brain Cogn. https://doi.org/10.1016/j.bandc.2015.11.003.

Strauss M, Sitt JD, King J-R, Elbaz M, Azizi L, Buiatti M, Naccache L, van Wassenhove V, Dehaene S (2015) Disruption of hierarchical predictive coding during sleep. Proc Natl Acad Sci USA 112:E1353–E1362. https://doi.org/10.1073/pnas.1501026112.

Szymanski FD, Garcia-Lazaro JA, Schnupp JWH (2009) Current source density profiles of stimulus-specific adaptation in rat auditory cortex. J Neurophysiol 102:1483–1490. https://doi.org/10.1152/jn.00240.2009.

Taaseh N, Yaron A, Nelken I (2011) Stimulus-specific adaptation and deviance detection in the rat auditory cortex. PLoS One 6:e23369. https://doi.org/10.1371/journal.pone.0023369.

Todorovic A, de Lange FP (2012) Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. J Neurosci 32:13389–13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012.

Todorovic A, van Ede F, Maris E, de Lange FP (2011) Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J Neurosci 31:9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011.

Todorovic A, Schoffelen JM, Van Ede F, Maris E, De Lange FP (2015) Temporal expectation and attention jointly modulate auditory oscillatory activity in the beta band. PLoS One 10. https://doi.org/10.1371/journal.pone.0120288.

Uhrig L, Dehaene S, Jarraya B (2014) A hierarchy of responses to auditory regularities in the macaque brain. J Neurosci 34:1127–1132. https://doi.org/10.1523/JNEUROSCI.3165-13.2014.

Uhrig L, Janssen D, Dehaene S, Jarraya B (2016) Cerebral responses to local and global auditory novelty under general anesthesia. Neuroimage 141:326–340. https://doi.org/10.1016/j.neuroimage.2016.08.004.

Ulanovsky N, Las L, Farkas D, Nelken I (2004) Multiple time scales of adaptation in auditory cortex neurons. J Neurosci

24:10440–10453. https://doi.org/10.1523/JNEUROSCI.1905-04.2004.

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6:391–398. https://doi.org/10.1038/nn1032.

Umbricht D, Koller R, Vollenweider FX, Schmid L (2002) Mismatch negativity predicts psychotic experiences induced by NMDA receptor antagonist in healthy volunteers. Biol Psychiatry 51:400–406. https://doi.org/10.1016/S0006-3223(01)01242-2.

van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M-A, Poort J, van der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci USA 111:14332–14341. https://doi.org/10.1073/pnas.1402773111.

van Pelt S, Heil L, Kwisthout J, Ondobaka S, van Rooij I, Bekkering H (2016) Beta- and gamma-band activity reflect predictive coding in the processing of causal events. Soc Cogn Affect Neurosci 11:973–980. https://doi.org/10.1093/scan/nsw017.

Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci USA 108:20754–20759. https://doi.org/10.1073/pnas.1117807108.

Wacongne C, Changeux J-P, Dehaene S (2012) A neuronal model of predictive coding accounting for the mismatch negativity. J Neurosci 32:3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012.

Winer JA (1985) Structure of layer II in cat primary auditory cortex (AI). J Comp Neurol 238:10–37. https://doi.org/10.1002/cne.902380103.

Winkler I, Czigler I (1998) Mismatch negativity: deviance detection or the maintenance of the ''standard''. Neuroreport 9:3809–3813.

Winkler I, Czigler I (2012) Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. Int J Psychophysiol. https://doi.org/10.1016/j.ijpsycho.2011.10.001.

Winkler I, Denham SL, Nelken I (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci. https://doi.org/10.1016/j.tics.2009.09.003.

Winkler I, Schröger E (2015) Auditory perceptual objects as generative models: Setting the stage for communication by sound. Brain Lang 148:1–22. https://doi.org/10.1016/j.bandl.2015.05.003.

Yabe H, Tervaniemi M, Reinikainen K, Näätänen RN (1997) Temporal window of integration revealed by MMN to sound omission. Neuroreport 8:1971–1974. https://doi.org/10.1097/00001756-199705260-00035.

Yaron A, Hershenhoren I, Nelken I (2012) Sensitivity to complex statistical regularities in rat auditory cortex. Neuron 76:603–615. https://doi.org/10.1016/j.neuron.2012.08.025.

Yu AJ (2014) Bayesian Models of Attention. In: The Oxford Handbook of Attention. Oxford: Oxford University Press.

Yu AJ, Dayan P (2005a) Inference, attention, and decision in a Bayesian neural architecture. Adv Neural Inf Process Syst:1577–1584.

Yu AJ, Dayan P (2005b) Uncertainty, neuromodulation, and attention. Neuron 46:681–692. https://doi.org/10.1016/j.neuron.2005.04.026.

Zhao J, Al-Aidroos N, Turk-Browne NB (2013) Attention is spontaneously biased toward regularities. Psychol Sci 24:667–677. https://doi.org/10.1177/0956797612460407.