

Research report

Semantic prediction-errors are context-dependent: An ERP study

Bradley N. Jack^{a,*}, Mike E. Le Pelley^a, Oren Griffiths^{a,b}, David Luque^{a,c}, Thomas J. Whitford^a^a School of Psychology, UNSW Sydney, Sydney, Australia^b School of Psychology, Flinders University, Adelaide, Australia^c Department of Basic Psychology, University of Malaga, Malaga, Spain

HIGHLIGHTS

- Predictive coding assumes an interactive hierarchy of prediction-errors.
- We combined the auditory oddball and visual semantic priming ERP paradigms.
- We found that the N400 was larger when preceded by a standard than a deviant.
- These results are consistent with hierarchical prediction-error processing.

ARTICLE INFO

Keywords:

Prediction

Perception

Semantics

Mismatch negativity (MMN)

N400

Event-related potentials (ERPs)

ABSTRACT

The human brain is an efficient, adaptive, and predictive machine, constructing a generative model of the environment that we then perceive and become conscious of. Here, we show that different types of prediction-errors – the discrepancies between top-down expectations and bottom-up sensory input – are integrated across processing levels and sensory modalities of the cortical hierarchy. We designed a novel, hybrid protocol in which five prediction-establishing sounds were played in rapid succession (e.g., “meow”, “meow”, “meow”, etc.), followed by either a *standard* (e.g., “meow”) or a *deviant* (e.g., “woof”) prime sound, then a visual target word that was either *congruent* or *incongruent* (e.g., “cat” or “dog”) with the prime sound. We found that the deviants elicited a more negative voltage than the standards at about 150 ms – the mismatch negativity (MMN), an event-related potential (ERP) sensitive to low-level perceptual violations – and that the incongruent words elicited a more negative voltage than the congruent words at about 350 ms – the N400, an ERP sensitive to high-level semantic violations. We also found that the N400 was context-dependent: the N400 was larger when the target words were preceded by a standard than a deviant. Our results suggest that perceptual prediction-errors modulate subsequent semantic prediction-errors. We conclude that our results are consistent with one of the most important assumptions of predictive coding theories: hierarchical prediction-error processing.

1. Introduction

The concept of predictive processing has a long tradition in psychology and neuroscience. For example, [Helmholtz \(1867\)](#) proposed that perception is the result of unconscious inferences: logical interpretations of sensory input based on internal representations. Similarly, [Craik \(1943\)](#) discussed cognition in terms of mental models: internal representations of the relationships that exist between the various parts of reality. Recent years have seen a resurgence of theoretical and empirical work framing psychological and neuroscientific domains, such as perception and cognition, in terms of predictive coding ([Clark, 2013](#); [Friston, 2010](#)). The predictive coding theory of brain function postulates that the brain is hierarchically organised, with top-down neural

signals transmitting the prediction – generated by constantly updated models of our environment – and bottom-up neural signals transmitting the prediction-error – the discrepancy between the prediction and sensory input. Each layer in the hierarchy compares the prediction and sensory input, and if they match, then neural processing for the sensory input is reduced (or “explained away”; [Summerfield and Egner, 2009, p. 407](#)); if they do not match, then the resulting prediction-error is used to optimise our model of the environment ([Barlow, 1961](#); [Mumford, 1992](#)). Because predictions prepare sensory and associative cortices to receive sensory input, less cortical activation is required to process predicted sensory input than unpredicted sensory input ([Garrido et al., 2013](#); [Hsu et al., 2015](#); [Schröger et al., 2015](#)).

Hierarchical prediction-error processing is implicated in a

* Corresponding author.

E-mail address: bradley.jack@unsw.edu.au (B.N. Jack).<https://doi.org/10.1016/j.brainres.2018.10.034>

Received 28 May 2018; Received in revised form 29 October 2018; Accepted 30 October 2018

Available online 31 October 2018

0006-8993/ © 2018 Published by Elsevier B.V.

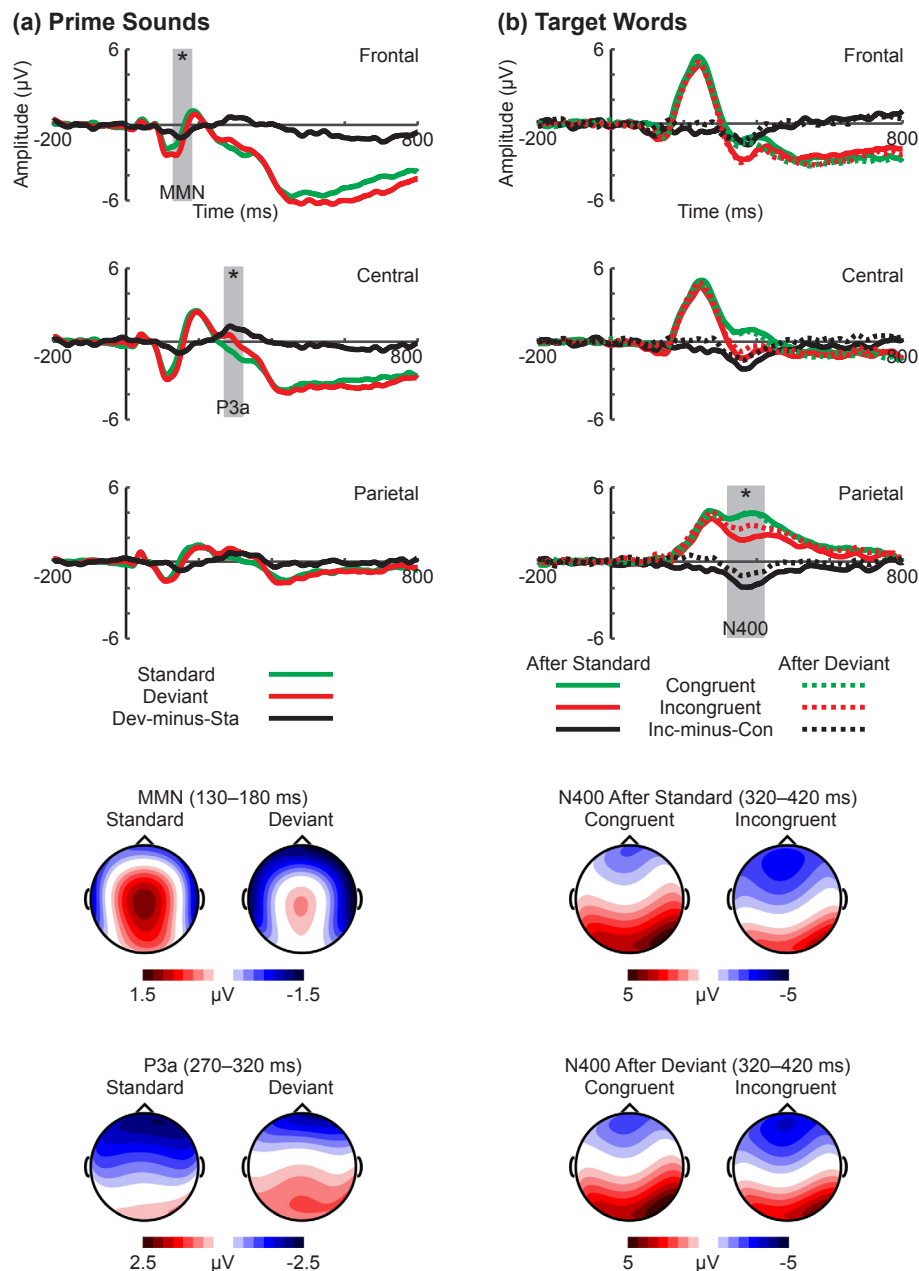


Fig. 1. Results. (a) The graph shows the ERPs and difference waves for the standard and deviant sounds averaged over frontal (F3, Fz, and F4), central (C3, Cz, and C4), and parietal (P3, Pz, and P4) electrodes, showing time (ms) on the x-axis, with 0 indicating the onset of the sound, and voltage (μV) on the y-axis, with positive voltages plotted upwards. The grey bar shows the analysed time-windows, and the voltage maps show the distribution of voltages over the scalp during the time-windows. (b) The graph shows the ERPs and difference waves for the congruent and incongruent words.

“bewildering breadth” (den Ouden et al., 2012, p. 1) of studies in contemporary psychology and neuroscience, including visual (Rao and Ballard, 1999), auditory (Baldeweg, 2006), and somatosensory perception (Akatsuka et al., 2007), as well as in action (Blakemore et al., 1998), language (Kutas and Hillyard, 1980), memory (Erickson and Desimone, 1999), learning (Schultz and Dickinson, 2000), attention (Feldman and Friston, 2010), cognitive control (Alexander and Brown, 2011), and motivational value processing (Hare et al., 2008). Indeed, the ubiquitous nature of prediction-errors has led to the recent, intriguing suggestion that prediction-errors are the brain’s *lingua franca*: the common, computational principles underlying interrelated processes of perception, action, and cognition (Clark, 2013; den Ouden et al., 2012; Friston, 2010). Although there is a plethora of studies demonstrating hierarchical prediction-error responses in the brain (Bendixen et al., 2009; Chennu et al., 2013; Costa-Faidella et al., 2011; Dürschmid et al.,

2016; Garrido et al., 2013; Hsu et al., 2015; Jiang et al., 2016; Kok et al., 2016; Todorovic et al., 2011; Wacongne et al., 2011), according to Egner and Summerfield (2013), there are no studies showing that different types of prediction-errors are integrated across processing levels and sensory modalities of the cortical hierarchy. We sought to address this.

On the assumption of a shared neural network, prediction-error processing ought to modulate subsequent prediction-error processing. To test this, we studied the effects of one event-related potential (ERP) signature of predictive processing, the auditory-evoked mismatch negativity (MMN; Näätänen et al., 2007), on another, the visually-evoked N400 (Federmeier, 2007). The MMN is a negative ERP deflection that is maximal at frontal electrodes and peaks at about 150 ms after sound onset (Näätänen et al., 2007), and it has its neural sources in the auditory cortex, as well as in the frontal and temporal lobes (Alho, 1995;

Deouell, 2007). The MMN occurs when a sequence of repeating *standards* is randomly and unpredictably interrupted by a rare *deviant* (Näätänen et al., 1978); thus, it reflects a low-level perceptual prediction-error (Baldeweg, 2006; Garrido et al., 2009; Wacongne et al., 2012; Winkler, 2007). The N400 is a negative ERP deflection that is maximal at parietal electrodes and peaks at about 350 ms after target word onset (Federmeier, 2007), and it has its neural sources in a wide-range of brain areas, including the frontal, temporal, and parietal lobes (Lau et al., 2008; Van Petten and Luka, 2006). The N400 occurs when a target word is *incongruent* rather than *congruent* with a prime word (Kutas and Hillyard, 1980; Sanquist et al., 1980); thus, it reflects a high-level semantic prediction-error (DeLong et al., 2005; Federmeier, 2007; Kutas and Federmeier, 2011; Rabovsky et al., 2016; Rabovsky and McRae, 2014).

In the present study, we used a novel, hybrid protocol in which an individual trial consisted of five prediction-establishing sounds played in rapid succession (e.g., “meow”, “meow”, “meow”, etc.), followed by either a standard (e.g., “meow”) or a deviant (e.g., “woof”) prime sound, then a visual target word that was either congruent or incongruent (e.g., “cat” or “dog”) with the prime sound. We expected the deviants to elicit a more negative voltage than the standards at about 150 ms – the MMN (Näätänen et al., 2007) – and we expected the incongruent words to elicit a more negative voltage than the congruent words at about 350 ms – the N400 (Federmeier, 2007). Importantly, this design allowed us to compare congruent and incongruent words when preceded by a standard (i.e., *congruent-after-standard* and *incongruent-after-standard* trials) with the same words when preceded by a deviant (i.e., *congruent-after-deviant* and *incongruent-after-deviant* trials). Thus, if a shared neural network is used to compute different types of prediction-errors across processing levels and sensory modalities of the cortical hierarchy (or if prediction-errors are computed independently but are then shared across levels/domains), then the N400 should show differential effects depending on whether it is preceded by a standard or a deviant. If, however, independent neural networks are used, then the N400 should show no modulation.

2. Results

2.1. Behavioural results

At the end of every trial, participants indicated whether the target word was semantically related – specifically, “meow” and “cat” or “woof” and “dog” – or unrelated – specifically, “meow” and “dog” or “woof” and “cat” – to the prime sound. Participants’ responses were close to ceiling for all four trial types: accuracy was 98.9% ($SD = 2.4\%$) in congruent-after-standard trials, 98.6% ($SD = 1.8\%$) in incongruent-after-standard trials, 98.2% ($SD = 2.4\%$) in congruent-after-deviant trials, and 98.7% ($SD = 2.0\%$) in incongruent-after-deviant trials. This suggests that participants generally performed the task as instructed.

2.2. ERP results

Fig. 1a shows the ERPs and difference wave for the standard and deviant prime sounds averaged over frontal (F3, Fz, and F4), central (C3, Cz, and C4), and parietal (P3, Pz, and P4) electrodes, as well as the voltage maps. We identified the MMN as the negative peak in the difference wave at frontal electrodes in the time-window of 130–180 ms, and we analysed this time-window with a one-way ANOVA. We found that the deviants elicited a more negative voltage than the standards, $F(1, 30) = 8.64$, $p = .006$, $\eta_p^2 = 0.22$ – this difference is the MMN (Näätänen et al., 2007).

The MMN is typically followed by the P3a (Escera and Corral, 2007; Horváth et al., 2008), a positive ERP deflection that is maximal at central electrodes and peaks at about 300 ms after sound onset (Polich, 2007). The P3a has its neural sources in the auditory cortex, parietal lobe, and prefrontal regions (Linden, 2005), and it is thought to reflect

the orienting of attention toward a novel stimulus (Escera et al., 2000; Friedman et al., 2001; Horváth et al., 2008; Knight and Scabini, 1998; Schröger and Wolff, 1996). Thus, we identified the P3a as the positive peak in the difference wave at central electrodes in the time-window of 270–320 ms, and we analysed this time-window with a one-way ANOVA. We found that the deviants elicited a more positive voltage than the standards, $F(1, 30) = 4.78$, $p = .037$, $\eta_p^2 = 0.14$ – this difference is the P3a (Polich, 2007).

Fig. 1b shows the ERPs and difference waves for the congruent and incongruent target words averaged over frontal, central, and parietal electrodes, as well as the voltage maps. We identified the N400 as the negative peak in the difference waves at parietal electrodes in the time-window of 320–420 ms, and we analysed this time-window with a repeated-measures ANOVA using the factors target word (congruent, incongruent) and prime sound (standard, deviant). We found that the incongruent words elicited a more negative voltage than the congruent words, $F(1, 30) = 7.79$, $p = .009$, $\eta_p^2 = 0.21$ – this difference is the N400 (Federmeier, 2007). Even though the main effect of prime sound was not significant, $F(1, 30) = 2.29$, $p = .141$, $\eta_p^2 = 0.07$, there was a significant interaction between target word and prime sound, $F(1, 30) = 4.58$, $p = .041$, $\eta_p^2 = 0.13$.

Post-hoc comparisons for the significant interaction found that incongruent-after-standard words elicited a more negative voltage than congruent-after-standard words, $t(30) = 2.91$, $p = .007$, $d = 0.52$, and that incongruent-after-deviant words elicited a more negative voltage than congruent-after-deviant words, $t(30) = 2.17$, $p = .038$, $d = 0.39$. This shows that the target words elicited the N400 regardless of whether the target words were preceded by a standard or a deviant. Post-hoc comparisons also found that incongruent-after-standard words elicited a more negative voltage than incongruent-after-deviant words, $t(30) = 2.49$, $p = .019$, $d = 0.45$, but that there was no difference between congruent-after-standard words and congruent-after-deviant words, $t(30) = 0.12$, $p = .908$, $d = 0.02$. That is, we found that the N400 was larger when the target words were preceded by a standard than a deviant, and that this interaction was driven by the incongruent words.

2.3. Exploratory analyses

We also conducted some exploratory analyses. First, we tested whether the MMN could be observed at locations other than frontal electrodes. Thus, we analysed the mean amplitudes of the standard and deviant ERPs in the MMN time-window at central and parietal electrodes separately with a one-way ANOVA. We found that the deviants elicited a more negative voltage than the standards at central, $F(1, 30) = 13.14$, $p = .001$, $\eta_p^2 = 0.31$, and parietal, $F(1, 30) = 7.88$, $p = .009$, $\eta_p^2 = 0.21$, electrodes. We then tested whether the P3a could be observed at locations other than central electrodes. Thus, we analysed the mean amplitudes of the standard and deviant ERPs in the P3a time-window at frontal and parietal electrodes separately with a one-way ANOVA. We found that there were no differences between standards and deviants at frontal, $F(1, 30) = 1.35$, $p = .254$, $\eta_p^2 = 0.04$, or parietal, $F(1, 30) = 2.88$, $p = .100$, $\eta_p^2 = 0.09$, electrodes.

Next, we tested whether the N400 could be observed at locations other than parietal electrodes. Thus, we analysed the mean amplitudes of the congruent and incongruent ERPs in the N400 time-window at frontal and central electrodes separately with a repeated-measures ANOVA. There was a significant main effect of target word at frontal, $F(1, 30) = 9.18$, $p = .005$, $\eta_p^2 = 0.23$, and central, $F(1, 30) = 11.81$, $p = .002$, $\eta_p^2 = 0.28$, electrodes; however, the main effect of prime sound was not significant at frontal, $F(1, 30) = 1.39$, $p = .248$, $\eta_p^2 = 0.04$, or central, $F(1, 30) = 0.45$, $p = .509$, $\eta_p^2 = 0.02$, electrodes, and the interaction between target word and prime sound was not significant at frontal, $F(1, 30) = 0.17$, $p = .680$, $\eta_p^2 = 0.01$, or central, $F(1, 30) = 1.36$, $p = .252$, $\eta_p^2 = 0.04$, electrodes.

Finally, we have characterised the difference between the standards

and deviants as reflecting a low-level perceptual prediction-error. However, the prime sounds must also contain semantic information; otherwise, why would the congruent and incongruent words show differential modulation at the N400? Moreover, if there is conceptual priming between sounds and words (Griffiths et al., 2016; Orgs et al., 2006; Van Petten and Rheinfelder, 1995), then it is also possible that conceptual priming could occur between two sounds. Thus, we analysed the mean amplitudes of the standard and deviant ERPs at frontal electrodes in the N400 time-window with a one-way ANOVA; the difference between them was not significant, $F(1, 30) = 0.10$, $p = .754$, $\eta_p^2 < 0.01$. We found the same pattern of results at central, $F(1, 30) = 0.62$, $p = .437$, $\eta_p^2 = 0.02$, and parietal, $F(1, 30) = 0.27$, $p = .610$, $\eta_p^2 < 0.01$, electrodes. This suggests that the prime sounds do not elicit a high-level semantic prediction-error during the N400 time-window.

3. Discussion

Our results suggest that perceptual prediction-errors modulate subsequent semantic prediction-errors. This conclusion relies on three important findings. The first is that we found that the deviants elicited a more negative voltage than the standards at about 150 ms – the MMN (Näätänen et al., 2007), an ERP sensitive to low-level perceptual violations (Baldegweg, 2006; Garrido et al., 2009; Wacongne et al., 2012; Winkler, 2007). The second is that we found that the incongruent words elicited a more negative voltage than the congruent words at about 350 ms – the N400 (Federmeier, 2007), an ERP sensitive to high-level semantic violations (DeLong et al., 2005; Federmeier, 2007; Kutas and Federmeier, 2011; Rabovsky et al., 2016; Rabovsky and McRae, 2014). The third is that we found that the N400 was context-dependent: the N400 was larger when the target words were preceded by a standard than a deviant. Our results suggest that different types of prediction-errors are integrated across processing levels and sensory modalities of the cortical hierarchy, and are consistent with one of the most important, yet relatively untested, assumptions of predictive coding: hierarchical prediction-error processing (Clark, 2013; Friston, 2010).

The predictive coding theory of brain function postulates that different types of prediction-errors are integrated across processing levels of the cortical hierarchy to optimise our model of the environment (Clark, 2013; Friston, 2010). Our results are consistent with this view in that the amplitude of the N400 was reduced by the deviant compared to the standard. The most parsimonious explanation for this is that prediction-errors are precision-weighted; that is, as evidence for the generative model accumulates, the amount of prediction-error generated by its violation also increases (Garrido et al., 2013; Hsu et al., 2015; Schröger et al., 2015). Specifically, we suspect that evidence for the model accumulated during the initial, prediction-establishing-sounds phase of the trial, and that this model was tested during the intermediate, prime-sound phase of the trial: the standards increased precision because they fulfilled the model whereas the deviants decreased precision because they violated the model. We also suspect that this decrease in precision reduced (but did not abolish) subsequent semantic prediction-error processing during the later, target-word phase of the trial. This is akin to disrupting a meaningful sentence with incoherent text immediately prior to the final word. Moreover, and consistent with predictive coding, our results suggest that different types of prediction-errors are integrated across sensory modalities of the cortical hierarchy (Talsma, 2015), because the prime sounds and target words were received by different sensory organs.

What factors are known to modulate the N400? It is well-known that the amplitude of the N400 is larger if the target words are infrequent in their everyday usage (Van Petten and Kutas, 1990), have many other physically similar items (Holcomb et al., 2002), or have lower expectancy, or *cloze* (Kutas and Hillyard, 1984). Cloze is typically operationalised as the percentage of participants who would continue a sentence fragment or word pair with a particular word (Taylor, 1953).

However, we can rule out these possible explanations for our results because we used a 2×2 factorial design, because we counter-balanced our stimuli across our design, and because we compared physically-identical prime sounds and target words. That is, the only difference between our two N400s is that one was preceded by a standard whereas the other was preceded by a deviant. Another possible explanation is the distraction of attention. Specifically, it is well-known that the presentation of a deviant triggers involuntary (auditory) attention switching, as indexed by the P3a (Escera et al., 2000; Friedman et al., 2001; Horváth et al., 2008; Knight and Scabini, 1998; Schröger and Wolff, 1996), meaning that participants may be distracted from (visual) attention to the target words. This is important, because the amplitude of the N400 decreases as attention to the target words decreases (Erlbeck et al., 2014; Holcomb and Grainger, 2009; McCarthy and Nobre, 1993; Schacht et al., 2014). Although we cannot rule out that distraction of attention contributed to our results, we think that it is rather unlikely, because the effects of distraction usually last for about 500 ms (Martens and Wyble, 2010; Raymond et al., 1992); this time is considerably shorter than the 2 s period between the onset of the prime sound and the onset of the target word.

If one accepts that perceptual prediction-errors modulate subsequent semantic prediction-errors, and if one accepts that this modulation is driven by the concept of precision-weighting, then it is worth considering where in the brain these different types of prediction-errors might meet and interact. The MMN has its neural sources in the auditory cortex, as well as in the frontal and temporal lobes (Alho, 1995; Deouell, 2007), particularly in the inferior frontal gyrus (IFG) and superior temporal gyrus (STG; Opitz et al., 2002). The N400 has its neural sources in a wide-range of brain areas spanning the frontal, temporal, and parietal lobes (Lau et al., 2008; Van Petten and Luka, 2006), including the IFG and STG (Halgren et al., 2002; Helenius et al., 2002; Maess et al., 2006). Given this, one candidate site for where different types of prediction-errors are integrated (at least in the context of our study) is the IFG and STG; however, future neuroimaging studies with higher spatial resolution than that afforded by EEG are required to test this possibility. Future research may also wish to investigate the integration of different types of prediction-errors that do not share neural sources. Such experiments might allow researchers to investigate the “spreading” of prediction-errors across independent neural networks, similar to what has already been observed in the visual modality for object features (Jiang et al., 2016) and spatial locations (Kok et al., 2016).

In summary, we found that a perceptual prediction-error at a lower level of the cortical hierarchy modulated a subsequent semantic prediction-error at a higher level. Our results are consistent with the view that different types of prediction-errors propagate forward within a shared neural network, eventually meeting and interacting to optimise our model of the environment, and are consistent with the recent, intriguing suggestion that prediction-errors are the brain's *lingua franca*: the common, computational principles underlying interrelated processes of perception, action, and cognition (Clark, 2013; den Ouden et al., 2012; Friston, 2010). Importantly, our results cannot be explained by the opposing view: that the calculation of perceptual and semantic prediction-errors are independent, modular functions that do not interact. We conclude that our study provides crucial evidence for one of the most important, yet relatively untested, assumptions of the predictive coding theory of brain function: hierarchical prediction-error processing.

4. Methods and materials

4.1. Participants

Thirty-four undergraduate students from UNSW Sydney participated in our study for course credit. All participants gave written informed consent prior to the experiment and reported having normal

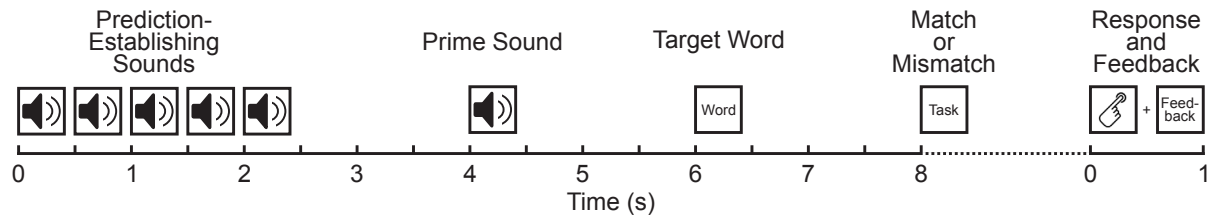
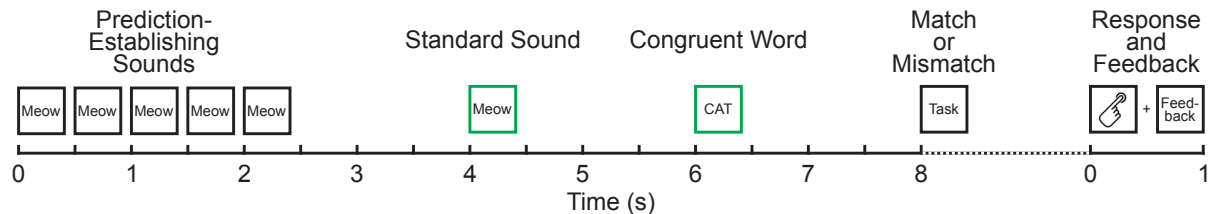
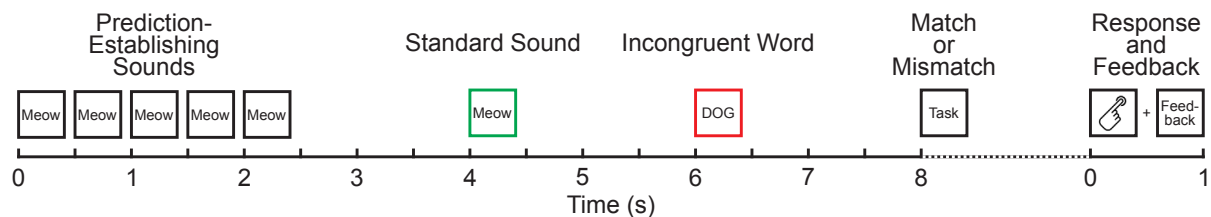
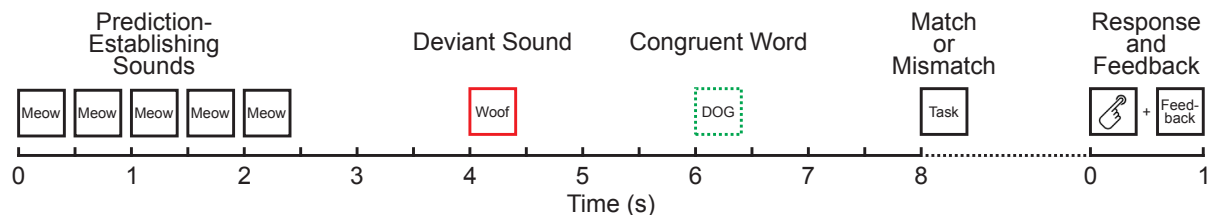
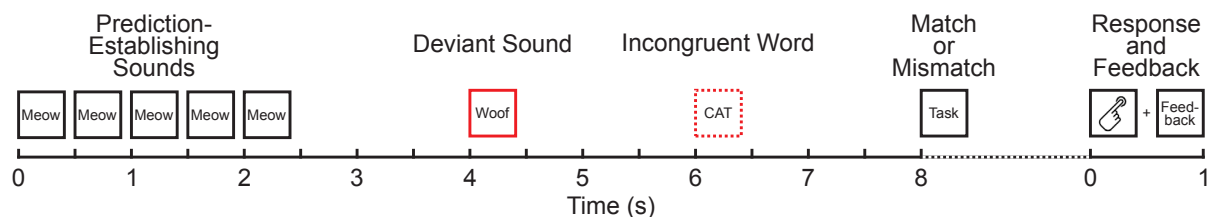
(a) General Procedure**(b) Congruent-after-Standard Trial****(c) Incongruent-after-Standard Trial****(d) Congruent-after-Deviant Trial****(e) Incongruent-after-Deviant Trial**

Fig. 2. Experimental paradigm. (a) An individual trial consisted of five prediction-establishing sounds played in rapid succession, followed by a prime sound that either matched (a standard sound) or mismatched (a deviant sound) the previous sounds, then a visual target word that either matched (a congruent word) or mismatched (an incongruent word) the prime sound. Participants were asked to press one of two buttons on a keyboard indicating whether the target word was semantically related – specifically, “meow” and “cat” or “woof” and “dog” – or unrelated – specifically, “meow” and “dog” or “woof” and “cat” – to the prime sound. Following the response, participants were provided with visual feedback. (b) An example of congruent-after-standard trials, in which the prime sound was a standard and the target word was congruent with the prime. (c) An example of incongruent-after-standard trials, in which the prime sound was a standard and the target word was incongruent with the prime. (d) An example of congruent-after-deviant trials, in which the prime sound was a deviant and the target word was congruent with the prime. (e) An example of incongruent-after-deviant trials, in which the prime sound was a deviant and the target word was incongruent with the prime.

hearing in both ears. Data from three participants were excluded from further analyses due to excessive artefacts in the electroencephalogram (EEG) recording (> 25% of epochs meeting the rejection criteria; see below). Mean age of the remaining participants, 20 of whom were female and 27 of whom were right-handed, was 20 ($SD = 4$) years. The study was approved by UNSW Sydney’s Human Research Ethics Advisory Panel (Psychology) and was conducted in accordance with the ethical standards laid down in the Declaration of Helsinki (World

Medical Association, 2004).

4.2. Apparatus and stimuli

Participants sat in a quiet, dimly-lit room, approximately 1 m in front of a computer monitor (BenQ XL2420T) and directly in front of a low-latency keyboard (Ducky Shine 4) while wearing headphones (AKG K77). Auditory stimuli were two sound samples obtained from the

Internet (from www.freesound.org): specifically, a cat meowing (“meow”) and a dog woofing (“woof”). The sounds were adjusted using Audacity to be 386 ms long and 75 dB SPL, including a 45 ms to 75% peak volume rise time (these sounds are identical to those used in our previous work; see Griffiths et al., 2016). Visual stimuli were written words on a black background. The font type was Helvetica and the font size was 30 points. The words “CAT”, “DOG”, and “MATCH or MISMATCH” were written in white font, the word “CORRECT” was written in green font, and the word “INCORRECT” was written in red font.

4.3. Design and procedure

The experiment consisted of eight blocks, each of which contained 40 trials. An individual trial (see Fig. 2a) consisted of five prediction-establishing sounds played in rapid succession (e.g., “meow”, “meow”, “meow”, etc.), followed by a prime sound that either matched (a standard sound; e.g., “meow”) or mismatched (a deviant sound; e.g., “woof”) the previous sounds (Grau et al., 1998; Ruzzoli et al., 2012, 2016), then a visual target word that either matched (a congruent word) or mismatched (an incongruent word) the prime sound (e.g., “cat” or “dog”; Griffiths et al., 2016; Orgs et al., 2006; Van Petten and Rieffers, 1995). The prediction-establishing sounds were played once every 500 ms from 0 to 2000 ms after trial onset, the prime sound was played 4000 ms after trial onset, and the target word was shown 6000 ms after trial onset for 500 ms. At 8000 ms after trial onset, the prompt “MATCH or MISMATCH” appeared, and participants pressed one of two buttons on a keyboard indicating whether the target word was semantically related – specifically, “meow” and “cat” or “woof” and “dog” – or unrelated – specifically, “meow” and “dog” or “woof” and “cat” – to the prime sound. This task ensured that participants attended to the target words. This is important, because it is well-known that the N400 decreases as attention to the target words decreases (Erlbeck et al., 2014; Holcomb and Grainger, 2009; McCarthy and Nobre, 1993; Schacht et al., 2014). Following the response, visual feedback (e.g., “CORRECT” or “INCORRECT”) was provided for 1000 ms. The time between the offset of the feedback and the start of the next trial varied randomly between 500 and 2500 ms.

The probability of either a set of “meow” or “woof” prediction-establishing sounds was 50% each, the probability of either a standard or a deviant prime sound was 50% each, and the probability of either a congruent or incongruent target word was 50% each. Thus, after collapsing across the different sounds and words, there were four trial-types (see Fig. 2b–e): (1) congruent-after-standard trials, in which the prime sound was a standard and the target word was congruent with the prime; (2) incongruent-after-standard trials, in which the prime sound was a standard and the target word was incongruent with the prime; (3) congruent-after-deviant trials, in which the prime sound was a deviant and the target word was congruent with the prime; and (4) incongruent-after-deviant trials, in which the prime sound was a deviant and the target word was incongruent with the prime. This design allowed us to compare congruent and incongruent words when preceded by a standard with the same words when preceded by a deviant. The order of trials was random and different within each block as well as random and different for each participant.

4.4. EEG acquisition

EEG was recorded with a BioSemi ActiveTwo system from 64 Ag/AgCl active electrodes placed according to the extended 10–20 system (FP1, FPz, FP2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P9, P7, P5, P3, P1, Pz, P2, P4, P6, P8, P10, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, Iz). A vertical electrooculogram (EOG) was recorded by placing an electrode above (we used FP1) and below the left eye; a horizontal EOG was recorded by placing an electrode on the outer canthus of each eye.

We also placed an electrode on the tip of the nose and on each mastoid. The sampling rate of the EEG was 2048 Hz.

4.5. EEG processing and ERP analysis

We processed the EEG data according to the recommendations in Duncan et al. (2009). We re-referenced the data offline to the average of the electrodes on the mastoids, and we filtered the data using a half-amplitude 0.1 to 30 Hz phase-shift free Butterworth filter (48 dB/Oct slope) and a 50 Hz Notch filter. We extracted the epochs from –200 to 800 ms relative to stimulus onset, and we baseline-corrected all epochs to their mean voltage from –200 to 0 ms. We corrected the epochs for eye-blink and movement artefacts using the technique described in Gratton et al. (1983) and Miller et al. (1988), and we excluded all epochs with signals exceeding peak-to-peak amplitudes of 200 μ V at any EEG channel. On average, ERPs were computed from 157 ($SD = 4$) standard, 157 ($SD = 5$) deviant, 76 ($SD = 4$) congruent-after-standard, 75 ($SD = 4$) incongruent-after-standard, 75 ($SD = 4$) congruent-after-deviant, and 75 ($SD = 5$) incongruent-after-deviant epochs. We identified the MMN as the negative peak in the deviant-minus-standard difference wave at frontal electrodes in the time-window of 130–180 ms, and we identified the P3a as the positive peak in the deviant-minus-standard difference wave at central electrodes in the time-window of 270–320 ms. These electrodes and time-windows are consistent with the literature on the MMN and P3a (Duncan et al., 2009; Näätänen et al., 2007; Polich, 2007). We identified the N400 as the negative peak in the incongruent-minus-congruent difference waves at parietal electrodes in the time-window of 320–420 ms. These electrodes are consistent with the literature (Duncan et al., 2009; Federmeier, 2007; Kutas and Federmeier, 2011), and the time-window was selected on the basis of a collapsed localiser: we averaged the congruent and incongruent words separately, we computed an incongruent-minus-congruent difference wave, and we used the collapsed waveforms to define the time-window for the non-collapsed waveforms (Luck and Gaspelin, 2017).

Acknowledgements

This work was supported by the Australian Research Council (DP140104394, DP170103094, FT100100260, and DE150100667) and by the National Health and Medical Research Council of Australia (APP1090507).

Conflict of interest

The authors declare no competing financial interests.

References

- Akatsuka, K., Wasaka, T., Nakata, H., Kida, T., Kakigi, R., 2007. The effect of stimulus probability on the somatosensory mismatch field. *Exp. Brain Res.* 181, 607–614.
- Alexander, W.H., Brown, J.W., 2011. Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* 14, 1338–1344.
- Alho, K., 1995. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear.* 16, 38–51.
- Baldeweg, T., 2006. Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn. Sci.* 10, 93–94.
- Barlow, H.B., 1961. Possible principles underlying the transformation of sensory messages. In: Rosenblith, W.A. (Ed.), *Sensory Communication*. MIT Press, Cambridge, MA.
- 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.
- Blakemore, S.J., Goodbody, S.J., Wolpert, D.M., 1998. Predicting the consequences of our own actions: the role of sensorimotor context estimation. *J. Neurosci.* 18, 7511–7518.
- Chennu, S., Noreika, V., Gueorguiev, D., Blenkman, A., Kochen, S., Ibáñez, A., Bekinschtein, T.A., 2013. Expectation and attention in hierarchical auditory prediction. *J. Neurosci.* 33, 11194–11205.
- Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204.
- 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.
- Craik, K.J.W., 1943. *The Nature of Explanation*. Cambridge University Press, Cambridge, UK.
- DeLong, K.A., Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121.
- den Ouden, H.E.M., Kok, P., de Lange, F.P., 2012. How prediction errors shape perception, attention, and motivation. *Front. Psychol.* 3 (548), 1–12.
- Deouell, L.Y., 2007. The frontal generator of the mismatch negativity revisited. *J. Psychophysiol.* 21, 188–203.
- Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Näätänen, R., Van Petten, C., 2009. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clin. Neurophysiol.* 120, 1883–1908.
- Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H.-J., Knight, R.T., 2016. Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6755–6760.
- Egner, T., Summerfield, C., 2013. Grounding predictive coding models in empirical neuroscience research. *Behav. Brain Sci.* 36, 210–211.
- Erickson, C.A., Desimone, R., 1999. Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J. Neurosci.* 19, 10404–10416.
- Erlbeck, H., Kübler, A., Kotchoubey, B., Vesper, S., 2014. Task instructions modulate the attentional mode affecting the auditory MMN and the semantic N400. *Front. Hum. Neurosci.* 8 (654), 1–16.
- Escera, C., Alho, K., Schröger, E., Winkler, I., 2000. Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol. Neuro-Otol.* 5, 151–166.
- Escera, C., Corral, M.J., 2007. Role of mismatch negativity and novelty-P3 in involuntary auditory attention. *J. Psychophysiol.* 21, 251–264.
- Federmeier, K.D., 2007. Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505.
- Feldman, H., Friston, K.J., 2010. Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4 (215), 1–23.
- Friedman, D., Cycowicz, Y.M., Gaeta, H., 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci. Biobehav. Rev.* 25, 355–373.
- Friston, K., 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.
- Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463.
- Garrido, M.I., Sahani, M., Dolan, R.J., 2013. Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS Comput. Biol.* 9 (e1002999), 1–10.
- Gratton, G., Coles, M.G.H., Donchin, E., 1983. A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484.
- Grau, C., Escera, C., Yago, E., Polo, M.D., 1998. Mismatch negativity and auditory sensory memory evaluation: a new faster paradigm. *NeuroReport* 9, 2451–2456.
- Griffiths, O., Le Pelley, M.E., Jack, B.N., Luque, D., Whitford, T.J., 2016. Cross-modal symbolic processing can elicit either an N2 or a protracted N2/N400 response. *Psychophysiology* 53, 1044–1053.
- Halgren, E., Dhond, R.P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J.D., Dale, A.M., 2002. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage* 17, 1101–1116.
- Hare, T.A., O'Doherty, J., Camerer, C.F., Schultz, W., Rangel, A., 2008. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* 28, 5623–5630.
- Helenius, P., Salmelin, R., Service, E., Connolly, J.F., Leinonen, S., Lyytinen, H., 2002. Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *J. Neurosci.* 22, 2936–2944.
- Helmholtz, H.V., 1867. *Handbuch der physiologischen optik*. Voss, Leipzig, Germany.
- Holcomb, P.J., Grainger, J., 2009. ERP effects of short interval masked associative and repetition priming. *J. Neurolinguistics* 22, 301–312.
- Holcomb, P.J., Grainger, J., O'Rourke, T., 2002. An electrophysiology study of the effects of orthographic neighbourhood size on printed word perception. *J. Cognit. Neurosci.* 14, 938–950.
- Horváth, J., Winkler, I., Bendixen, A., 2008. Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol. Psychol.* 79, 139–147.
- Hsu, Y.-F., Bars, S.L., Hämäläinen, J.A., Waszak, F., 2015. Distinctive representation of mispredicted and unpredicted prediction errors in human electroencephalography. *J. Neurosci.* 35, 14653–14660.
- Jiang, J., Summerfield, C., Egner, T., 2016. Visual prediction error spreads across object features in human visual cortex. *J. Neurosci.* 36, 12746–12763.
- Knight, R.T., Scabini, D., 1998. Anatomic bases of event-related potentials and their relation to novelty detection in humans. *J. Clin. Neurophysiol.* 15, 3–13.
- Kok, P., van Lieshout, L.L.F., de Lange, F.P., 2016. Local expectation violations result in global activity gain in primary visual cortex. *Sci. Rep.* 6 (37706), 1–10.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647.
- Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.
- Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.
- Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933.
- Linden, D.E.J., 2005. The P300: where in the brain is it produced and what does it tell us? *Neuroscientist* 11, 563–576.
- Luck, S.J., Gaspelin, N., 2017. How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology* 54, 146–157.
- Maess, B., Herrmann, C.S., Hahne, A., Nakamura, A., Friederici, A.D., 2006. Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Res.* 1096, 163–172.
- Martens, S., Wyble, B., 2010. The attentional blink: past, present, and future of a blind spot in perceptual awareness. *Neurosci. Biobehav. Rev.* 34, 947–957.
- McCarthy, G., Nobre, A.C., 1993. Modulation of semantic processing by spatial selective attention. *Electroencephalogr. Clin. Neurophysiol.* 88, 210–219.
- Miller, G.A., Gratton, G., Yee, C.M., 1988. Generalized implementation of an eye movement correction procedure. *Psychophysiology* 25, 241–243.
- Mumford, D., 1992. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern.* 251, 241–251.
- Näätänen, R., Gaillard, A.W.K., Mäntysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329.
- 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.
- Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D.Y., Schröger, E., 2002. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *NeuroImage* 15, 167–174.
- Orgs, G., Lange, L., Dombrowski, J.H., Heil, M., 2006. Conceptual priming for environmental sounds and words: an ERP study. *Brain Cogn.* 62, 267–272.
- Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148.
- Rabovsky, M., McRae, K., 2014. Simulating the N400 ERP component as semantic network error: insights from a feature-based connectionist attractor model of word meaning. *Cognition* 132, 68–89.
- Rabovsky, M., Hansen, S.S., McClelland, J.L., 2016. N400 amplitudes reflect change in a probabilistic representation of meaning: evidence from a connectionist model. In: Papafragou, A., Grodner, D., Miran, D., Trueswell, J.C. (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Cognitive Science Society, Austin, TX.
- Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87.
- Raymond, J.E., Shapiro, K.L., Arnell, K.M., 1992. Temporal suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol.: Human Perception Performance* 18, 849–860.
- Ruzzoli, M., Pirulli, C., Brignani, D., Maioli, C., Miniussi, C., 2012. Sensory memory during physiological aging indexed by mismatch negativity (MMN). *Neurobiol. Aging* 33 (625), e21–e30.
- Ruzzoli, M., Pirulli, C., Mazza, V., Miniussi, C., Brignani, D., 2016. The mismatch negativity as an index of cognitive decline for the early detection of Alzheimer's disease. *Sci. Rep.* 6 (33167), 1–11.
- Sanquist, T.F., Rohrbaugh, J.W., Syndulko, K., Lindsay, D.B., 1980. Electrocortical signs of levels of processing: perceptual analysis and recognition memory. *Psychophysiology* 17, 568–576.
- Schacht, A., Sommer, W., Shmuelovitch, O., Martínez, P.C., Martín-Loeches, M., 2014. Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS One* 9 (e91226), 1–7.
- Schröger, E., Wolff, C., 1996. Mismatch response of the human brain to changes in sound location. *NeuroReport* 7, 3005–3008.
- Schröger, E., Marzecová, A., SanMiguel, I., 2015. Attention and prediction in human audition: a lesson from cognitive psychophysiology. *Eur. J. Neurosci.* 41, 641–664.
- Schultz, W., Dickinson, A., 2000. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* 23, 473–500.
- Summerfield, C., Egner, T., 2009. Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* 13, 403–409.
- Talsma, D., 2015. Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Front. Integr. Neurosci.* 9 (19), 1–13.
- Taylor, W.L., 1953. Cloze procedure: a new tool for measuring readability. *J. Bull.* 30, 415–433.
- Todorovic, A., van Ede, F., Maris, E., de Lange, F.P., 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J. Neurosci.* 31, 9118–9123.
- Van Petten, C., Kutas, M., 1990. Interactions between sentence context and word frequency in event-related brain potentials. *Memory Cogn.* 18, 380–393.
- Van Petten, C., Luka, B.J., 2006. Neural localization of semantic context effects in electrophysiological and hemodynamic studies. *Brain Lang.* 97, 279–293.
- Van Petten, C., Rieffelder, H., 1995. Conceptual relationships between spoken words and environmental sounds: event-related brain potential measures. *Neuropsychologia* 33, 485–508.
- 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. U.S.A.* 108, 20754–20759.
- Wacongne, C., Changeux, J.-P., Dehaene, S., 2012. A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.* 32, 3665–3678.
- Winkler, I., 2007. Interpreting the mismatch negativity. *J. Psychophysiol.* 21, 147–163.
- World Medical Association, 2004. *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*.