Attention allocation in a language with post-focal prominences

Caterina Ventura[a], Martine Grice[b], Michelina Savino[c], Diana Kolev[a], Ingmar Brilmayer[a] and Petra B. Schumacher[a]

Accentuation influences selective attention and the depth of semantic processing during online speech comprehension. We investigated the processing of semantically congruent and incongruent words in a language that presents cues to prosodic prominences in the region of the utterance occurring after the focussed information (the post-focal region). This language is Italian, in particular the variety spoken in Bari. In this variety, questions have a compressed, post-focal accent, whereas in statements there is a low-level pitch in this position. Using event-related potentials, we investigated the processing of congruent and incongruent target words with two prosodic realizations (focussed with accentuation, post-focal realization) and in two-sentence modalities (statement, question). Results indicate an N400 congruence effect that was modulated by position (focal, post-focal) and modality (statement, question): processing was deeper for questions in narrow focus than in post-focal position, while statements showed similar pronounced N400 effects across positions. The attenuated N400 difference for post-focal targets in questions was accompanied by a more enhanced late positivity when they were incongruent, indicating that attentional resources are allocated during updating of speech act information. *NeuroReport* 31: 624–628 Copyright © 2020 The Author(s). Published by Wolters Kluwer Health, Inc.

NeuroReport 2020, **31**:624–628

Keywords: attention, Italian, prominence, prosody, rERPs, sentence modality

[a]Department of German Language and Literature I – Linguistics, [b]Department of Linguistics, Institute of Linguistics – Phonetics, University of Cologne, Cologne, Germany and [c]Department of Education, Psychology, Communication, University of Bari 'Aldo Moro', Bari, Italy

Correspondence to Caterina Ventura, MA, Department of German Language and Literature I – Linguistics, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany
Tel: +49 0221 470 89914; fax: +49 0221 470 5107;
e-mail: cventura@uni-koeln.de

Received 2 March 2020 **Accepted** 22 March 2020

Introduction

The language processing system attends to only selected parts of the information influx with which it is confronted. This is possible because language provides cues for attention orienting, for example, prosodic prominences (accents) on focussed words. We investigate how far attention is allocated to words that are prosodically marked as most informative (narrow focus) and to words occurring after the focus (post-focal position) that are less informative. We examine how different prosodic cues guide selective attention and lead to deeper processing [1], such that semantic incongruence is detected.

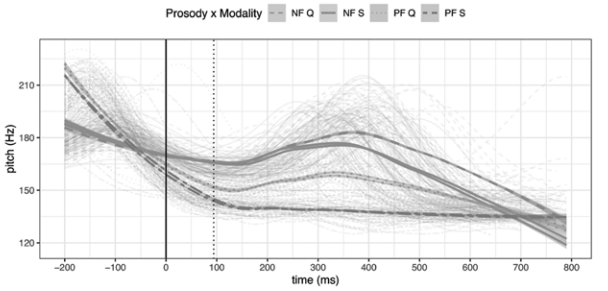
Event-related potential (ERP) studies [2–5] have shown that prosodic marking of information is processed online. Higher degrees of prosodic prominence, characterized by major excursions in fundamental frequency (F0), increased duration and energy, lead to the allocation of selective attention during online speech processing [6,7]. Conversely, the absence of acoustic cues to prominence (deaccentuation) leads to shallow processing. Additionally, Wang *et al.* [8] suggest that in Dutch an early focal accent draws attention away from the following part of the utterance, which is usually deaccented, causing semantic information in post-focal position to be processed in a shallow way [9]. The present experiment tests whether the presence of fine-grained cues to prominence in this position could play a role in reorienting attention towards this part of the utterance. We investigate the Italian variety spoken in Bari, in which polar questions (Fig. 1) regularly bear an F0 rise in post-focal position, signaling sentence modality. In this case, the mere presence of prominence might preclude shallow processing of post-focal material.

We investigate online comprehension of semantically congruent and incongruent words realized in utterances in two-sentence modalities, questions and statements, and two prosodic conditions, narrow focus and post-focal. In Bari Italian, words in narrow focus have an accent (high prominence) irrespective of modality. Words in post-focal position in statements are realized with a flat F0 shape, whereas words in the same position in questions bear a compressed rising-falling F0 contour, presenting enhanced cues to prominence in comparison to those in statements. Note that this F0 shape is crucial for questioning.

0959-4965 Copyright © 2020 The Author(s). Published by Wolters Kluwer Health, Inc. DOI: 10.1097/WNR.0000000000001453

Fig. 1

	Per aprire la porta,
	To open the door,
NF Q	bisogna girare la **MANIGLIA** (congruent) vs. **FRITTATA** (incongruent)?
PF Q	bisogna GIRARE la **maniglia** (congruent) vs. **frittata** (incongruent)?
	do you have to turn the handle vs. omelet?
	Per arieggiare la stanza,
	To ventilate the room,
NF S	bisogna aprire la **FINESTRA** (congruent) vs. **LATTINA** (incongruent),
PF S	bisogna APRIRE la **maniglia** (congruent) vs. **lattina** (incongruent),
	you need to open the window vs. can.



Left panel: Examples of stimuli per condition and their English translations. Right panel: F0 contours of critical words. Thicker lines represent the means of the contours. Time in milliseconds starts from 200 ms before the determiner (*la*) of the critical word. Solid vertical line at 0 ms indicates the onset of the determiner, dotted vertical line at 100 ms indicates the onset of the critical word.

Generally, we predict an effect of semantic congruence on the N400 ERP component, known to be sensitive to semantic mismatches [10]. The N400 is a negative deflection that peaks around 400 ms after stimulus onset. Its amplitude is more enhanced for unexpected vs. expected information; thus, a more pronounced N400 for incongruent over congruent target words is expected. Since the ERP amplitude elicited by attended stimuli is larger compared to unattended ones [1], we predict the relative difference of congruence to be further modulated by prosody, which is expected to orient attentional processes [6]. We predict a greater difference in the N400-amplitude between incongruent and congruent conditions when processing of the incongruence is facilitated by prominence: accent in narrow focus will facilitate processing of the incongruence (large N400 difference between incongruent and congruent targets); this will not differ between statements and questions. The hypothesis for words in post-focal position differs for each sentence modality: for statements, the missing acoustic cues to prominence should lead to shallow processing of the word; for questions, the acoustic cues to modality should prevent such shallow processing (larger N400 difference in questions compared to statements). In addition, signal-driven attention orienting may evoke a positive ERP deflection and result in updating of mental representations [11,12]. Alternatively, since Italian tends to place prominent information in sentence-final position, attention may be allocated to this position by default. In this case, the incongruence would instead be processed deeply (large N400 difference).

Method

Thirty-two right-handed, monolingual native speakers of Bari Italian (seven men) participated in the ERP study after giving written informed consent in accordance with the Declaration of Helsinki. They were students from the University of Bari (mean age: 22.72; range: 19–32). None of them self-reported any auditory, visual or neurological impairment.

The study involved three factors with two levels each: (1) prosody: narrow focus (NF) or post-focal (PF); (2) semantic congruence: congruent (C+) or incongruent (C–) within the utterance; (3) modality: realization of the sentences as statement (S) or question (Q). Examples of utterances in the different conditions are shown in Fig. 1 (left), where the critical word is indicated in bold and the narrow focus in capital letters.

Stimuli were recorded by a Bari Italian trained female phonetician in a sound-attenuated cabin (44 100 Hz sampling rate, 16-bit resolution). To ensure segmental comparability, all critical words were trisyllabic with lexical stress on the penultimate syllable. Figure 1 (right) shows the mean F0 contours of all target words and the individual F0 contours of all targets superimposed on each other. Fillers were also produced, which were realized either in broad focus (default condition for hearing utterances without context) or with a narrow focus on a noncritical word.

Each experimental session contained 360 trials. They involved 240 critical items (60 lexically different sentences × 2 intonation contours × 2 sentence modalities) plus 120 filler items. Critical and filler items were pseudo-randomized.

Participants performed a word recognition task after each auditory stimulus. Probe words were equally selected from the first part of the sentence, the inflected verb (*bisogna*, one needs) or the infinite verb (e.g. *girare*, to turn); incongruences were never addressed directly. The expected yes/no responses were equally distributed across stimuli and conditions. The stimuli were presented in eight blocks with pauses in-between. Each block contained either questions or statements to discourage participants from focusing on sentence modality. To prevent repetition effects, test sentences with the same lexical material were assigned to different experimental blocks. To avoid systematic order effects, experimental stimuli were presented in different condition sequences across the blocks.

Participants were instructed to focus on a fixation star on the monitor while the auditory stimuli were presented over loudspeakers. The electroencephalogram (EEG) was recorded and digitized (500 Hz) by means of 24 Ag/AgCl electrodes. EEGs were referenced online to the left mastoid (ground: AFz). To control for eye-movement artifacts, the electrooculogram was recorded by electrode pairs placed above and below the participant's left eye and at the outer canthus of each eye. Impedances were kept below 5 kΩ.

Data were analyzed using a Python3 implementation of MNE-Python version 0.19 [13]. EEGs were re-referenced offline to linked mastoids. Since the auditory signal shows differences between conditions prior to critical word onset, the EEG was filtered with a 0.3–45 Hz filter to counter prestimulus-evoked activity [14]. Eye artifacts were automatically detected and portions of raw data containing blinks were excluded from further analysis. The data were epoched from –200 to 1000 ms relative to determiner-onset of the target and resampled to 100 Hz for further analysis. Trials with false or time-out responses to the task were excluded from the analysis (7%).

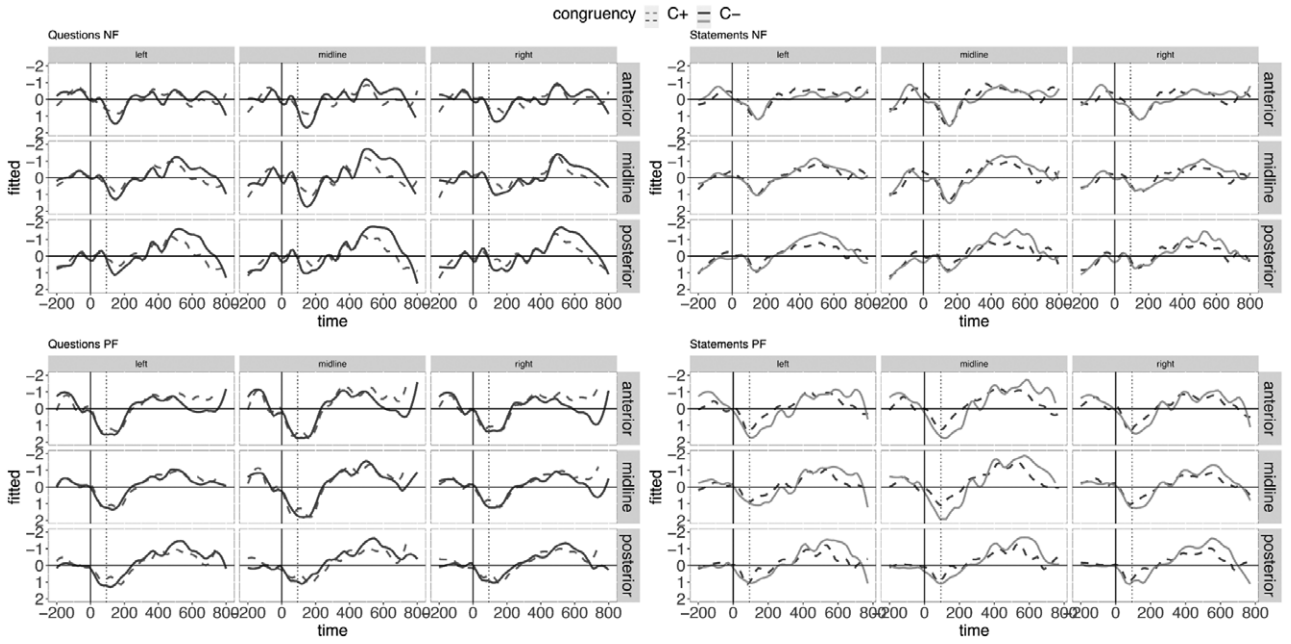
Due to design-immanent differences in the acoustic properties across the critical conditions, we performed a regression-based ERP (rERP) analysis [15,16] using the lm() function in R [17]. We calculated linear models by subject, channel and sample (i.e. for time points in 10 ms steps) with factors PROSODY (narrow focus, post-focal), CONGRUENCE (C+, C–) and MODALITY (Q, S) as well as PITCH (Hertz, continuous) and PERIODIC ENERGY (dB, continuous) [18]. Pitch and periodic energy were extracted for periods of 10 ms from the audio files using PRAAT [19] from –200 to 1000 ms relative to determiner-onset (the sentence-final critical words ended at 800 ms; the interval from 800 to 1000 ms was filled with silence). We calculated linear mixed-effect models using the lmer() function from the lme4 package [20] for R with mean fitted values in the windows 400–600 ms and 600–800 ms. Models included three fixed factors PROSODY, CONGRUENCE and MODALITY and two continuous factors SAGITTALITY and LATERALITY based on the planar coordinates of the standard BESA system. The models assumed random intercepts for subjects as well as by subject random slopes for the effect of PROSODY.

Results

Figure 2 depicts the grand-averaged rERPs (fitted microvolt values) for the eight experimental conditions. We time-locked the rERPs to the determiner because determiner + noun constitutes a prosodic word; yet, the determiner does not provide information about the semantic (in)congruence. In the window from 400 to 600 ms, all focal conditions show a more pronounced negative potential for the incongruent words (solid lines) over posterior regions. Post-focal targets in questions furthermore show a more pronounced positivity between 600 and 800 ms for incongruent words over anterior sites.

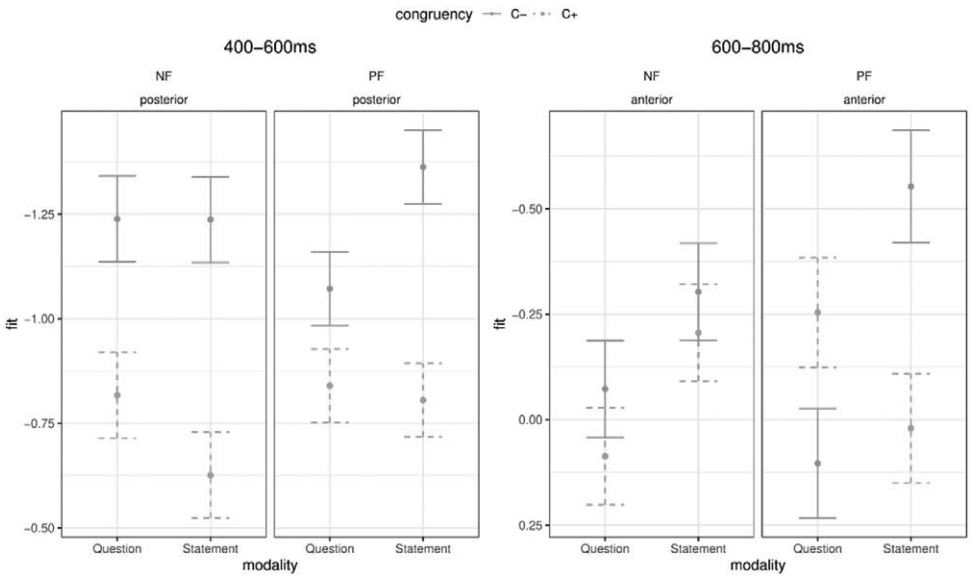
Statistical analysis for the 400–600 ms window registered an interaction of prosody, modality, congruence

Fig. 2



Grand-average rERPs (negativity plotted upwards). First line shows results for NF, second line for PF. Left panels show questions, right panels shows statements. Dashed lines indicate C+ condition, solid line C–. Electrodes are grouped by centrality, laterality and sagittality. Time course on horizontal axis spans from 200 ms before until 800 ms after the onset of the determiner (= vertical solid bar). Vertical dotted bar indicates the onset of the critical noun.

Fig. 3



Significant congruence effects of rERPs for the time window from 400 to 600 ms (left panel) and from 600 to 800 ms (right panel). Error bars represent 83% confidence intervals. Negativity is plotted upwards. Dotted lines=C+, solid lines=C–.

and sagittality ($\chi^2 = 21.85$, $P < 0.0001$). Contrast obtained with emmeans() [21] function show that for incongruent words in narrow focus for both questions and statements rERPs deviate from congruent words in a negative direction in the posterior regions (congruence effect for narrow focus-Q: $\beta = -0.42$, $P < 0.0001$; narrow focus-S: $\beta = -0.61$, $P < 0.0001$). The same holds for post-focal conditions (post-focal-Q: $\beta = -0.23$, $P < 0.0001$; post-focal-S: $\beta = -0.56$, $P < 0.0001$). The pairwise comparison (Fig. 3) shows that the congruence effect is higher in statements than in questions for both prosodic conditions. In questions, the effect of congruence is higher for narrow focus than post-focal whereas in statements the effect of congruence does not differ between the conditions.

Statistical analysis for the time window 600–800 ms registered an interaction of prosody, modality, congruence and sagittality ($\chi^2 = 16.47$, $P < 0.001$). Contrast obtained with emmeans() function shows that in questions the incongruent critical word in post-focal exhibits a more pronounced positivity over anterior sites ($\beta = 0.36$, $P < 0.0001$) (Fig. 3). The stimuli, data and scripts for the analysis are available at https://osf.io/zepfa/.

Discussion

In this EEG study, we tested the hypothesis whether allocation of resources (triggered by prosodic prominence) modulates processing depth. We investigated the processing of semantic incongruencies presented with different degrees of prosodic prominence in the absence of context with the aim of identifying signal-driven processes.

Results revealed that semantically incongruent words evoked a more pronounced N400 relative to congruent words for all contrasts. The magnitude of this effect was modulated by sentence modality and prosody. In questions, the effect was larger in narrow focus than in post-focal position. Since the narrow focus accent is higher in prominence than the post-focal one, our results are consistent with previous studies [6–8]: a strong prosodic prominence results in more attention allocation to the accented stimulus, reflected in deeper processing of the incongruence. However, in statements, the effect of the incongruence did not differ when comparing focal and post-focal positions. In post-focal position, the incongruence seems to be more deeply processed in statements than in questions: the negativity for incongruent words increases in statements. Yet, questions in post-focal condition engendered an additional late positive deflection for incongruent over congruent items.

In the present study, the N400 congruence effect reflects the deep processing of semantic information. The relative difference between congruent and incongruent targets suggests that the N400 effect is stable in both modalities in narrow focus, indicating that prosodic marking of the target attracts attention preventing shallow semantic processing. Incongruences in statements in post-focal position also engendered a pronounced N400 effect. This may indicate that, given the high variability in degree of prominence for this position in Italian, attention could be directed there by default, drawn by top-down expectations of prominence (i.e. the tendency to place prominent words in final position). This could,

in turn, prevent shallow processing of the incongruence. In contrast, the N400 effect in post-focal position was least pronounced for questions. In this condition, the compressed rising-falling accent may draw attention to the sentence modality and its corresponding illocutionary force rather than on semantic congruence. This suggests that the compressed accent on the post-focal word prioritizes the speech act of requesting at the expense of semantic processing. Illocutionary information must be encoded in the mental representation, which results in updating mechanisms, and may initiate an action (answering) reflected in a later positivity [4,5].

This late positivity emerged between 600 and 800 ms for incongruent over congruent targets. This suggests that attentional resources are allocated to this condition after all and that the reorienting towards this part of the utterance is modulated by congruence. Modality-specific prosodic cues thus lead to signal-induced attention allocation.

Conclusion

This study reveals that in languages that place prominences in post-focal position, attention can be drawn to focal and post-focal information. The rERPs show N400 congruence effects that are attenuated by modality-specific demands in post-focal position. Modality-induced mechanisms, in turn, give rise to a late positivity. In focal position, attention is entirely allocated to the target word, rendering a pronounced N400 congruence, suggesting enhanced precision through the interplay of attention and prediction [22]. The same applies to Italian words in post-focal position, which is the preferred location for prominent information in this language. Crucially, post-focal information is further modulated by modality. Cues that convey a specific illocutionary function (i.e. request) reduce the attention allocated to semantic processing and instead consume resources for updating of sentence modality information. Our findings thus indicate that different prosodic cues (to focus or for modality) influence selective attention in discrete ways.

Acknowledgements

We would like to thank Brita Rietdorf, Annalisa Palmisano and Mariagrazia Violante for invaluable help in EEG data acquisition. We are also very grateful to Mario Refice for generous help in setting up the EEG lab in the Department of Education, Psychology, Communication of the Bari University, and Davide Rivolta for support in recruiting experiment participants. Lastly, we would like to thank Aviad Albert for calculating periodic energy.

This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252 'Prominence in Language' in the project A01 'Intonation and attention orienting: Neurophysiological and behavioural correlates' at the University of Cologne, Germany.

Conflicts of interest

There are no conflicts of interest.

References

1 Luck SJ, Kappenman ES. ERP components and selective attention. In: Luck SJ, Kappenman ES, editors. *The Oxford Handbook of Event-Related Potential Components*. New York: Oxford University Press; 2012. pp. 295–327.
2 Heim S, Alter K. Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiol Exp* 2006; **66**:55.
3 Toepel U, Pannekamp A, Alter K. Catching the news: processing strategies in listening to dialogs as measured by ERPs. *Behav Brain Funct* 2007; **3**:53.
4 Schumacher PB, Baumann S. Pitch accent type affects the N400 during referential processing. *Neuroreport* 2010; **21**:618–622.
5 Baumann S, Schumacher PB. (De-)accentuation and the process of information status: evidence from event-related brain potentials. *Lang Speech* 2012; **55**:361–381.
6 Li XQ, Ren GQ. How and when accentuation influences temporally selective attention and subsequent semantic processing during on-line spoken language comprehension: an ERP study. *Neuropsychologia* 2012; **50**:1882–1894.
7 Kristensen LB, Wang L, Petersson KM, Hagoort P. The interface between language and attention: prosodic focus marking recruits a general attention network in spoken language comprehension. *Cereb Cortex* 2013; **23**:1836–1848.
8 Wang L, Bastiaansen M, Yang Y, Hagoort P. The influence of information structure on the depth of semantic processing: how focus and pitch accent determine the size of the N400 effect. *Neuropsychologia* 2011; **49**:813–820.
9 Sanford GM, Sanford AJ, Molle J, Emmott C. Shallow processing and attention capture in written and spoken discourse. *Discourse Processes* 2006; **42**:109–130.
10 Kutas M, Federmeier KD. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol* 2011; **62**:621–647.
11 Picton TW. The P300 wave of the human event-related potential. *J Clin Neurophysiol* 1992; **9**:456–479.
12 Polich J. Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 2007; **118**:2128–2148.
13 Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, *et al*. MEG and EEG data analysis with MNE-Python. *Front Neurosci* 2013; **7**:267.
14 Maess B, Schröger E, Widmann A. High-pass filters and baseline correction in M/EEG analysis. Commentary on: "How inappropriate high-pass filters can produce artefacts and incorrect conclusions in ERP studies of language and cognition". *J Neurosci Methods* 2016; **266**:164–165.
15 Hauk O, Davis MH, Ford M, Pulvermüller F, Marslen-Wilson WD. The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage* 2006; **30**:1383–1400.
16 Smith NJ, Kutas M. Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology* 2015; **52**:157–168.
17 R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. URL https://www.R-project.org/. [Accessed 20 December 2019]
18 Albert A, Cangemi F, Grice M. Using periodic energy to enrich acoustic representations of pitch in speech: a demonstration. *Proceedings Speech Prosody* 2018; **9**:13–16.
19 Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version 6.0.43. 2018. URL http://www.praat.org/. [Accessed 15 December 2019]
20 Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed effects models using lme4. *J Stat Softw* 2015; **67**:1–48.
21 Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.1. 2019. URL https://CRAN.R-project.org/package=emmeans. [Accessed 21 December 2019]
22 Kok P, Rahnev D, Jehee JF, Lau HC, de Lange FP. Attention reverses the effect of prediction in silencing sensory signals. *Cereb Cortex* 2012; **22**:2197–2206.