



HAL
open science

Intention-based and sensory-based predictions

Álvaro Darriba, Yi-Fang Hsu, Sandrien van Ommen, Florian Waszak

► **To cite this version:**

Álvaro Darriba, Yi-Fang Hsu, Sandrien van Ommen, Florian Waszak. Intention-based and sensory-based predictions. *Scientific Reports*, 2021, 11 (1), 10.1038/s41598-021-99445-z . hal-03424555

HAL Id: hal-03424555

<https://hal.science/hal-03424555>

Submitted on 16 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



OPEN

Intention-based and sensory-based predictions

Álvaro Darriba¹✉, Yi-Fang Hsu^{2,3}, Sandrien Van Ommen⁴ & Florian Waszak¹

We inhabit a continuously changing world, where the ability to anticipate future states of the environment is critical for adaptation. Anticipation can be achieved by learning about the causal or temporal relationship between sensory events, as well as by learning to act on the environment to produce an intended effect. Together, sensory-based and intention-based predictions provide the flexibility needed to successfully adapt. Yet it is currently unknown whether the two sources of information are processed independently to form separate predictions, or are combined into a common prediction. To investigate this, we ran an experiment in which the final tone of two possible four-tone sequences could be predicted from the preceding tones in the sequence and/or from the participants' intention to trigger that final tone. This tone could be congruent with both sensory-based and intention-based predictions, incongruent with both, or congruent with one while incongruent with the other. Trials where predictions were incongruent with each other yielded similar prediction error responses irrespectively of the violated prediction, indicating that both predictions were formulated and coexisted simultaneously. The violation of intention-based predictions yielded late additional error responses, suggesting that those violations underwent further differential processing which the violations of sensory-based predictions did not receive.

The ability to make predictions about future states of the environment allows humans to adapt their perception and optimize their behaviour. According to predictive coding models^{1,2}, the brain represents predictions as probability distributions that are continuously compared with actual evidence and adjusted correspondingly. Predictions are primarily based on knowledge and experience about the relation between events in a given context. This includes causal and temporal relationships between sensory events, but also knowledge of the effects of our actions on those events. That is, we adapt to the environment by anticipating upcoming events and by producing expected events through our actions. This ability to generate sensory-based and intention-based predictions enables us to interact with our environment with great flexibility³. The present study aimed to use event-related potentials (ERPs) to investigate how these sensory-based and intention-based predictions interact and, particularly, whether or not they are made simultaneously and independently of each other.

Sensory-based predictions are based on global probabilities about how events happen in a given context. According to predictive coding models, the brain infers the causes of the inputs it receives and predicts future inputs accordingly⁴. Predictive processes are thought to progressively minimise the difference between predicted and actual sensory data, i.e., prediction error (PE). PE reduction is achieved through a hierarchically organized loop in which backward projections from one processing level to its subordinate provide a representation of the expected input, while reciprocal forward projections convey PE reporting the difference between that representation and the actual input⁵. Error signals are used to correct the representation, which is again provided to the subordinate level for comparison with the actual input. Sensory-based predictions are often studied employing variants of the oddball paradigm^{6–10}, where deviant, oddball stimuli occur infrequently and irregularly among standard, repetitive stimuli¹¹. However, since PE reduction partly manifests in these paradigms as repetition suppression, i.e., repeated stimuli generating reduced neural activity compared to novel stimuli^{12–15}, designs where apprehending the probabilistic structure of the environment and anticipating an upcoming stimulus require more complex computations^{16–19} are more suitable to study sensory predictions in relation to intention-based predictions, as is the goal of the present study.

In intention-based predictions, PE is reduced by producing desired (predicted) sensory states through action (i.e., active inference²⁰). This idea is closely related to ideomotor theories^{21,22}, which propose that actions are represented by the sensory effects they have been experienced to produce. Performing an action would result in a bidirectional association between the action itself and its effects on the environment, which would integrate

¹Université de Paris, INCC UMR 8002, CNRS, F-75006 Paris, France. ²Department of Educational Psychology and Counselling, National Taiwan Normal University, 10610 Taipei, Taiwan. ³Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, 10610 Taipei, Taiwan. ⁴Department of Basic Neurosciences, University of Geneva, Biotech Campus, Geneva, Switzerland. ✉email: alvarodarriba@gmail.com

into a common code. After integration, an agent can activate this code to select the adequate motor patterns to produce a wanted effect and, conversely, anticipate the effect of a given action on the environment before executing it²³. During action selection the predicted effect is internally anticipated or simulated. This prediction is then matched against the actual effect, the difference being sent to a higher level as PE and used to update the internal model^{24,25}. Although originally formulated to explain action-unrelated sensory predictions, predictive coding models have recently been extended to the study of action-control²⁶. Perhaps due to this late incorporation, research on sensory-based and intention-based predictions have followed largely separate lines.

Nevertheless, a series of brain responses have been demonstrated to occur with both types of predictions. In the present work we studied some of these responses through electroencephalography (EEG), focusing on the analyses of the N1, P2, N2b, P3a, and P3b ERP components, since they have reliably shown modulations in relation to the predictability of auditory stimuli. Sensory-based and intention-based predictions produce attenuated brain responses in the N1-P2 time range in relation to predicted stimuli (for reviews, see^{3,27}). N1 is considered to reflect processing leading to conscious detection and orientation towards a sensory event^{28,29}. Although the functional interpretation of P2 is debated³⁰, it has been suggested to reflect higher order stimulus evaluation and classification processes^{31,32}. In studies on sensory-based predictions, standard, predicted stimuli have been observed to yield attenuated brain responses in this time range compared to deviant, unpredicted stimuli, in experiments typically employing oddball paradigms^{7,9,12,15,33}. Research on intention-based predictions has shown that self-generated tones elicit attenuated N1 and P2 responses in comparison to externally generated tones³⁴⁻³⁷. In this context, the N1b and Tb subcomponents of the N1 have been particularly related to predictions about the sensory characteristics of the stimuli, their amplitude being suppressed in response to self-generated (and therefore predicted) sounds and enhanced in response to a mismatch between the predicted and the actual auditory stimulus^{36,38,39}. Interestingly, the N1-P2 attenuation does not seem to occur when actions are induced by transcranial magnetic stimulation (TMS)⁴⁰, which suggests that it is the intention, rather than the action itself, that is necessary for generating a prediction about the effect of an action. N2b and P3 brain responses have also received substantial attention in this context. The amplitude of the N2b component, classically shown to be larger for infrequent than for frequent auditory stimuli^{41,42}, has more recently been shown to be sensitive to the probability of occurrence^{43,44}, as well as to conflict and mismatch detection^{45,46}. Moreover, its amplitude is enhanced in response to unexpected stimuli triggered by intentional actions, thus signalling a PE response⁴⁷. Finally, modulations in the P3 ERP responses, including P3a and P3b⁴⁸, have been consistently observed to follow the aforementioned brain responses in both types of predictions. Amplitude enhancements of P3a have been extensively related to mismatch, surprise, and novelty processing, and consequently to PE⁴⁹⁻⁵¹. P3b has also been linked to a wide variety of processes, including target detection, contextual change evaluation⁵², decision confidence⁵³, and the updating of perceptual evidence⁵⁴, all of which may arguably constitute manifestations, in different experimental setups, of a prediction updating process following PE detection. Increases of P3b amplitude have been directly associated to PE in different studies^{33,55-57}.

Since the brain responses obtained from both prediction types are largely similar, one can reasonably suppose that a common, shared mechanism is responsible for their implementation. Conversely, it is also reasonable to hypothesize that these responses may depend on whether the eliciting predictions are based on sensory evidence or on the agent's intention, since even when both predictions refer to the same event, the underlying predictive models may cover partly different aspects of it in relation to the environment. Furthermore, sensory-based and intention-based predictions often cooccur, concurring or even competing to anticipate upcoming events. It is an open question whether in these cases converging predictions integrate or, rather, are carried out in an independent manner, their effects adding up instead.

In this experiment we manipulated sensory-based and intention-based predictions independently to investigate whether or not the two sources of information generate two independent predictions in the brain. In each trial, participants were presented with one of two possible sequences of four tones, the last tone (either a high or a low tone) being predicted by the preceding ones, as taught in a series of sensory-training blocks before and between experimental blocks. Unlike previous research, we used random tone sequences as sensory-based predictive information instead of a standard oddball context to prevent repetition suppression effects. In addition, participants were cued at the beginning of each trial to generate either the high or the low tone by pressing one of two possible keys, in synchrony with the last tone of the sequence. Prior to the experimental trials, and between blocks, participants ran a series of action-training blocks in which they learned to associate the left and right keypresses with the same two high and low tones they learned as possible final tones of the sequences. Therefore, a sensory-based and an intention-based prediction converged to anticipate the last tone. Both predictions could be congruent or incongruent with each other, anticipating the same or different tones respectively. In addition, the tone eventually played could fulfil or not any of the predictions made, yielding four experimental conditions: both predictions fulfilled; both predictions violated; intention-based prediction fulfilled but sensory-based prediction violated; and intention-based prediction violated but sensory-based prediction fulfilled. Please, note that the critical comparisons involved predicted and mispredicted (rather than unpredicted) stimuli, an important differentiation given that mispredicted and unpredicted stimuli may have dissociable neurophysiological mechanisms^{19,58}. We hypothesized three possible scenarios. First, one prediction prevails over the other in a winner takes it all manner. In this scenario, the pattern observed when the prevailing prediction is violated would be similar to that obtained when both predictions are violated, and the pattern observed when the prevailing prediction is fulfilled would be similar to that obtained when both predictions are fulfilled. We would then observe only one main effect. Second, both predictions are made and interact in one of two ways. In the conservative way, both predictions need to be fulfilled to avoid a PE. Therefore, error responses would be identical when both predictions are violated and when sensory-based and intention-based predictions compete (one fulfilled, one violated), compared to when both predictions are fulfilled. In the liberal way, only one prediction needs to be fulfilled to avoid a PE. Consequently, error responses would be the largest when both predictions

	C4	D5	E5	F5	G5	B5
Frequency (Hz)	261.63	587.33	659.26	698.46	783.99	987.77

Table 1. Frequency of each tone.

are violated, and identical and smallest when sensory-based and intention-based predictions compete and when both predictions are fulfilled. In these two cases we would observe the corresponding interactions. Third, both sensory-based and intention-based predictions are made independently of each other and their effect on PE is additive. In this scenario, error responses would be largest when both predictions are violated and intermediate when sensory-based and intention-based predictions compete, all compared to when both predictions are fulfilled. In this last scenario we would find two independent main effects.

Materials and methods

Participants. A total of 21 healthy volunteers (age: mean = 21.71, SD = 1.42; 8 males; 19 right-handed) participated in the experiment with no history of neurological, neuropsychiatric, or visual/hearing impairments as indicated by self-report. All participants gave written informed consent and were paid for participation. The study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics Committee at National Taiwan Normal University.

Stimuli and procedures. Six sinusoidal tones were generated using Sound Forge Pro 10.0 (Sony Creative Software Inc.). The duration of each tone was 50 ms (including 5 ms rise/fall times). The frequency of each tone was within the range of 261.63–987.73 Hz, matching the absolute frequency of a series of 6 natural keys on a modern piano (i.e., C4 D5 E5 F5 G5 B5) (Table 1).

E-prime version 2.0 (Psychology Software Tools) was used for stimulus presentation. Auditory stimulation was delivered binaurally via headphones (Sennheiser PX200-II) with an intensity of maximum 83.3 dB (56–82.7 dBA; 65–83.3 dBC). Visual stimuli (cue and feedback) were presented at the centre of a 16-inch cathode-ray tube (CRT) screen placed at a viewing distance of 120 cm. Participants were instructed to place their thumbs on the “1” and “3” keys on a numeric keypad.

For each participant, two sequences of four tones were generated at the beginning of the experiment, according to the following restrictions: first, one of the sequences ended with the lowest tone (C4) while the other ended with the highest tone (B5); second, for each participant the preceding three tones in each sequence were randomly chosen among the other four possible tones (D5, E5, F5, G5), with the only restriction that no tones could be repeated within a sequence.

Prior to the experiment, participants were presented with four sensory-based and four intention-based prediction training blocks, containing 20 trials each (Fig. 1, upper panel A). Sensory-based and intention-based blocks were presented in an alternate manner. Sensory-based blocks intended to allow participants to learn the tones that complete each of the two possible sequences they would be presented with during the experiment. Participants just looked at the screen and passively listen to each of the two possible sequences, randomly interspersed. In intention-based blocks participants were required to press one of the keys every time it was indicated on the screen, so that the keypress would trigger a given tone. These blocks aimed at allowing participants to learn the association between each one of two actions (right- or left-hand keypresses) and each one of two possible tones (the low or the high tone, i.e., C4 or B5), and thus the probability of each keypress triggering always the same tone was 100%. The hand/tone association was counterbalanced between participants. Besides the initial training blocks, participants were presented with 10 sensory-based and 10 intention-based training trials prior to every experimental block.

Figure 1 (upper panel B) shows a schematic illustration of the design. In each trial participants were presented with one of the two sequences of four tones and were required to generate the fourth and last tone by pressing one of the two keys after hearing the first three tones of the sequence. A cue presented at the beginning of each trial indicated participants whether they had to generate the high or the low tone, according to what they had learned in the action-based training blocks, independently of the presented sequence. The cue was presented for 500 ms, and consisted of one or two possible tone symbols indicating a high or a low tone. Participants were instructed about the meaning of these symbols before the experimental blocks. The tone that participants were asked to generate could be coherent or not with the tone expected to complete the sequence on the basis of what participants had learned in the sensory-based prediction training blocks. Therefore, the last tone could be coherent with the intention-based prediction, with the sensory-based prediction, with both or with none according to the probabilities shown in the lower panel of Fig. 1, yielding the four possible experimental conditions. Participants were instructed to press the key to complete the sequence with approximately the same timing the tone would be played if a keypress was not required (500 ms after the previous tone). When participants pressed too quickly (less than 300 ms before the expected timing, that is, less than 200 ms after the preceding tone) or too slowly (more than 1000 ms after the preceding tone), a feedback message (“Too fast” or “Too slow”, respectively, was displayed). The next trial started 700 ms after either the tone is played or the feedback is presented. The experiment consisted of 8 blocks containing 70 trials each, for a total of 560 trials, including 320 trials with both predictions fulfilled, and 80 for each of the three other conditions (sensory-based prediction fulfilled and intention-based violated; sensory-based violated and intention-based fulfilled; both predictions violated). 80 additional catch trials were randomly intermingled with the 560 experimental trials in order to keep participants actively involved in the experiment. These trials were similar to the experimental ones with the only difference

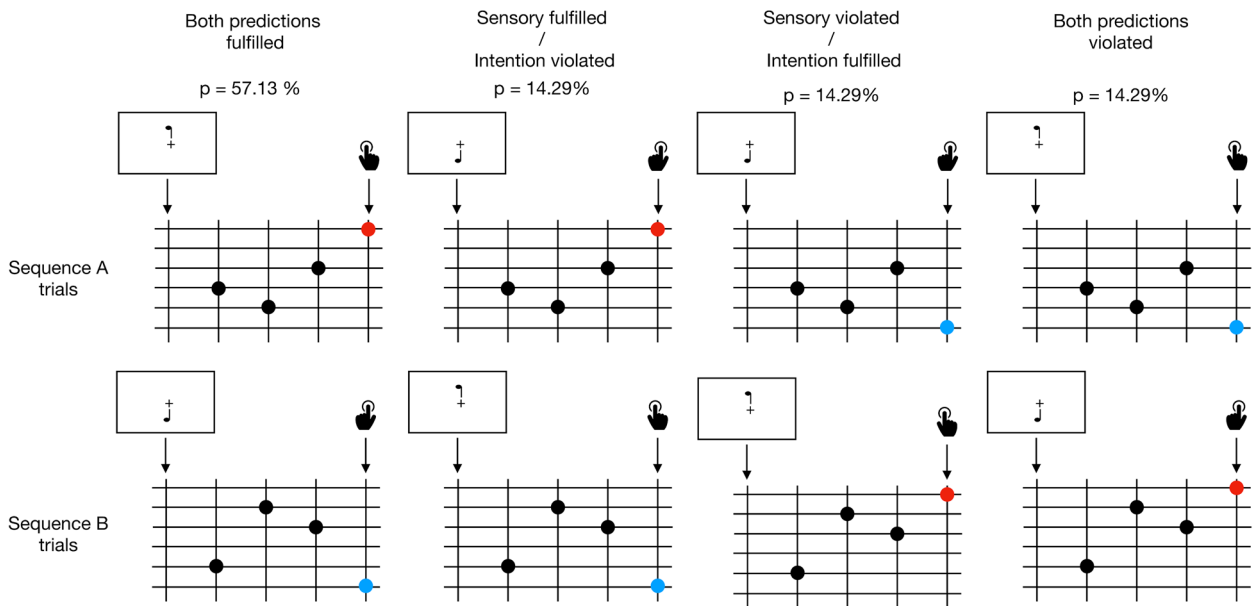
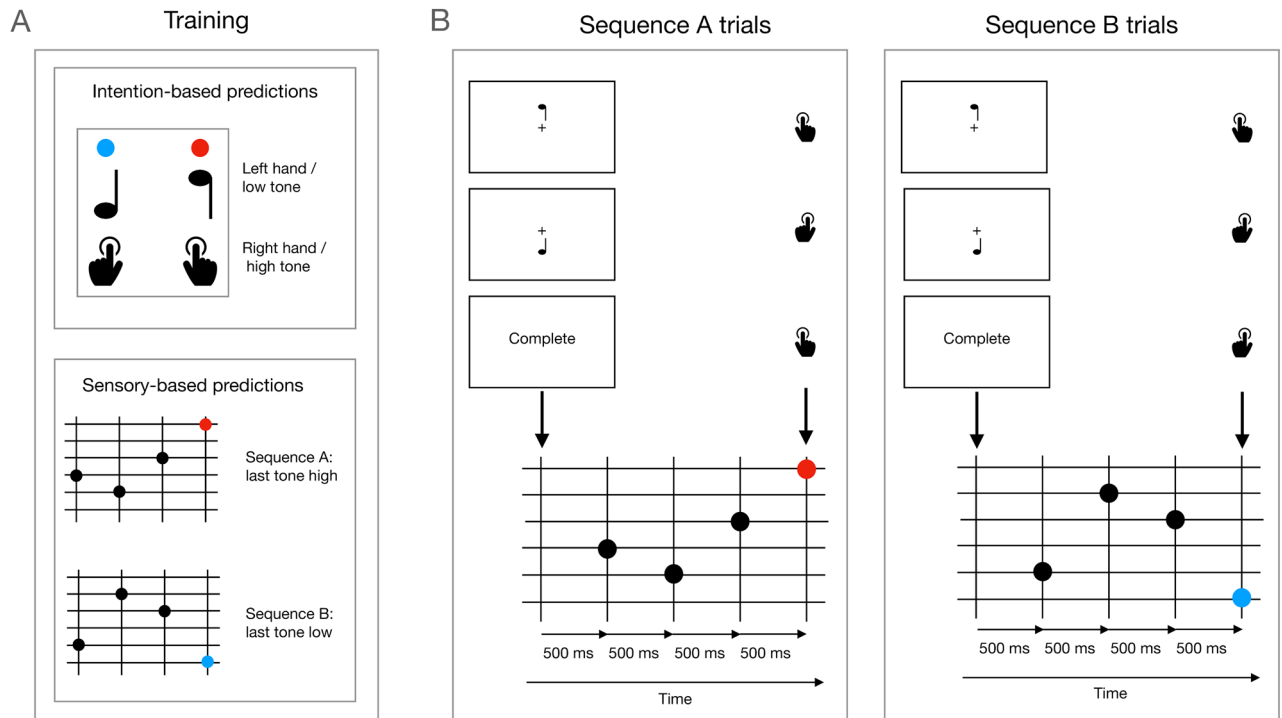


Figure 1. *Upper panel* Schematic description of the task. (A) Prior to the experiment, participants performed a series of training blocks aimed at allowing them to associate left and right keypresses with high and low tones (intention-based predictions) and to learn the two possible sequences they could be presented with (sequence A, sequence B), each ending either with the same high or low tones (sensory-based predictions) participants associated with each of the keypresses. Participants performed four blocks of 20 trials for each type of prediction. Additional, shorter training (one block of 10 trials for each type of prediction) was provided before each experimental block. (B) On each experimental trial, participants were presented with either sequence A or B, randomly intermingled between trials. Every trial commenced with a cue signal (500 ms), prompting participants to press a key in order to produce the last tone of the sequence in synchrony with the moment in which it should be played. This cue indicated participants to press either the key that produces the high tone, the key that produces the low tone, or the key that produces the tone that completes the sequence (catch trials). Therefore, the last tone could be predicted on the basis of the preceding sequence of tones (sensory-based predictions) and on the basis of the executed action (intention-based predictions), yielding four experimental conditions: sensory and intention predictions fulfilled (57.13% of trials), neither sensory nor intention predictions fulfilled (14.29%), sensory fulfilled but intention prediction violated (14.29%), and sensory violated but intention prediction fulfilled (14.29%). *Lower panel* Representation of the four possible experimental conditions for trials from two given sequences A and B, as indicated in the upper panel, together with the relative probabilities of each condition over the total number of trials.

that the initial cue prompted participants to produce the tone that completed the presented sequence instead of a particularly cued one.

EEG recording and pre-processing. EEG was recorded from 62 sintered Ag/AgCl electrodes, placed according to the extended 10–20 system on a Neuroscan quik-cap. The reference electrode was placed between Cz and CPz, and the ground electrode was set at the AFz site. Four additional electrodes were placed above and below the left eye and at the outer canthi of both eyes, and bipolarized online to register vertical and horizontal electrooculogram (EOG), respectively. A Neuroscan Synamps 2 amplifier (Compumedics Neuroscan, USA) was used to amplify and online filter (0.1–100 Hz) all signals, which were recorded at a 500 Hz sampling rate.

EEG data were processed using EEGLAB⁵⁹ v2021.0 running under MATLAB R2020a (Mathworks, Navick, MA). Images depicting the ERP waveforms and the topographic distribution of voltage were obtained using the same software. Pre-processing was performed as follows. EEG data were re-referenced offline to linked mastoids. Data were filtered using a 0.1-Hz high-pass and 45-Hz low-pass windowed sinc finite impulse response filter (hamming window, filter order 8250 [high pass] and 166 [low pass])⁶⁰. Bad channels were then identified by visual inspection and excluded from processing. On average, 0.52 ± 0.73 channels were removed. Epochs for each stimulus type were extracted from -200 to $+1000$ ms with respect to the target stimulus in each sequence, and were inspected for non-stereotyped artifacts and removed if present ($1.70\% \pm 1.87$ of trials removed). Stereotyped artifacts, including blinks, eye movements, and muscle artifacts were deleted via independent component analysis (ICA) using the extended infomax algorithm (Bell and Sejnowski 1995). Components containing those artifacts were rejected by visual inspection and based on measures computed with FASTER⁶¹, ADJUST⁶², and SASICA⁶³. The average number of independent components removed was $3.45 (\pm 0.95 \text{ SD})$. The remaining components were then projected back into electrode space. On average, the minimum number of trials available for analysis was $314.6 (\pm 4.64 \text{ SD})$ in the both predictions fulfilled condition, and $78.64 (\pm 1.49 \text{ SD})$ in all the other conditions. EEG data were then transformed using a surface Laplacian filter (smoothing = 10^{-5} , number of iterations = 10, spherical spline order = 4) to reduce volume conduction effects in EEG electrode space using the CSD Toolbox⁶⁴. Finally, channels that were deemed bad were reintroduced by interpolating data between neighbouring electrodes using spherical spline interpolation⁶⁵.

EEG analyses. ERP analyses were performed on ICA-corrected CSD-transformed epochs time-locked to the onset of each target (-200 to $+1000$ ms). To minimize the influence of individual differences in topographies as well as the effects of performing multiple statistical comparisons, the analyses of the ERP components were performed on different ROIs of relevant sites, selected on the basis of both the grand average visual detection of the maximal peak electrodes and the topographical distribution of the activity on the scalp (see Figs. 3, 4). The time windows of interest were also determined based on the observed grand averages. Following this procedure, P2, N2b, and P3a were measured on a frontocentral cluster including FC1, FCz, and FC2 in 20 ms windows with regards to the most negative and the most positive points in the latency range of 120–180 ms, 180–220 ms, and 250–300 ms respectively. The electrodes and time windows selected for analyses are compatible with those selected in previous works^{35,36,38,40,66}. Regarding N1, given that it is known to consist of several subcomponents, we ran analyses on the N1a (Na), N1b and N1c (Tb)^{28,36,38,67}. Following Tonnquist-Uhlen et al.⁶⁷, N1b was measured on C3 and C4 electrodes, where it showed its largest amplitudes (120–170 ms), and Na and Tb were identified as the first (60–110 ms) and second (120–170 ms) negative peak after stimulus onset on electrodes T7 and T8. The N1 subcomponents were analysed in 20 ms windows with regards to the most negative point in the indicated latency ranges. Finally, P3b was measured on a centroparietal cluster including CP1, CPz, and CP2 in a 50 ms window with regards to the most positive point in the latency range of 300–400 ms⁶⁸. Baseline was designated in every case from -200 to 0 ms relative to stimulus onset.

Statistical analyses. Results were analysed with a Bayesian linear mixed-model (LMM) analysis using the package brms⁶⁹, a high-level interface on Stan⁷⁰ in R⁷¹. Plots were made using brms and ggplot2⁷². An advantage of LMMs over traditional approaches such as repeated measures ANOVA and paired sample t-tests is that a single model can take all sources of variance into account simultaneously. Furthermore, comparisons between conditions can easily be implemented in a single model. LMMs (of which t-tests and ANOVA are specific examples) allow for modelling complex data structures and taking correlations in data structures into account. Bayesian LMMs do so in a more powerful way than maximum likelihood models, even with small sample sizes. With weakly informative priors, Bayesian analysis gives insight in the range of possible effect sizes, reduces possible overinterpretation of sampling error, and allows for direct comparison of effect sizes. It is theoretically distinct from frequentist statistics in its inferences. The coefficient estimates are expressed in credible intervals. Credible intervals reflect the intuitive notion of the value of a parameter falling within that interval with a given probability, 95% in this case.

We used a predefined model reflecting our experimental design⁷³, and we kept this model structure the same across ERP components. Participant amplitudes were normally distributed and did not need transformation to their logarithmic function⁷⁴. Amplitudes were scaled for ease of interpretation and comparison. In the model, observations were predicted by Sensory (violated vs. fulfilled) and Intention (violated vs. fulfilled) in a full interaction. The model additionally included individual participant intercepts and slopes of Sensory and Intention in order to account for individual variation. Contrasts of all categorical factors were centred⁷⁵, so the intercept of the model represents the grand mean. Planned pairwise comparisons were conducted via Bayesian hypothesis testing using the function Hypothesis in brms with Bonferroni correction. We used a generic weakly informative prior with mean 0 and 1 SD over the fixed effects and kept all other priors at default. We used 4 chains of 3000 iterations each per model, of which 1000 per chain were used for warm-up only, a maximum tree depth of 15

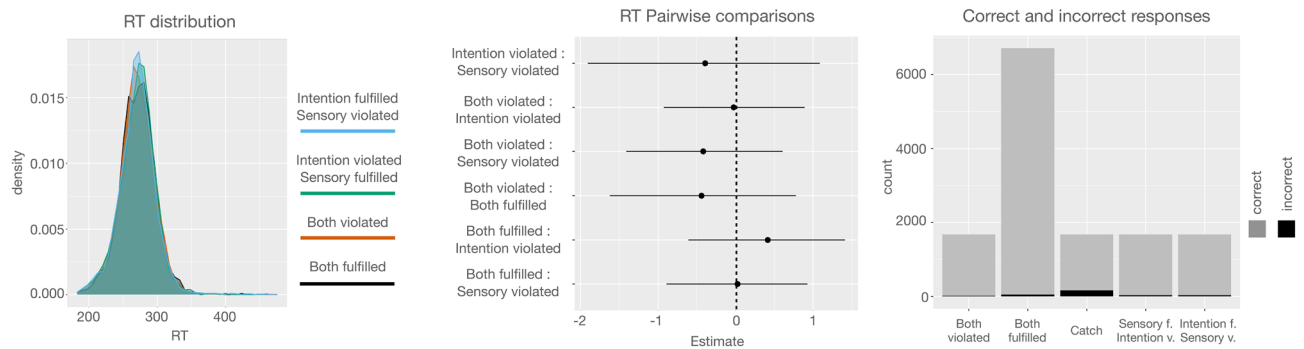


Figure 2. Behavioural results. Left panel shows the distribution of RTs per condition. Central panel illustrates the RT pairwise comparisons. No significant differences were observed. Right panel shows the rate of correct and incorrect responses per condition. Results indicate that participants correctly followed the instructions throughout the task. The low rate of incorrect responses in the catch trials indicates that participants were able to correctly identify the two possible two sequences.

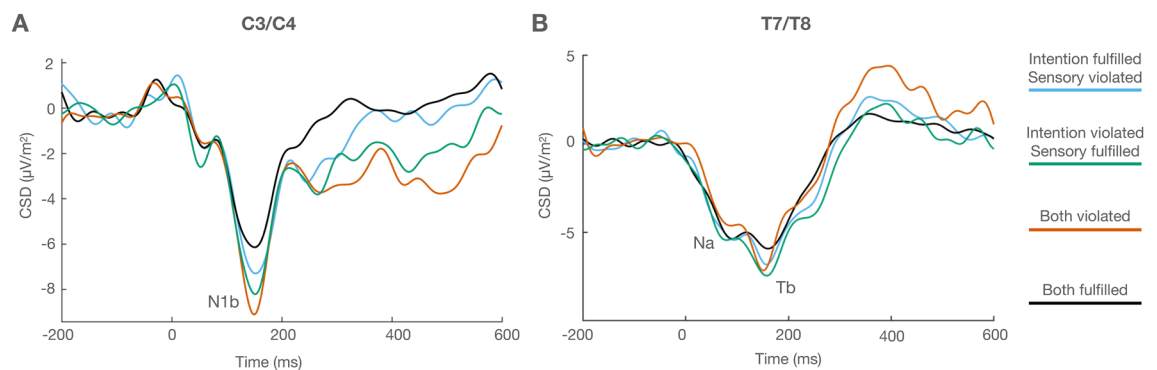


Figure 3. (A) ERP waveforms corresponding to the average of C3 and C4 electrodes, showing N1b (148 ms). (B) ERP waveforms corresponding to the average of T7 and T8 electrodes, depicting Na (80 ms) and Tb (158 ms).

and a target acceptance rate (adapt delta) of 0.95. Convergence was verified through visual inspection of trace plots, and the Rhat of 1.00 for each parameter.

The model was specified as follows,

$$brm(formula : amplitude \sim sensory * intention + (1 + sensory * intention|participant)).$$

Results

Behavioural performance. Behavioural results (Fig. 2) showed that participants followed the instructions appropriately throughout the task, with a high rate of correct keypresses according to the initial cue and within the indicated time window (0.99 ± 0.02 SD), and similar rates across conditions: both fulfilled (0.99 ± 0.02 SD), sensory fulfilled/intention violated (0.98 ± 0.04 SD), sensory violated/intention fulfilled (0.98 ± 0.04 SD), both violated (0.99 ± 0.02 SD). Reaction times did not significantly differ between conditions either (both fulfilled, $272 \text{ ms} \pm 27 \text{ SD}$; sensory fulfilled/intention violated, $271 \text{ ms} \pm 27 \text{ SD}$; sensory violated/intention fulfilled, $271 \text{ ms} \pm 27 \text{ SD}$; both violated, $271 \text{ ms} \pm 28 \text{ SD}$). The rate of correct keypresses in catch trials was also high (0.9 ± 0.12 SD), showing that participants were able to learn and recognize the two possible sequences accurately.

ERP results. Figure 3 depicts the ERP waveforms at the average of C3 and C4 electrodes, where N1b reached its largest amplitude (peaking at 148 ms), and the ERPs at the average of T7 and T8, where Na (peaking at 80 ms) and Tb (peaking at 158 ms) were identified. Figure 4 (upper panel) shows the ERP waveforms corresponding to the frontocentral cluster (FC1, FCz, FC2) where P2 (140 ms in the both violated condition, 148 ms in the sensory violated/intention fulfilled and sensory fulfilled/intention violated conditions, 176 ms in the both fulfilled condition), N2b (198 ms), and P3a (262 ms) were maximal. The lower panel of Fig. 4 displays the ERPs at the centro-parietal cluster (CP1, CPz, CP2), where P3b (356 ms) reached largest amplitudes. As stated in the Introduction, we hypothesized three possible scenarios. First, one prediction prevails over the other, and violating or fulfilling the prevailing prediction generates the same pattern as simultaneously violating or fulfilling both predictions. In this scenario we would observe only one main effect. Second, both predictions are made and interact in one

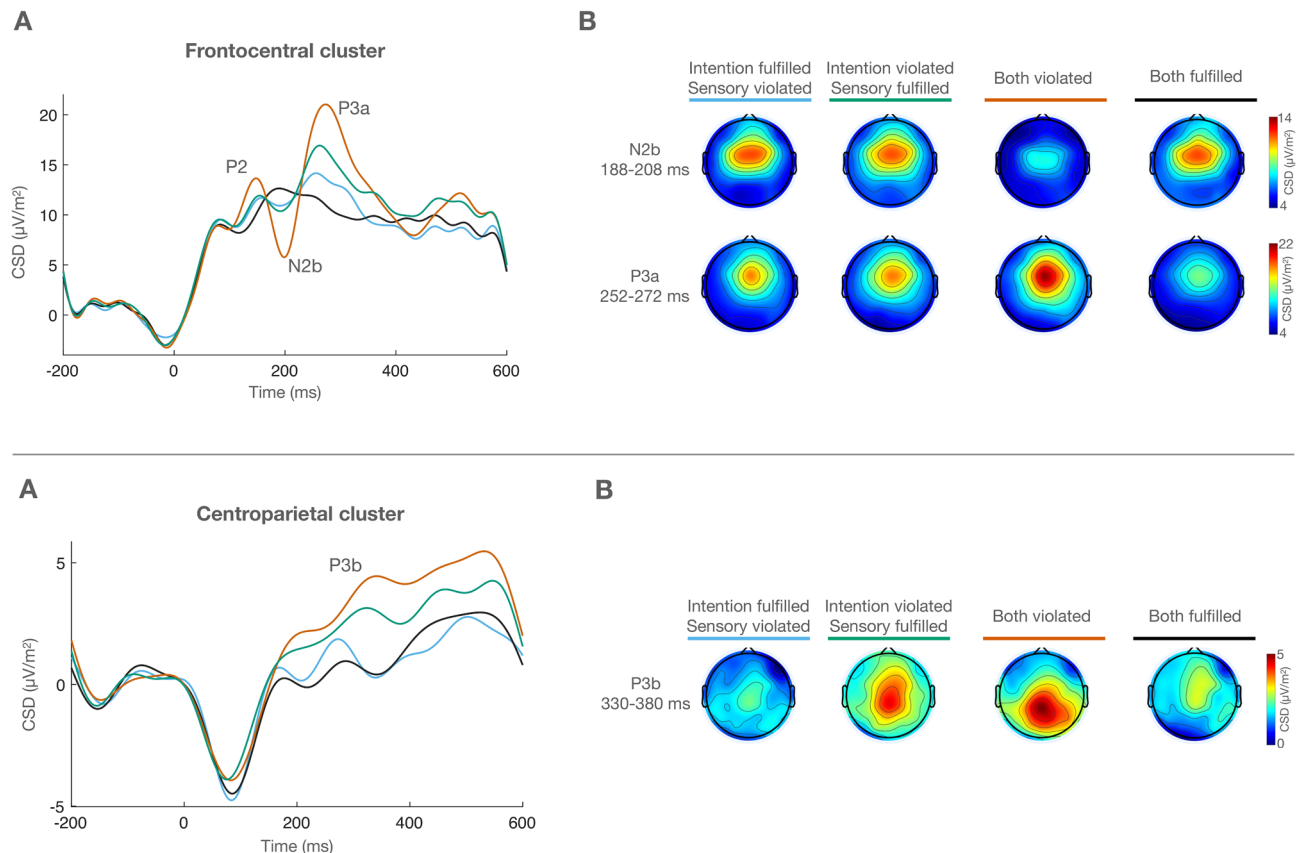


Figure 4. Upper panel (A) ERP waveforms corresponding to the frontocentral electrode cluster, showing P2 (between 120 and 180 ms), N2b (198 ms) and P3a (262 ms). Note that the P2 peak is delayed in the both fulfilled condition, as later confirmed by the statistical analyses. (B) Scalp voltage distribution maps of N2b and P3a for every condition. Lower panel (A) ERP waveforms corresponding to the centroparietal electrode cluster. P3b peaked at 356 ms. (B) P3b scalp voltage distribution maps for every condition.

of two ways. In the conservative way, both predictions need to be fulfilled to avoid a PE. Consequently, error responses would be identical when both predictions are violated and when sensory-based and intention-based predictions compete (one fulfilled, one violated), compared to when both predictions are fulfilled. In the liberal way, only one prediction needs to be fulfilled to avoid a PE. Therefore, error responses would be the largest when both predictions are simultaneously violated, and identical and smallest when sensory-based and intention-based predictions compete and when both predictions are fulfilled. In these two cases we would observe the corresponding interactions. Third, both sensory-based and intention-based predictions are made independently and their effect on PE is additive. In this case, error responses would be largest when both predictions are violated and intermediate when sensory-based and intention-based predictions compete, all compared to when both predictions are fulfilled. In this last scenario we would find two independent main effects. No significant results were obtained in the analyses of the amplitudes of Na, Tb, and P2. The analyses of N1b, N2b, P3a, and P3b, did show significant results. The results obtained in the statistical analyses are graphically and numerically illustrated in Figs. 5 and 6.

N1b results. The analysis of N1b amplitudes revealed main Sensory and Intention effects but no significant interaction, thus indicating that sensory-based and intention-based predictions independently modulated this component's amplitude. The planned comparisons showed that both types of predictions yielded PE responses of similar size when independently violated. Additionally, each type of prediction elicited a N1b of medium size compared to the both fulfilled and both violated conditions, with sensory-based and intention-based predictions significantly differing from the former and the latter, respectively.

P2 results. The analysis of P2 amplitudes did not show any significant results. However, visual inspection of the ERP waveforms suggested possible differences in the latency of this component. The analysis of the latencies revealed that this was indeed the case. A linear mixed effects model of latencies in ms, including the four-level predictor of 'Condition' using successive differences coding, and a participant intercept revealed a difference in latencies between conditions ($\beta = -26.86$, $t = -2.94$, $p < 0.005$). A pairwise comparison corrected with the Tukey method showed that the latency of P2 was significantly longer in the 'both fulfilled' than in the 'both violated' condition, ($EM = 26.86$, $t = 2.94$, $p = 0.024$) with no further differences between conditions (both violated,

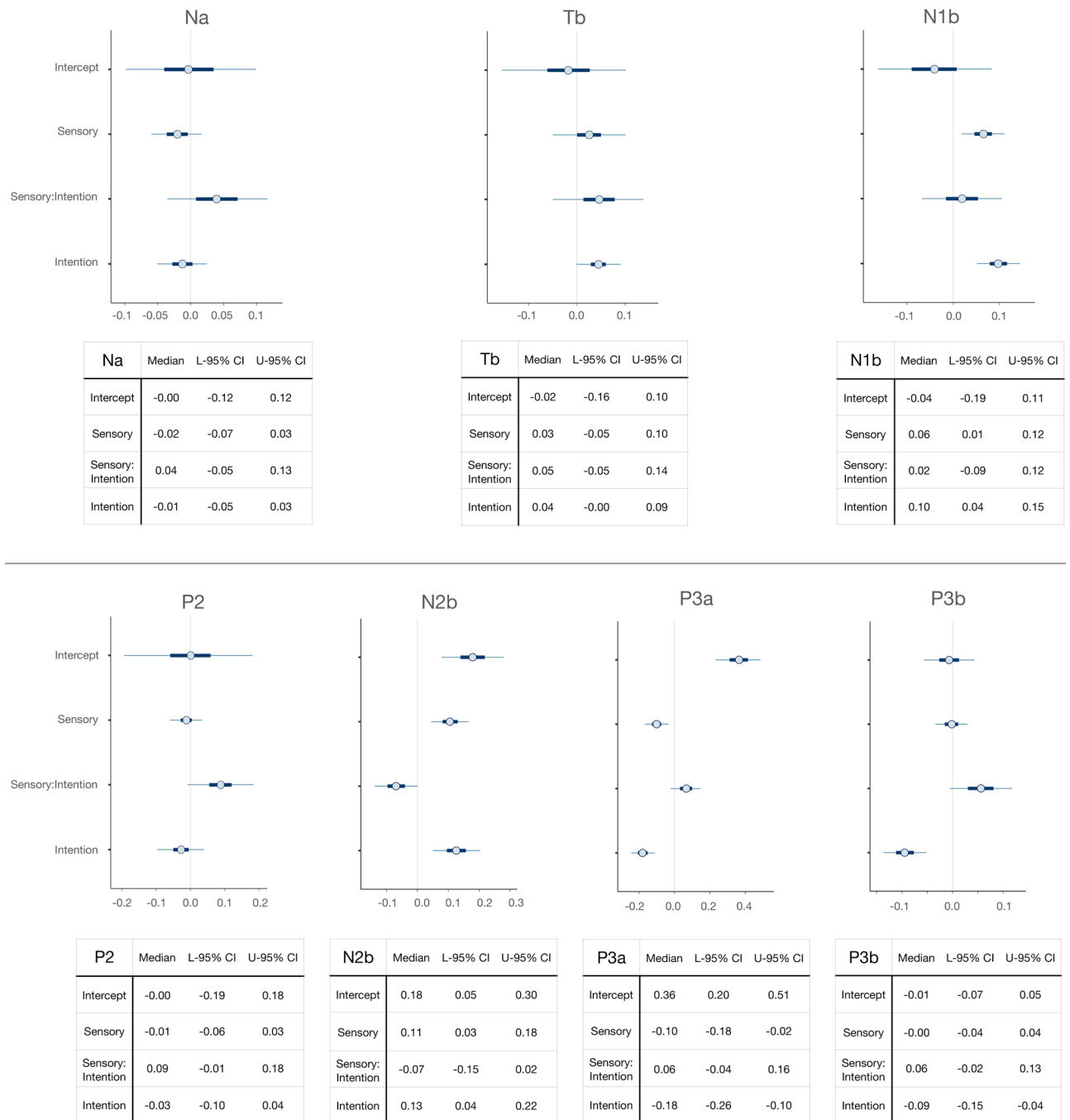


Figure 5. Medians and credible intervals (On the plots: 50%, thick line; 90% thin line. On the tables: 95%) of parameter values in Na, Tb, N1b, P2, N2b, P3a, and P3b. Intervals that do not include zero have the denoted probability to be a true effect.

140 ms; both fulfilled, 176 ms; sensory fulfilled/intention violated, 148 ms; sensory violated/intention fulfilled, 148 ms).

N2b results. Main Sensory and Intention effects were also observed in the analysis of N2b. As for N1b, these factors did not significantly interact, indicating that sensory-based and intention-based predictions independently modulated N2b amplitude. The planned comparisons showed, on the one hand, that the independent violation of sensory-based and intention-based predictions provoked a similar PE response and, on the other, that the magnitude of these PE responses was medium in size between the simultaneous violation of both predictions and the baseline set by their simultaneous fulfilment (i.e., both fulfilled condition).

P3a results. Similar to N1b and N2b, the analysis of P3a revealed main Sensory and Intention effects, and no significant interaction, pointing again to sensory-based and intention-based predictions modulating the ampli-

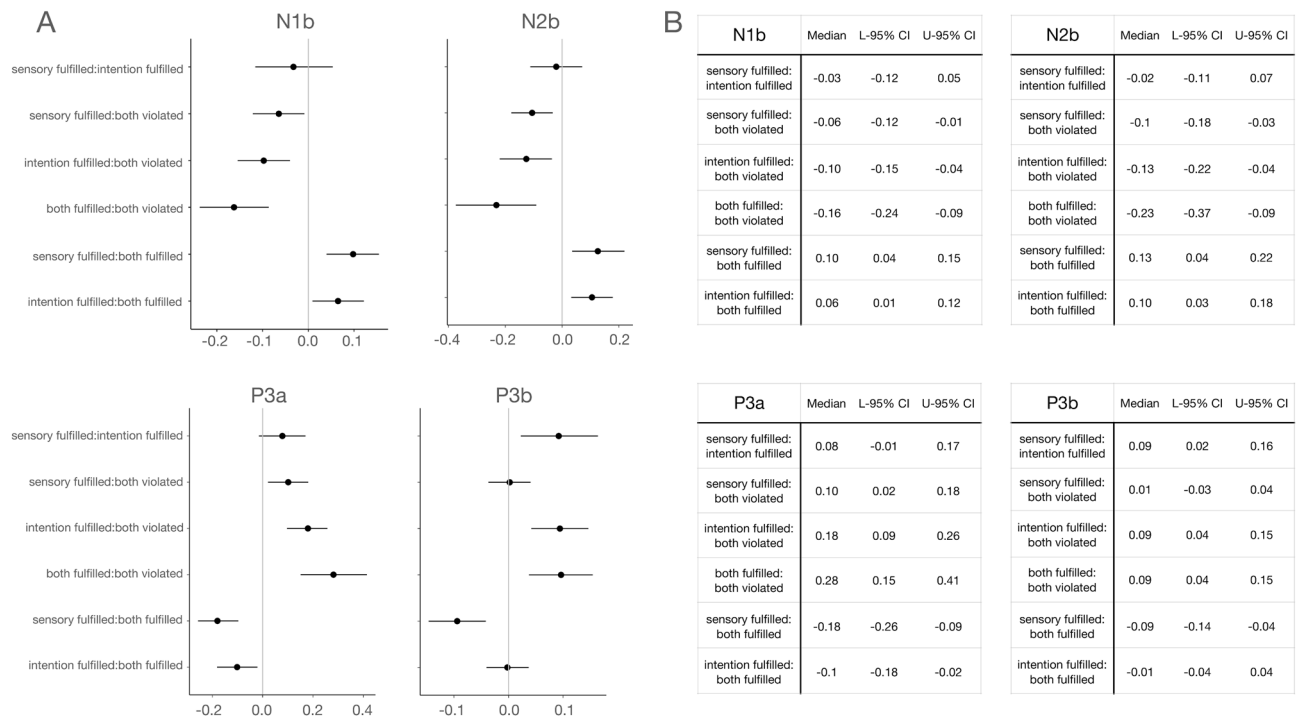


Figure 6. Medians and credible intervals (95%) of planned pairwise comparisons of condition estimates, calculated with the function ‘hypothesis’ in the R package ‘brms’, for components in which significant effects were observed. Intervals that do not include zero have the denoted probability to be a true effect, i.e., differences between the indicated factors to be significant. Note that the labels have been simplified for the sake of simplicity, so that sensory fulfilled refers to the sensory fulfilled/intention violated condition, and intention fulfilled to the sensory violated/intention fulfilled condition.

tude of P3a in an independent manner. As observed in the analysis of N2b, the planned comparisons indicated that the PE responses evoked by the violation of sensory-based and intention-based predictions were similar in size. These comparisons also revealed that the amplitude of P3a was in those conditions significantly larger and smaller than when both predictions were simultaneously fulfilled or violated, respectively.

P3b results. A main effect of Intention, together with the absence of both a main effect of Sensory and a Sensory*Intention interaction, indicated that the amplitude of P3b was modulated by the violation of intention-based but not of sensory-based predictions, independently of whether or not a sensory-based prediction was simultaneously violated. Accordingly, the planned comparisons showed that the amplitude of P3b was significantly larger in the two conditions in which an intention-based prediction was violated (both violated, intention-based violated) than in any other condition. The size of P3b did not differ between intention-based violated and both violated, confirming that the violation of sensory-based predictions had no impact on the size of P3b.

Test for homogeneity of variances. The results indicate that in the conditions in which the two predictions were not congruent the violation of one of them yield error responses of medium size compared to the condition in which both predictions were fulfilled and that where both were violated (on N1b, N2b, and P3a). However, it is possible that in the conditions in which predictions are contradictory the medium error responses are due to one type of prediction prevailing over the other on a trial basis. This would cause each of these two conditions (sensory fulfilled while intention violated; sensory violated while intention fulfilled), to contain trials in which the participants’ prediction was actually violated and trials in which it was not. As a consequence, some trials would contain a PE response while some trials would not, thus yielding a reduced PE response when averaged together. In this scenario, one would expect the distribution of amplitudes across trials to differ between conditions, being more widespread or perhaps bimodal in the conditions in which only one prediction is violated compared to those in which both predictions were either fulfilled or violated. In order to rule out this possibility, we ran the Levene’s test for homogeneity of variances on the amplitudes of the four ERP components in which significant effects were found, defining variance as absolute differences from the median. These tests could not ($\alpha=0.05$) reject the null hypothesis of equal variances for N1b [$F(3)=0.78$, $p=0.51$], N2b [$F(3)=0.61$, $p=0.61$], P3a [$F(3)=2.10$, $p=0.10$], and P3b [$F(3)=0.61$, $p=0.61$], thus excluding the possibility explained above. Figure 7 depicts the amplitude distributions of each ERP component. Our results suggest that sensory-based and intention-based predictions provoked independent effects on the PE responses, and that these effects added up when both predictions were simultaneously violated.

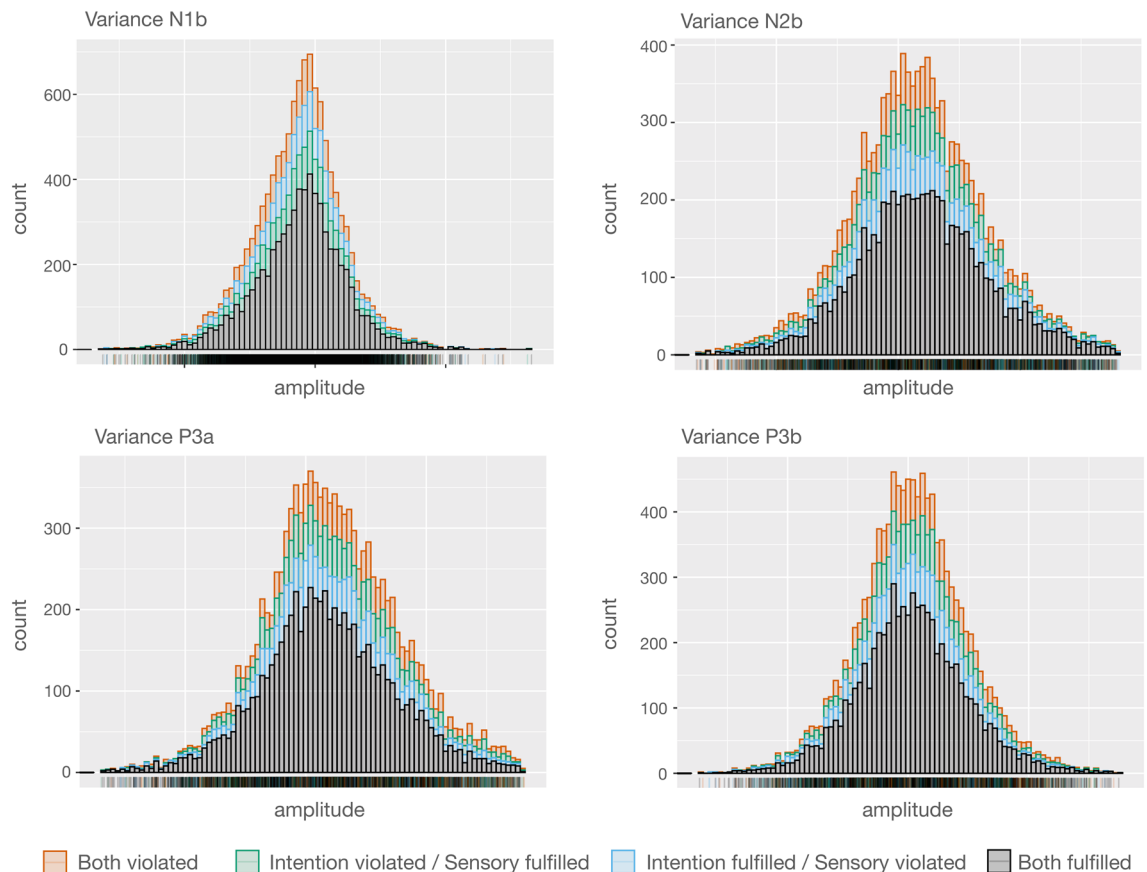


Figure 7. Distribution of amplitudes of each condition in stacked histograms for the four components where significant results were found. The Levene's test confirmed that variances did not differ between conditions for each component.

Discussion

In this work we investigated the interaction between sensory-based and intention-based predictions. We recorded EEG while participants performed a task in which they could predict a given stimulus on the basis of their intentional actions and/or agreeing or competing predictive sensory information. The critical comparisons thus involved predicted and mispredicted stimuli. We found two different patterns of brain activity. First, results showed main effects of sensory-based and intention-based predictions on the amplitudes of the N1b, N2b, and P3a components, and no interaction between these effects. ERP error responses were largest when both predictions concurred and were violated, compared to when both concurred and were fulfilled. When one of the two predictions was violated and the other fulfilled, results showed error responses of intermediate size. This pattern indicates that both predictions were formulated and coexisted simultaneously, so that PE was elicited when any of the predictions was violated. Second, results showed only a main effect of intention-based prediction on the amplitude of P3b. Specifically, P3b was significantly larger, compared to the trials in which both predictions were fulfilled, in those conditions in which the intention-based prediction was violated (both predictions violated, sensory-based fulfilled), but not when only the sensory-based prediction was violated. This suggests that the violation of intention-based predictions underwent further differential processing, which the violations of sensory-based predictions did not receive.

The earliest effects on the ERPs were observed as enhanced N1b amplitude when either sensory-based or intention-based predictions were violated. Interestingly, while both violations elicited a similar effect on N1b size its amplitude was significantly larger when both predictions concurrently failed. N1 has been shown to be reduced when predictions about the expected stimulus are fulfilled, and enhanced when those predictions are violated, presumably reflecting the operation of an internal predictive mechanism in sensory-based^{7,8,12} and intention-based^{36,40,76} predictions. More specifically, N1 is thought to reflect a lower-level cortical process involved in encoding simple physical features of the stimuli⁸, within a hierarchically organized deviance processing system in the auditory stream⁷⁷⁻⁷⁹. In agreement with those studies, the N1b modulations observed in the present experiment would reflect an early stage of PE processing, namely the detection of a discrepancy at the level of those physical features between the predicted and the actual sensory input. N1b amplitude did not differ depending on whether the violated prediction was sensory- or intention-based. This indicates that participants used both sources of information to make their predictions in a manner that rendered the final tone equally predictable on the basis of any of those sources. Whereas the effects observed on N1b largely agree with most of the literature, the lack of effects on Tb is difficult to interpret as it does not coincide with previous works in which effects on

Tb were observed in response to stimuli that could be anticipated^{36,38,39}. SanMiguel et al.³⁶ have suggested that, in contrast to the unspecific N1 component of the auditory ERP, which seems to simply reflect orienting responses towards a sensory event, the N1b and Tb components reflect actual predictions about the sensory features of the stimuli. However, beyond this distinction, there is not a clear functional interpretation of the T-complex (in which Tb is included) in the context of prediction, and most hypotheses posed to explain N1 attenuation actually do not differentiate between N1b and Tb attenuations³⁰. Even when analysed separately, they are usually discussed together. It is possible that the absence of an effect on Tb is related to the design we employed here. Most of the studies describing modulations of Tb in a predictive context have observed amplitude reduction of this component in a very specific context (of sensory attenuation), namely in response to self-generated compared to externally generated tones in a context of high regularity (where the same single tone was played in every condition throughout the experiments)^{36,40}. The design employed in those studies, and the comparisons made to reveal the effects on Tb, are very different from the design we used in our experiment, where there was not a constant repetition of the same stimulus and specific predictions about what the tone would be had to be made by choosing between different alternatives, with those predictions being fulfilled or violated across trials throughout the task. Interestingly, Korka et al.³⁸, employing a design aimed at investigating the interaction of sensory-based and action-based predictions, based on a high contextual regularity, found N1b enhancements in response to violations of sensory-based (regularity-based) and concurrent sensory- and action-based predictions, but, contrary to what could be expected on the basis of self-generation studies, Tb enhancements for the violations of sensory-based (regularity-based) predictions, but not for violations of action-based or simultaneous sensory- and action-based predictions. The results obtained in the studies cited above suggest that Tb modulations are particularly sensitive to contexts of high perceptual regularity. Although difficult to interpret, the absence of Tb modulations observed here could be related to the significantly lower contextual regularity in the design we employed, compared to those previous works.

P2 followed N1 in time. The literature on P2 has provided mixed results⁸⁰, with some studies showing the enhancement or attenuation of brain responses in the P2 time range in response to unpredicted or predicted events on sensory-based^{7,9,12,15,19,33} and intention-based predictions^{34–37}, and others failing to show such effects⁸⁰. It has been proposed that P2 modulations would represent a second stage in the formation of a prediction, so that while the N1 reflects PE related to a lower-level forward prediction that prepares the auditory cortex, P2 represents a more “cognitive” PE response⁸⁰, such as the processing of more complex characteristics of the stimulus⁷⁷, a conscious detection of the predicted stimulus⁸¹, or even PE attenuation during the formation of memory traces as part of the repetition positivity phenomenon⁸². Unlike previous studies, our results did not show any effect on P2 amplitude. We can speculate about the reasons for this unexpected result. On the one hand, it might indicate that the kind of processing indexed by P2 was not necessary in the experimental context we employed. On the other hand, it could be due to P2 being overlaid by a N2b component in trials where any prediction was violated compared to those where both predictions were fulfilled⁸⁰. This latter explanation would be supported by the effect observed on P2 latency, indicating that it peaked significantly later in trials where both predictions were fulfilled, which additionally showed a very small N2b response compared to trials containing any prediction violation, and particularly to trials where both predictions were violated, which showed a very large N2b response. The early detection of the PE, as signalled by N1b, would have activated the next stages of PE detection, improving the processing of the mispredicted stimuli, more informative than the less relevant predicted ones, which would be reflected in significant N2b responses in those trials, as explained below.

The initial detection of a PE reflected in the N1b was followed by effects on the N2b component. Specifically, as observed for N1b, N2b amplitude was significantly enhanced by the violation of any prediction compared to trials in which no violation occurred. Also, similarly to N1b, the effects of the violation of both prediction types on N2b amplitude added up instead of integrating when both predictions concurred. Several negative components occurring in the time-range of N2b, such as the mismatch negativity (MMN), the error-related negativity (ERN), or the feedback-related negativity (FRN) have been related to error signalling processes in different experimental contexts, while the possible functional relationships or equivalences between them has been the object of significant discussion (for reviews on this topic, see^{43,44}). Beyond this debate, N2b amplitude is known to be sensitive to the probability of occurrence of the stimuli, showing larger amplitudes in response to deviant or unexpected stimuli^{41,83}, and to require attention to the eliciting stimulus in order to be triggered⁴³. N2b modulations have been consistently shown in relation to PE resulting from violated sensory-based and intention-based predictions^{35,46,47,66,84}. The modulations observed here would reflect the conscious detection and registration of a mismatch between the expected and the actual sensory event, as suggested by previous research^{36,67}. As with the above-mentioned lower-level detection process signalled by N1b, this conscious detection would result from the independent monitoring of sensory-based and intention-based predictions, as suggested by the fact that N2b amplitude was enhanced whenever any of the two predictions were violated.

The error detection stages reflected in N1b and N2b likely constitute the first steps in a processing sequence⁷⁶, followed by further effects on the frontocentral P3a and the centroparietal P3b. P3a is considered to reflect the engagement of frontal attention mechanisms for evaluating unexpected, deviant stimuli, which presumably makes those stimuli available to consciousness and behavioural control^{48,85}. According to this view, the larger P3a component to mispredicted stimuli would be the consequence of the PE manifested at earlier processing stages reaching a certain threshold and thus activating an additional processing stage, which involves an involuntary capture of attention^{66,86}. We found, as with N1b and N2b, larger P3a amplitudes for stimuli mispredicted on the basis of either sensory or intention information compared to trials where both predictions were fulfilled. Our results agree with previous research, showing similar P3a enhancement signalling PE in response to violated sensory-based^{86–88} and intention-based predictions^{35,81,89–91}. This indicates that the violations of both types of predictions were equally salient, triggering a similar attention orienting response aimed at further processing

the deviance, as both violations were unexpected and therefore very informative for a hypothetical subsequent prediction adjustment^{66,92}.

Finally, the effects described above were followed by differences in the centroparietal P3b. Unlike what we observed in the frontocentral N1b, N2b, and P3a, the amplitude of P3b was significantly larger when the intention-based predictions were violated (i.e., both predictions violated, sensory-based fulfilled/intention-based violated). However, when only the sensory-based prediction was violated the amplitude of P3b did not significantly differ from trials in which both predictions were fulfilled. Among many other processes⁴⁸, P3b has been related to uncertainty or surprise^{55,83}, to decision processes⁵⁶, and to context-updating operations^{48,52,54,93}, including the updating of an internal prediction model⁹⁴ and, specifically, the reactivation of well-established stimulus–response links⁹⁵. Given its relationship with context updating, P3b has often been taken as a marker that stimulus processing is completed⁹⁶. In the present study the enhancement in the posterior P3b component would reflect the operation of such a mechanism, reactivating and re-evaluating, rather than updating⁹⁷, the link between the motor action and the stimulus when the association action-effect is violated. This step would constitute the final step of stimulus processing in those trials.

Our data indicate that predictions based on sensory and intention information were equally efficient in anticipating an upcoming stimulus, as reflected in the amplitude modulations of N1b, N2b, and P3a. The sensitivity of these components to PE in relation to sensory-based and intention-based predictions has been shown by previous works, as explained above, and they have been suggested to be correlates of successive processing stages within a hierarchically organized deviance detection system in the auditory domain^{77,78}. According to these views, different levels of information would be serially processed at different time intervals, with simple regularities being evaluated at early processing stages and more complex and integrated features being encoded in later time intervals, the predictive signals passing through hierarchically organized regions, in accordance to predictive coding notions. Our goal was to take a further step by addressing the question of whether both types of predictions can exist simultaneously. The results obtained support this hypothesis, indicating that a specific prediction was generated independently for sensory-based (bottom-up) and intention-based (top-down) predictions, and that the input was compared to both in parallel. This would be in line with previous studies suggesting that more than one regularity representation can be active at the same time, competing until one is selected^{98–100}. The mechanisms processing PE in both types of predictions seem to have operated in a largely independent manner and share similar processing stages, as shown by the effects observed on the amplitudes of N1b, N2b, and P3a. Only in the final stage of stimulus processing, marked by the P3b component, stimulus processing seemed to diverge between trials in which intention predictions were violated and trials where they were not, probably because those violations tapped into the learned relationship between motor actions and their effects, reactivating the corresponding action-effect maps for re-evaluation, as explained above.

The results presented here might appear at odds with recent data from Korcka et al.³⁸ who compared regularity-based (i.e., sensory-based), intention-based, and joint regularity- and intention-based predictions. These authors showed the amplitudes of N1b and Tb components to be modulated by violations of tone regularity only, while violations of either regularity-based, action-based, or both, resulted in similar effects on the amplitudes of the MMN and the P3a. However, a direct comparison of the results is difficult. Unlike in the present experiment, where the two types of prediction are manipulated orthogonally, these authors compared three different types of experimental blocks which tested, respectively, regularity-based predictions, intention-based predictions, and simultaneous regularity-based and intention-based predictions. Moreover, in these latter blocks both types of predictions were always congruent with each other, so that they were always either fulfilled or violated simultaneously. Furthermore, participants were instructed to press one of the keys in 80% of the trials in order to create a tone regularity that generated both a sensory and an intention prediction, while the other key was pressed only 20% of the times. For these and other reasons related to differences in the designs employed, more research is needed to fully understand the differences in the results. These differences are particularly relevant with regard to the hypothesis on whether or not the simultaneous violation of sensory-based and intention-based predictions has additive effects on PE. Korcka et al.³⁸ reported no additive PE when regularity and intention were violated concurrently (i.e., the error responses were not larger compared to when only one of the predictions was violated), which led them to suggest that the two predictions integrated into a single one. In the present experiment, however, the simultaneous violation of sensory-based and intention-based predictions did elicit larger error responses than the violation of any of the predictions alone, as reflected on the amplitudes of N1b, N2b, and P3a, thus suggesting that both predictions, rather than integrate, remained independent, with the PE from the violation of each prediction adding up to produce larger PE responses. This interpretation would be in line with another study that investigated the interaction between bottom-up predictions made on the basis of auditory regularities (one of the two possible predicted tones had an overall much higher probability than the other) and top-down predictions determined by a visual cue presented at the beginning of each trial⁹⁸. As in our study, in each trial the two sources of information could agree or disagree in their predictions. Their results showed that the violations of bottom-up and top-down predictions provoked independent PE responses, and that those responses added to each other when both predictions were concurrently violated. Although it could be argued that the top-down predictions these authors used might not be comparable to those we employed in the present experiment¹⁰¹ (i.e., visual-auditory associations versus intention-based predictions), our results are in agreement insofar as in both cases they suggest that predictions made on the basis of bottom-up and top-down information were made simultaneously and independently from each other, rather than integrating into a single prediction. Future research should investigate whether different sources of top-down and bottom-up predictions yield similar results.

Despite our results fitting in this interpretation, there is a possible alternative account of the larger effects observed when the two predictions were concurrently violated that must be taken into account. The generation of two parallel predictions and their separate comparison with the sensory input could have been favoured by

the design we employed. On the one hand, participants were given the instruction to generate one or the other tone at the beginning of each trial. On the other hand, participants could quickly learn to anticipate what the last tone in the sequence should be on a sensory basis as soon as the first tones were played. It is possible, therefore, that participants detected the incongruence between both predictions very early in the trial, maybe as soon as the sequence started. Specifically, they may have learned that the probability of both predictions being congruent is more likely than the probability of predictions being incongruent (71.42% vs. 28.58%, respectively), so that the first stimuli in the sequence could generate an immediate PE response when predicting a last tone different from that indicated by the cue. This PE could be processed before the last tone is presented, which would result in reduced PE in response to the last tone. Such processing would not occur when both predictions are congruent, which would result in larger PE when they are concurrently violated, compared to when only one is violated. In order to test this, we analysed the ERPs in response to the first stimuli in the sequence (Fig. S1). We did not find significant differences between conditions in any component, which argues against the possibility of differences in PE responses prior to the presentation of the last tone explaining the pattern of results obtained. The absence of such differences, together with results indicating that both predictions were made simultaneously and coexisted independently from each other (as shown by the PE in response to any violation when predictions are contradictory and by the statistical model, revealing independent effects of both sensory and intention predictions on the ERP amplitudes), and with results showing that P3b was modulated by intention, but not by sensory predictions, suggest the independence of both predictions and support the interpretation of the larger amplitudes observed when sensory-based and intention-based were concurrently violated as reflecting the additive effects of their respective PE. However, the absence of a statistically significant Sensory-Intention interaction does not completely exclude the actual presence of such interaction. Moreover, given the differences in the relative probabilities between conditions and the well-established sensitivity of PE to probability, we cannot rule out the contribution of those probability differences to the larger effects observed when both predictions were concurrently violated. Therefore, the interpretation of larger error responses when both predictions are violated as indicating additive effects of both types of PE should be taken with caution. Future research should be conducted to specifically address this shortcoming of the present design (Fig. S1).

To conclude, our results indicate that predictions based on sensory information and on intentional action were processed in a largely independent way, eliciting separate PE responses at different processing levels. The pattern of results obtained suggests that the predictive models underlying each prediction covered partly different aspects of the sensory event beyond the strict representation of its physical features, presumably related to the different origins of the information each prediction was built upon, in one case the understanding about how events happen in the environment, and in the other some sense of agency on what the sensory event will be. Although generating and actively maintaining separate predictions on the same given event may not seem the most efficient strategy in terms of processing resources, particularly when those are redundant, it may indeed be advantageous if, as suggested by our results, the predictive models covered partially different facets of that event. We live in a complex environment in which events are often determined by multiple causes, often contradictory, and the ability to consider multiple sources of information to make predictions may allow us to better and more accurately anticipate several possible future states of the environment, providing us with the necessary flexibility to adapt our behaviour accordingly. This becomes more important when intentional actions are involved, since agents must often monitor how events happen in their surroundings for their actions to cause the desired effect at the right time. The results presented here provide evidence in this regard, showing that agents can make and maintain separate predictions on a given event based simultaneously on sensory patterns and on the expected effects of their actions.

Data availability

All the analyses described were performed using EEGLAB functions⁵⁹, the CSD Toolbox⁶⁴, and R⁷¹, which are freely available. Data presented in this paper will be made available on request subject to a formal data sharing agreement.

Received: 12 April 2021; Accepted: 23 September 2021

Published online: 06 October 2021

References

1. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **360**, 815–836 (2005).
2. Rao, R. P. N. & Ballard, D. H. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
3. Bendixen, A., SanMiguel, I. & Schröger, E. Early electrophysiological indicators for predictive processing in audition: A review. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **83**, 120–131 (2012).
4. Friston, K. J., Stephan, K. E., Montague, R. & Dolan, R. J. Computational psychiatry: The brain as a phantastic organ. *Lancet Psychiatry* **1**, 148–158 (2014).
5. Bastos, A. M. *et al.* Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
6. Cacciaglia, R., Costa-Faidella, J., Zarnowiec, K., Grimm, S. & Escera, C. Auditory predictions shape the neural responses to stimulus repetition and sensory change. *Neuroimage* **186**, 200–210 (2019).
7. Garrido, M. I. *et al.* Repetition suppression and plasticity in the human brain. *Neuroimage* **48**, 269–279 (2009).
8. Hofmann-Shen, C. *et al.* Mapping adaptation, deviance detection, and prediction error in auditory processing. *Neuroimage* **207**, 116432 (2020).
9. Wacongne, C., Changeux, J.-P. & Dehaene, S. A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.* **32**, 3665–3678 (2012).
10. Winkler, I. & Czigler, I. 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.* **83**, 132–143 (2012).

11. Squires, N. K., Squires, K. C. & Hillyard, S. A. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr. Clin. Neurophysiol.* **38**, 387–401 (1975).
12. Aukstulewicz, R. & Friston, K. Repetition suppression and its contextual determinants in predictive coding. *Cortex* **80**, 125–140 (2016).
13. Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M. & Egner, T. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* **11**, 1004–1006 (2008).
14. Tang, M. F., Smout, C. A., Arabzadeh, E. & Mattingley, J. B. Prediction error and repetition suppression have distinct effects on neural representations of visual information. *Elife* **7**, e33123 (2018).
15. Todorovic, A. & de Lange, F. P. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J. Neurosci.* **32**, 13389–13395 (2012).
16. Bidet-Caulet, A. *et al.* Dynamics of anticipatory mechanisms during predictive context processing. *Eur. J. Neurosci.* **36**, 2996–3004 (2012).
17. Darriba, Á. & Waszak, F. Predictions through evidence accumulation over time. *Sci. Rep.* **8**, 494 (2018).
18. Domenech, P. & Dreher, J.-C. Decision threshold modulation in the human brain. *J. Neurosci.* **30**, 14305–14317 (2010).
19. Hsu, Y.-F., Bars, S. L., Hämäläinen, J. A. & Waszak, F. Distinctive representation of mispredicted and unpredicted prediction errors in human electroencephalography. *J. Neurosci.* **35**, 14653–14660 (2015).
20. Friston, K. What is optimal about motor control? *Neuron* **72**, 488–498 (2011).
21. Hommel, B., Müssele, J., Aschersleben, G. & Prinz, W. The theory of event coding (TEC): A framework for perception and action planning. *Behav. Brain Sci.* **24**, 849–878 (2001).
22. Shin, Y. K., Proctor, R. W. & Capaldi, E. J. A review of contemporary ideomotor theory. *Psychol. Bull.* **136**, 943–974 (2010).
23. Hommel, B. Theory of event coding (TEC) V2.0: Representing and controlling perception and action. *Atten. Percept. Psychophys.* **81**, 2139–2154 (2019).
24. Roussel, C., Hughes, G. & Waszak, F. Action prediction modulates both neurophysiological and psychophysical indices of sensory attenuation. *Front. Hum. Neurosci.* <https://doi.org/10.3389/fnhum.2014.00115> (2014).
25. Waszak, F., Cardoso-Leite, P. & Hughes, G. Action effect anticipation: Neurophysiological basis and functional consequences. *Neurosci. Biobehav. Rev.* **36**, 943–959 (2012).
26. Dogge, M., Custers, R. & Aarts, H. Moving forward: On the limits of motor-based forward models. *Trends Cogn. Sci.* **23**, 743–753 (2019).
27. Lange, K. The ups and downs of temporal orienting: A review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Front. Hum. Neurosci.* <https://doi.org/10.3389/fnhum.2013.00263> (2013).
28. Näätänen, R. & Picton, T. The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology* **24**, 375–425 (1987).
29. Picton, T. Hearing in time: Evoked potential studies of temporal processing. *Ear Hear.* **34**, 385–401 (2013).
30. Horváth, J. Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Res.* **1626**, 54–65 (2015).
31. Crowley, K. E. & Colrain, I. M. A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clin. Neurophysiol.* **115**, 732–744 (2004).
32. Tremblay, K., Ross, B., Inoue, K., McClannahan, K. & Collet, G. Is the auditory evoked P2 response a biomarker of learning? *Front. Syst. Neurosci.* <https://doi.org/10.3389/fnsys.2014.00028> (2014).
33. Bekinschtein, T. A. *et al.* Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1672–1677 (2009).
34. Horváth, J. The role of mechanical impact in action-related auditory attenuation. *Cogn. Affect. Behav. Neurosci.* **14**, 1392–1406 (2014).
35. Knolle, F., Schröger, E. & Kotz, S. A. Prediction errors in self- and externally-generated deviants. *Biol. Psychol.* **92**, 410–416 (2013).
36. Sanmiguel, I., Todd, J. & Schröger, E. Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. *Psychophysiology* **50**, 334–343 (2013).
37. van Elk, M., Salomon, R., Kannape, O. & Blanke, O. Suppression of the N1 auditory evoked potential for sounds generated by the upper and lower limbs. *Biol. Psychol.* **102**, 108–117 (2014).
38. Korka, B., Schröger, E. & Widmann, A. Action intention-based and stimulus regularity-based predictions: Same or different? *J. Cogn. Neurosci.* **31**, 1917–1932 (2019).
39. Timm, J., SanMiguel, I., Saupe, K. & Schröger, E. The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neurosci.* **14**, 2 (2013).
40. Timm, J., SanMiguel, I., Keil, J., Schröger, E. & Schönwiesner, M. Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *J. Cogn. Neurosci.* **26**, 1481–1489 (2014).
41. Näätänen, R., Simpson, M. & Loveless, N. E. Stimulus deviance and evoked potentials. *Biol. Psychol.* **14**, 53–98 (1982).
42. Näätänen, R. & Picton, T. W. N2 and automatic versus controlled processes. *Electroencephalogr. Clin. Neurophysiol. Suppl.* **38**, 169–186 (1986).
43. Folstein, J. R. & Petten, C. V. Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology* **45**, 152–170 (2008).
44. Holroyd, C. A note on the oddball N200 and the feedback ERN. In *Errors, Conflicts, and the Brain: Current Opinions on Performance Monitoring* (ed. Falkenstein, M.) 211–218 (MPI of Cognitive Neuroscience, 2004).
45. Kanske, P. & Kotz, S. A. Modulation of early conflict processing: N200 responses to emotional words in a flanker task. *Neuropsychologia* **48**, 3661–3664 (2010).
46. Näätänen, R. & Gaillard, A. W. K. 5 The orienting reflex and the N2 deflection of the event-related potential (ERP). In *Advances in Psychology* Vol. 10 (eds Gaillard, A. W. K. & Ritter, W.) 119–141 (North-Holland, 1983).
47. Le Bars, S., Darriba, Á. & Waszak, F. Event-related brain potentials to self-triggered tones: Impact of action type and impulsivity traits. *Neuropsychologia* **125**, 14–22 (2019).
48. Polich, J. Updating P300: An integrative theory of P3a and P3b. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **118**, 2128–2148 (2007).
49. Ehinger, B. V., König, P. & Ossandón, J. P. Predictions of visual content across eye movements and their modulation by inferred information. *J. Neurosci.* **35**, 7403–7413 (2015).
50. Feldman, H. & Friston, K. J. Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* <https://doi.org/10.3389/fnhum.2010.00215> (2010).
51. Kolossa, A., Kopp, B. & Fingscheidt, T. A computational analysis of the neural bases of Bayesian inference. *Neuroimage* **106**, 222–237 (2015).
52. Donchin, E. & Coles, M. G. H. Is the P300 component a manifestation of context updating? *Behav. Brain Sci.* **11**, 357–374 (1988).
53. Sawaki, R. & Katayama, J. Stimulus context determines whether non-target stimuli are processed as task-relevant or distractor information. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **117**, 2532–2539 (2006).
54. O’Connell, R. G., Dockree, P. M. & Kelly, S. P. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* **15**, 1729–1735 (2012).

55. Mars, R. B. *et al.* Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *J. Neurosci.* **28**, 12539–12545 (2008).
56. Summerfield, C. & de Lange, F. P. Expectation in perceptual decision making: Neural and computational mechanisms. *Nat. Rev. Neurosci.* **15**, 745–756 (2014).
57. Wacongne, C. *et al.* Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci.* **108**, 20754–20759 (2011).
58. Arnal, L. H. & Giraud, A.-L. Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* **16**, 390–398 (2012).
59. Delorme, A. & Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
60. Widmann, A., Schröger, E. & Maess, B. Digital filter design for electrophysiological data—A practical approach. *J. Neurosci. Methods* **250**, 34–46 (2015).
61. Nolan, H., Whelan, R. & Reilly, R. B. FASTER: Fully automated statistical thresholding for EEG artifact rejection. *J. Neurosci. Methods* **192**, 152–162 (2010).
62. Mogron, A., Jovicich, J., Bruzzone, L. & Buiatti, M. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* **48**, 229–240 (2011).
63. Chaumon, M., Bishop, D. V. M. & Busch, N. A. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *J. Neurosci. Methods* **250**, 47–63 (2015).
64. Kayser, J. & Tenke, C. E. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: II. Adequacy of low-density estimates. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **117**, 369–380 (2006).
65. Perrin, F., Pernier, J., Bertrand, O., Giard, M. H. & Echallier, J. F. Mapping of scalp potentials by surface spline interpolation. *Electroencephalogr. Clin. Neurophysiol.* **66**, 75–81 (1983).
66. Horváth, J., Roeber, U., Bendixen, A. & Schröger, E. Specific or general? The nature of attention set changes triggered by distracting auditory events. *Brain Res.* **1229**, 193–203 (2008).
67. Toninquist-Uhlen, I., Ponton, C. W., Eggermont, J. J., Kwong, B. & Don, M. Maturation of human central auditory system activity: The T-complex. *Clin. Neurophysiol.* **114**, 685–701 (2003).
68. Verleger, R., Cäsar, S., Siller, B. & Śmigasiwicz, K. On Why Targets Evoke P3 Components in prediction tasks: Drawing an analogy between prediction and matching tasks. *Front. Hum. Neurosci.* <https://doi.org/10.3389/fnhum.2017.00497> (2017).
69. Bürkner, P.-C. brms: An R package for bayesian multilevel models using stan. *J. Stat. Softw.* **80**, 1–28 (2017).
70. Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
71. RCore, T. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, 2016).
72. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
73. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
74. Baayen, R. H. & Milin, P. Analyzing reaction times. *J. Psychol. Res.* <https://doi.org/10.21500/20112084.807> (2010).
75. Baayen, R. H. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R* (Cambridge University Press, 2008).
76. Hughes, G., Desantis, A. & Waszak, F. Attenuation of auditory N1 results from identity-specific action-effect prediction. *Eur. J. Neurosci.* **37**, 1152–1158 (2013).
77. Escera, C., Leung, S. & Grimm, S. Deviance detection based on regularity encoding along the auditory hierarchy: Electrophysiological evidence in humans. *Brain Topogr.* **27**, 527–538 (2014).
78. Horváth, J., Winkler, I. & Bendixen, A. Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol. Psychol.* **79**, 139–147 (2008).
79. Grimm, S. & Escera, C. Auditory deviance detection revisited: Evidence for a hierarchical novelty system. *Int. J. Psychophysiol.* **85**, 88–92 (2012).
80. Knolle, F., Schwartze, M., Schröger, E. & Kotz, S. A. Auditory predictions and prediction errors in response to self-initiated vowels. *Front. Neurosci.* <https://doi.org/10.3389/fnins.2019.01146> (2019).
81. Knolle, F., Schröger, E., Baess, P. & Kotz, S. A. The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *J. Cogn. Neurosci.* **24**, 698–706 (2012).
82. Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H. & Baldeweg, T. Event-related brain potential correlates of human auditory sensory memory-trace formation. *J. Neurosci.* **25**, 10494–10501 (2005).
83. Duncan-Johnson, C. C. & Donchin, E. The P300 component of the event-related brain potential as an index of information processing. *Biol. Psychol.* **14**, 1–52 (1982).
84. Iwanaga, M. & Nittono, H. Unexpected action effects elicit deviance-related brain potentials and cause behavioral delay. *Psychophysiology* **47**, 281–288 (2010).
85. Friedman, D., Cycowicz, Y. M. & Gaeta, H. The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci. Biobehav. Rev.* **25**, 355–373 (2001).
86. Schröger, E., Marzecová, A. & SanMiguel, I. Attention and prediction in human audition: A lesson from cognitive psychophysiology. *Eur. J. Neurosci.* **41**, 641–664 (2015).
87. Chang, A., Bosnyak, D. J. & Trainor, L. J. Beta oscillatory power modulation reflects the predictability of pitch change. *Cortex* **106**, 248–260 (2018).
88. Max, C., Widmann, A., Schröger, E. & Sussman, E. Effects of explicit knowledge and predictability on auditory distraction and target performance. *Int. J. Psychophysiol.* **98**, 174–181 (2015).
89. Bäss, P., Jacobsen, T. & Schröger, E. Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **70**, 137–143 (2008).
90. Behroozmand, R., Liu, H. & Larson, C. R. Time-dependent neural processing of auditory feedback during voice pitch error detection. *J. Cogn. Neurosci.* **23**, 1205–1217 (2011).
91. Jones, A., Hughes, G. & Waszak, F. The interaction between attention and motor prediction. An ERP study. *Neuroimage* **83**, 533–541 (2013).
92. Ullsperger, M., Fischer, A. G., Nigbur, R. & Endrass, T. Neural mechanisms and temporal dynamics of performance monitoring. *Trends Cogn. Sci.* **18**, 259–267 (2014).
93. Kelly, S. P. & O'Connell, R. G. Internal and external influences on the rate of sensory evidence accumulation in the human brain. *J. Neurosci.* **33**, 19434–19441 (2013).
94. Fischer, A. G. & Ullsperger, M. Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron* **79**, 1243–1255 (2013).
95. Verleger, R., Hamann, L. M., Asanowicz, D. & Śmigasiwicz, K. Testing the S-R link hypothesis of P3b: The oddball effect on S1-evoked P3 gets reduced by increased task relevance of S2. *Biol. Psychol.* **108**, 25–35 (2015).
96. Hughes, G. & Waszak, F. ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. *Neuroimage* **56**, 1632–1640 (2011).
97. Verleger, R., Metzner, M. F., Ouyang, G., Śmigasiwicz, K. & Zhou, C. Testing the stimulus-to-response bridging function of the oddball-P3 by delayed response signals and residue iteration decomposition (RIDE). *Neuroimage* **100**, 271–280 (2014).

98. Pieszek, M., Widmann, A., Gruber, T. & Schröger, E. The human brain maintains contradictory and redundant auditory sensory predictions. *PLoS ONE* **8**, e53634 (2013).
99. Horváth, J., Czigler, I., Sussman, E. & Winkler, I. Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. *Brain Res. Cogn. Brain Res.* **12**, 131–144 (2001).
100. Winkler, I., Denham, S. L. & Nelken, I. Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* **13**, 532–540 (2009).
101. Press, C., Kok, P. & Yon, D. The perceptual prediction paradox. *Trends Cogn. Sci.* **24**, 13–24 (2020).

Acknowledgements

This research was funded by the French Agence Nationale de la Recherche (ANR) within the *Programme franco-allemand en Sciences humaines et sociales (FRAL)* 2016. PROJECT ID: ANR-16-FRAL-0008. This research was partly supported by IdEx Université de Paris ANR-18-IDEX-0001. This work was supported by Taiwan Ministry of Science and Technology (Grant Numbers MOST109-2636-H-003-001 and MOST110-2636-h-003-001) to YFH. We thank Miss S-H Lee and Mr C-A Tu for assistance with data collection.

Author contributions

Á.D.: Conceptualization, Methodology, Formal Analysis, Writing—original draft, Writing—review & editing, Visualization. Y.-F.H.: Conceptualization, Investigation, Resources, Writing—review & editing, Funding acquisition. S.V.O.: Methodology, Software, Formal Analysis, Writing—review & editing, Visualization. F. W.: Conceptualization, Writing—review & editing, Supervision, Funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-99445-z>.

Correspondence and requests for materials should be addressed to Á.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021