NeuroImage 201 (2019) 116011

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making

Jan Herding [a,b,*], Simon Ludwig [a], Alexander von Lautz [a,b], Bernhard Spitzer [c], Felix Blankenburg [a,b,c]

[a] Neurocomputation and Neuroimaging Unit, Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195, Berlin, Germany
[b] Bernstein Center for Computational Neuroscience Berlin, Philippstr. 13, 10115, Berlin, Germany
[c] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

ARTICLE INFO

Keywords:
Decision making
EEG
P300
Confidence
Vibrotactile

ABSTRACT

Recent studies suggest that a centro-parietal positivity (CPP) in the EEG signal tracks the absolute (unsigned) strength of accumulated evidence for choices that require the integration of noisy sensory input. Here, we investigated whether the CPP might also reflect the evidence for decisions based on a quantitative comparison between two sequentially presented stimuli (a signed quantity). We recorded EEG while participants decided whether the latter of two vibrotactile frequencies was higher or lower than the former in six variants of this task (n = 116). To account for biases in sequential comparisons, we applied a behavioral model based on Bayesian inference that estimated subjectively perceived frequency differences. Immediately after the second stimulus, parietal ERPs reflected the signed value of subjectively perceived differences and afterwards their absolute value. Strikingly, the modulation by signed difference was evident in trials without any objective evidence for either choice and correlated with choice-selective premotor beta band amplitudes. Modulations by the absolute strength of subjectively perceived evidence – a direct indicator of task difficulty - exhibited all features of statistical decision confidence. Together, our data suggest that parietal EEG signals first index subjective evidence, and later include a measure of confidence in the context of perceptual decision making.

1. Introduction

Recent studies have suggested a centro-parietal positivity (CPP) in the EEG signal (arguably identical to the classic P300 component) as a modality-independent proxy of accumulated evidence in perceptual decision making tasks (e.g., Kelly & O'Connell, 2015; Philiastides et al., 2014). In particular, when classifying a noisy sensory stimulus interval into one of two categories, the CPP increased faster and peaked earlier the weaker the noise was, i.e., the clearer the presented evidence (e.g., random dot motion (RDM) discrimination: Kelly & O'Connell, 2013; face-vs-car discrimination: Philiastides et al., 2014). Moreover, the CPP reached a fixed threshold at the time of the decision report, suggesting an accumulation-to-bound mechanism for response initiation (e.g., O'Connell et al., 2012; but see Philiastides et al., 2014). Together, these findings capture the hallmarks of popular sequential-sampling models of evidence accumulation (e.g., see Smith and Ratcliff, 2004), and may relate to similar, or even homologue neuronal processes as identified in the parietal cortex of non-human primates (e.g., Roitman and Shadlen, 2002; Gold and Shadlen, 2007; Shadlen and Kiani, 2013).

The link between decisional evidence and the CPP is not limited to decisions that require the accumulation of noisy sensory input over time. In an auditory four-stimulus oddball paradigm, the differences between 'deviant' and 'standard' stimuli (i.e., the evidence for a 'deviant' detection) modulated the CPP in the very same way as it was modulated by the strength of evidence in accumulation-based decisions (Twomey et al., 2015). Notably, the three 'deviant' stimuli in this task were always higher in pitch than the 'standard' stimulus, eliminating the necessity for participants to evaluate the sign of the difference (i.e., higher or lower) between 'deviant' and 'standard'.

In all of the aforementioned EEG studies, a parietal potential tracked the strength of evidence during perceptual decision making, however, without indicating for which choice alternative (i.e., unsigned evidence; e.g., Kelly and O'Connell, 2013; Philiastides et al., 2014). In the RDM task for instance, only the proportion of coherently moving dots

* Corresponding author. Neurocomputation and Neuroimaging Unit, Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195, Berlin, Germany.
E-mail address: jan.herding@bccn-berlin.de (J. Herding).

https://doi.org/10.1016/j.neuroimage.2019.116011
Received 16 November 2018; Received in revised form 9 July 2019; Accepted 10 July 2019
Available online 11 July 2019
1053-8119/

modulated the CPP, without differentiating between the directions in which the dots moved (Kelly and O'Connell, 2013). Here, we examined whether the CPP might also index the choice alternative, in addition to the strength of evidence, if we apply a sequential comparison task. In particular, does the CPP indicate the decision-relevant signed evidence for choices that involve a quantitative comparison? We used a classic vibrotactile two-alternative forced choice (2-AFC) task, in which participants compare two stimulus frequencies (f1 and f2) and decide whether the second one was higher or lower than the first one (comprehensive review on monkey electrophysiology in Romo and de Lafuente, 2013). In this paradigm, a choice-specific (i.e., binary) modulation of upper beta band (~20–30 Hz) amplitude in premotor cortex, decoupled from the motor response, was recently observed in human EEG recordings (Herding et al., 2016, 2017; Ludwig et al., 2018), replicating previous findings from monkey LFPs (Haegens et al., 2011, 2017). A representation of the graded differences between f1 and f2 (i.e., the signed evidence), however, has not yet been identified in the human EEG. For the current study, we pooled EEG data over six experiments, utilizing the same vibrotactile 2-AFC task while varying response modality, response timing, and response mapping (N = 116). We estimated subjective evidence and difficulty (i.e., the subjectively perceived signed and absolute difference, respectively, between f1 and f2) using a Bayesian inference model of choice behavior. This way, we accounted for known biases in sequential comparisons due to the so-called contraction-bias (e.g., Jou et al., 2004; Ashourian et al., 2011; Karim et al., 2012; Raviv et al., 2014) which is a direct consequence of the time-order effect/error (TOE; see Fechner, 1860; Woodrow, 1935; Hellström, 1985, 2003). Moreover, the behavioral model allowed us to derive a measure of confidence grounded in statistical decision theory (i.e., statistical decision confidence) obviating the need for subjective confidence reports. Using the behavioral model, we found that signals from the parietal cortex appeared to be first modulated by the subjectively perceived signed difference, and later by the absolute value (i.e., the absolute strength of evidence).

2. Materials and Methods

2.1. Experimental design

Participants: A total of 129 datasets were obtained from healthy, right-handed volunteers (21–40 years; 76 females) who participated in six different variants of the experiment. Most participants were students from the Freie Universität Berlin, and some participated in more than one variant of the experiment. All studies were approved by the local ethics committee at the Freie Universität Berlin, and participants gave written informed consent before an experiment started. Thirteen datasets were excluded due to chance-level behavioral performance (<55% correct answers) and/or excessive EEG artifacts, leaving 116 datasets for further analyses.

Stimuli and behavioral task: In all six variants of the experiment, stimuli and comparison task were identical. Only the response modality and response timing varied across experiments (Fig. 1). Supra-threshold vibrotactile stimuli with constant peak amplitude were applied to the left index finger using a piezoelectric Braille stimulator (QuaeroSys Medical Devices, Schotten, Germany). The stimuli consisted of amplitude-modulated sinusoids with a fixed carrier frequency of 133 Hz (137 Hz in Experiment 2). Amplitude-modulation of this carrier signal with frequencies between 12 and 32 Hz was used to create the sensation of tactile 'flutter' (see Talbot et al., 1968; Romo and Salinas, 2003), while limiting the spectrum of the physical driving signal to frequencies above 100 Hz (e.g., Tobimatsu et al., 1999). Thus, the risk of physical artifacts in the EEG analysis range of interest (<100 Hz) was minimized. The sound of the stimulator was masked by white noise of ~80 dB that was played throughout the experiment (e.g., see Spitzer et al., 2010; Spitzer and Blankenburg, 2011). Participants were comfortably seated ~60 cm in front of a TFT monitor. A fixation cross was displayed at the center of the screen to minimize eye movements. On each trial, two flutter stimuli were successively presented for 250 ms each (with frequencies f1 and f2), interleaved by a retention interval of 1000 ms (see Fig. 1). The frequencies of the first stimulus (f1) were randomly drawn from 16, 20, 24 or 28 Hz, whereas f2 differed from f1 by ± 2 or 4 Hz. In four variants of the experiment (Experiments 3–6), f2 was identical to f1 in 25% of the trials, without participants knowing. Participants were instructed to always decide whether f2 > f1 or f2 < f1.

In Experiments 1 and 2, participants indicated choices immediately after presentation of the second stimulus either by pressing one of two buttons with the right index or middle finger (Experiment 1), or by making a saccade to one of two target dots (Experiment 2). The target dots (diameter of ~0.5° visual angle) appeared on the left and on the right side of the screen (~12° visual angle off-center). Importantly, the response assignment of the two buttons and of the two saccade directions was reversed for half of the participants. This way, the mapping of choices onto specific motor responses (which might have been associated with specific motor preparatory signals) was fully counterbalanced across participants (see also Herding et al., 2016, 2017). In Experiments 3 and 4, participants reported choices analogously to Experiments 1 and 2, however, only after a delay of 2500 ms. In Experiments 5 and 6, an additional mapping of choices onto a color-code (blue vs. yellow) was required to report decisions after the delay. In the experiments with delayed responses (Experiments 3–6), 2000 ms after the presentation of f2, a blue and a yellow target dot (diameter of ~1° visual angle) appeared on the left and on the right side of the screen (fully counterbalanced across trials; ~12° visual angle off-center). In Experiments 3 and 4, the colors of the dots were irrelevant, and participants selected targets based on a fixed association between direction and choices (counterbalanced across participants). In Experiments 5 and 6, each color was associated with one of the two choice options (counterbalanced across participants). Participants selected a target based on its location (Experiments 3 and 4) or color (Experiments 5 and 6) after another 500 ms either by pressing the left-arrow or right-arrow button with the right index or middle finger (Experiments 3 and 5, see Ludwig et al., 2018), or by making a saccade onto the target (Experiments 4 and 6). See Fig. 1 for a graphical summary of the experimental designs.

In Experiments 1 and 2, participants received performance feedback after each trial, and completed seven blocks of 160 f1-vs-f2 comparisons (each block lasted ~15 min including eye-tracker calibration) for a total of 1120 trials. In Experiments 3–6, feedback based on the performance for trials with f1 ≠ f2 was provided after each block, and participants completed eight blocks of 128 frequency comparisons (each block lasted ~12 min including eye-tracker calibration) for a total of 1024 trials. Before each experiment, participants performed ~50 practice trials.

Note that the influence of the different response conditions was not subject to the current study. Oscillatory signatures in the EEG signal that are related to these response manipulations have been reported elsewhere (Herding et al., 2016, 2017; Ludwig et al., 2018).

Eye-tracking: In Experiment 2, a Tobii T60 eye-tracker (Tobii Technology, Danderyd, Sweden) was used to record eye movements of participants during each trial (binocular sampling at 60 Hz). The T60 is integrated into a 17″ TFT monitor and is able to track participants that are comfortably seated in front of the monitor (i.e., no chin rest required). In Experiments 4 and 6, eye movements were recorded (monocular sampling at 500 Hz) using an EyeLink 1000 Desktop Mount with a chin rest (SR Research, Ottawa, Canada). Online evaluation of the participants' gaze directions was implemented with custom code using the Tobii toolbox and psychtoolbox 3 for MATLAB (Brainard, 1997; Cornelissen et al., 2002). Thus, we were able to monitor that participants kept the gaze on the central fixation cross during each trial (with tolerance of ~3° visual angle) and displayed a warning message if this was not the case ("Please keep fixation throughout the trial"). Additionally, we read out participants' choices (200 ms fixation on target dot) and provided performance feedback online, either after each trial (experiment 2) or after each block (experiments 4 and 6). To maintain a high tracking accuracy, the eye-tracker was calibrated before the beginning of each block

using a standard 5-dot (Tobii T60) or 9-dot (EyeLink 1000) calibration procedure.

2.2. Statistical analysis

Behavioral model of choices and confidence: In order to explain the observed choice pattern, we fitted a Bayesian inference model to individual behavioral data, and thereby, estimated subjectively perceived frequency differences (SPFDs; Fig. 2A, for details see Herding et al., 2016; see also Ashourian and Loewenstein, 2011; Sanchez, 2014). In brief, the model targets to account for a known bias in sequential comparisons (e.g., see Hellström, 1985, 2003; Ashourian and Loewenstein, 2011). That is, participants tend to compare f2 not only with the physical value of f1, but also with the average frequency of all presented stimuli, as if the representation of f1 is subject to a contraction towards the mean (hence, the term 'contraction bias', see e.g., Jou et al., 2004; Preuschhof et al., 2010; Ashourian and Loewenstein, 2011; Karim et al., 2012; Raviv et al., 2014; see time-order effect for underlying core principle: e.g., Fechner, 1860; Woodrow, 1935, Hellström, 1985). In other words, the quantity that drives choices in the given task is best described by the difference between f2 and a representation of f1 that deviates from its physical value toward the mean frequency of the stimulus set. In our model, we introduce this shifted quantity – which we will call f1' – as a weighted average of the mean of all stimulus frequencies and the physical value of f1 – implemented in terms of Bayesian inference. In particular, f1′ is the expected value of the Gaussian posterior distribution of f1, assuming a Gaussian prior centered on the frequencies of the stimulus set (see Fig. 2A). The model was fitted to the choices of individual participants by optimizing three free parameters (i.e., variance of stimulus likelihood σ^2_{stim}, prior variance σ^2_{prior}, and a decision criterion *c*) using variational Bayes as implemented in the VBA toolbox (Daunizeau et al., 2014). In order to assess the model's goodness-of-fit, we computed Bayes Factors (BFs) to compare each model fit with a "null" model in which decisions were based on the physical stimulus differences (i.e., f2 – f1). Notably, the model of SPFDs as well as the "null" model followed Weber-Fechner's law and implemented the representation of frequency values on a logarithmic scale (see Herding et al., 2016).

Based on the individual model fits, we quantified the SPFD for each stimulus pair by the difference f2 – f1′, yielding 16 SPFD values for Experiments 1 and 2, and 20 SPFD values for Experiments 3–6. At the same time, the difference distribution between the likelihood of f2 and the posterior of f1 (centered on f1') additionally allowed us to compute a measure of confidence based on statistical decision theory (e.g., Drugowitsch, 2016; Hangya et al., 2016; Sanders et al., 2016, Fig. 2B). The difference distribution describes the distribution of percepts that are associated with a given stimulus pair, i.e., with the SPFD between both stimuli. According to statistical decision theory (or signal detection theory), a single percept can be conceived as a sample *d* from this distribution, and a choice based on this very percept depends on where the sample is located with respect to a decision criterion *c* (i.e., choose f2 > f1 if $d > c$). The distance between the sample and the criterion (i.e., $|d - c|$) can be transformed into the probability of a correct response given the percept *d*, which in turn is a measure for confidence (Lak et al., 2014; Urai et al., 2017, Fig. 2B). For each participant, we estimated average confidence based on this approach for SPFDs on the interval [-0.4, 0.4]. For each SPFD on this interval, we drew 100,000 samples from the individual difference distributions (i.e., based on the estimated parameters), and computed the associated confidence for each sample. Confidence values were then averaged separately for correct and incorrect trials. Since results were roughly symmetric across zero, the average confidence was grouped according to absolute values of SPFDs (e.g., −0.2 and 0.2), and respective mean values were computed. The illustration in Fig. 2B was obtained by simulating data from an unbiased observer (c = 0) with $\sigma^2_{stim} = 0.05$ and $\sigma^2_{prior} = 0.347$.

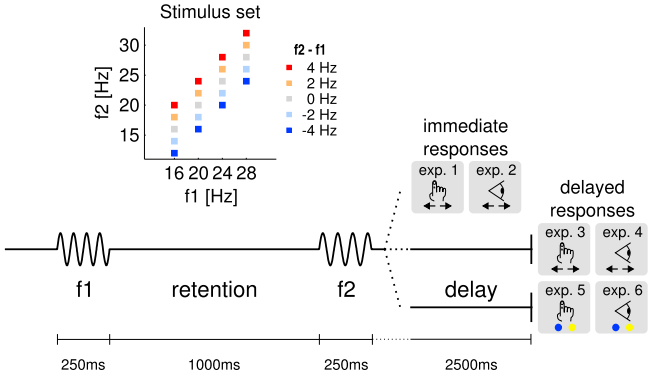

Fig. 1. Task and stimuli. One after another, two vibrotactile stimuli with frequencies f1 and f2 were briefly presented to the left index finger of participants who had to decide whether f2 > f1 or f2 < f1. Response timing (immediate/delayed), response modality (saccade/button press), and response mapping (direction/color) varied over six variants of the task (exp. 1 – exp. 6). **Inset**, The stimulus set that was used in all experiments, with the exception of zero-difference trials (gray) which were not used in exp. 1 and exp. 2. Each square represents one stimulus pair with f1 (x-axis) and f2 (y-axis). The color-code denotes the physical stimulus differences f2 – f1.

EEG recording and analysis: In all experiments, EEG (ActiveTwo; BioSemi) was recorded at 2048 Hz (offline down-sampled to 512 Hz) from 64 electrodes positioned in an elastic cap according to the extended 10–20 system. Individual electrode locations for each participant were obtained prior to the experiments using a stereotactic electrode-positioning system (Zebris Medical GmbH, Isny, Germany). Additional electrodes were used to register the horizontal and vertical electrooculogram (hEOG and vEOG). For preprocessing, EEG data were high- and low-pass filtered using a non-causal FIR filter (with cut-off frequencies of 0.1 and 30 Hz, respectively), and re-referenced to a common average montage. Eye blink artifacts in the EEG data were corrected using adaptive spatial filtering based on individual calibration data informed by the vEOG signal (see Ille et al., 2002). For experiment 2, in which participants gave immediate responses by saccades, we used the same approach informed by the hEOG signal to remove artifacts of horizontal saccades from the EEG signal. The artifact-free EEG data were segmented into epochs from −2250 to 2000 ms relative to the presentation time of the second stimulus in order to examine evoked EEG responses after the second stimulus as well as to compute control analyses after the first stimulus. Noisy trials were identified by careful visual inspection and were excluded from further analysis (14.8% of trials on average). The remaining single-trial data were baseline-corrected relative to the 100 ms preceding stimulus onset. All analyses were done in MATLAB (The MathWorks) using custom code, functions of the SPM12 toolbox (Wellcome Department of Cognitive Neurology, London; www.fil.ion.ucl.ac.uk/spm), and the FieldTrip toolbox for EEG/MEG data (Radboud University Nijmegen, Donders Institute; fieldtrip.fcdonders.nl).

Multiple regression and group-level analysis: For each participant, we implemented a multiple regression analysis of the preprocessed single-trial EEG data. At each time point, we regressed the EEG data onto the SPFDs (i.e., f2 – f1′) and their absolute values (i.e., |f2 – f1'|) over trials, separately for correct and incorrect choices. The resulting regression coefficients quantified how strongly the trial-specific values of the regressors (i.e., f2 – f1′ and |f2 – f1'|) were related to trial-by-trial variability in the EEG data. To identify time periods and channels for which this relation was consistently different from zero across participants, we used cluster-based permutation testing (Maris and Oostenveld, 2007). That is, we compared the summary statistics of the observed data (one-sample *t*-test of regression coefficients across all data sets at each time point) with a distribution of summary statistics obtained from 500 randomly sign-flipped permutations. A cluster was defined as a group of adjacent time points that all exceeded a cluster-defining threshold of $p_{threshold} < 0.005$ (uncorrected). Clusters that exceeded a cluster-based family-wise error (FWE) corrected threshold of $p_{FWE} < 0.05$ (corrected

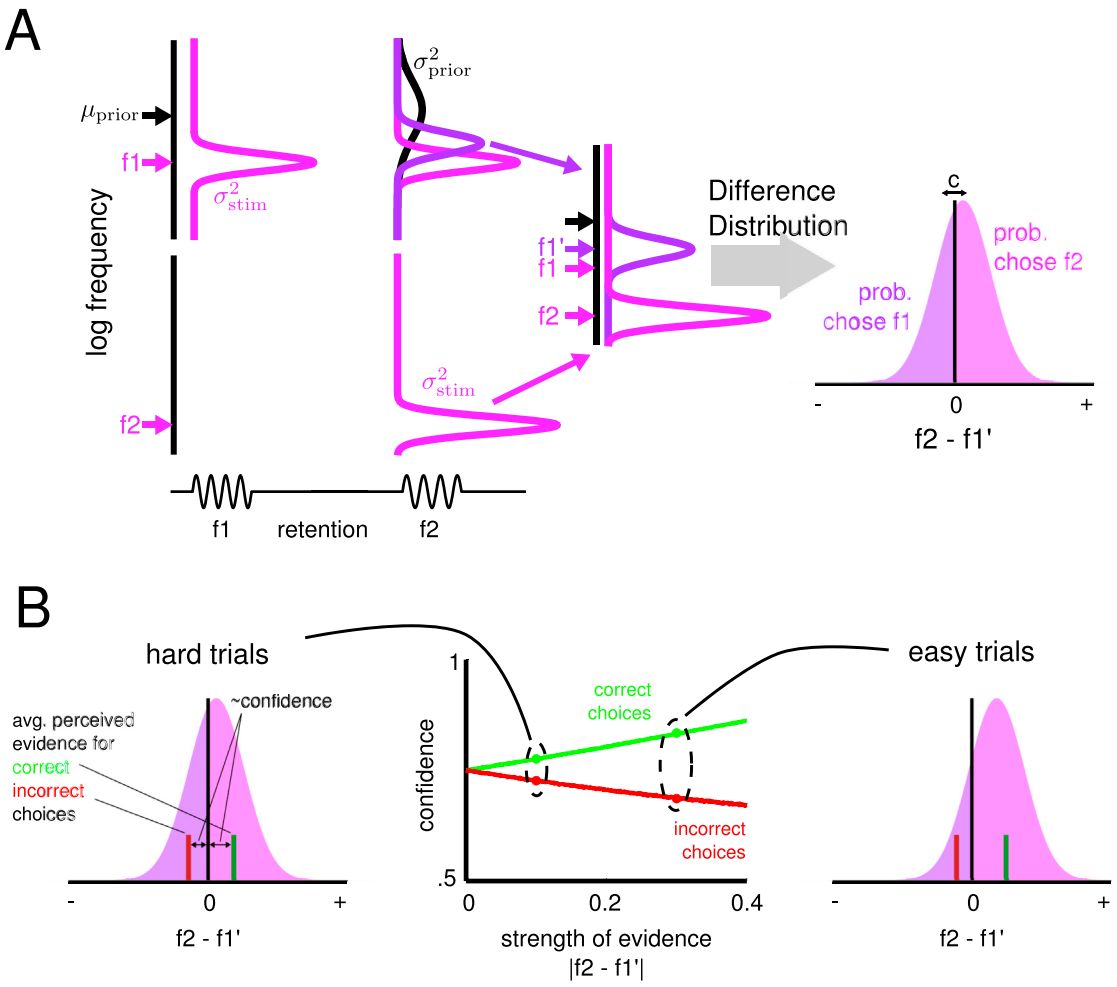

Fig. 2. Behavioral model for choices and confidence. **A**, Graphical illustration of the behavioral model based on Bayesian inference. Y-axes display frequencies on a logarithmic scale. Top: Representation of f1 during different stages of the task. Pink distribution represents the likelihood function of f1. Black distribution is the prior centered on the stimulus set. Purple distribution is the posterior of f1 with shifted mean f1'. Lower: The likelihood of f2 (pink distribution) is used for the comparison with the posterior of f1. Subtracting the posterior of f1 from the likelihood of f2, yields a difference distribution which is used to fit the probability to choose f1 to the behavioral data of each participant by optimizing σ^2_{stim},σ^2_{prior}, and decision criterion *c*. **B**, Intuition of statistical decision confidence. The distance between perceived evidence and decision criterion is proportional to confidence. Average perceived evidence is displayed separately for correct and incorrect trials (green and red bar, respectively). Difference distribution for hard (f2 – f1' = 0.1) and easy (f2 – f1' = 0.3) trials illustrate that confidence increases with evidence strength for correct trials but decreases for incorrect trials.

for time and channels) were considered statistically significant.

Event-related potentials (ERPs): To visualize the effects identified in the statistical analysis as classic ERPs, we binned the individual 16 values of SPFDs (i.e., differences of log-transformed stimulus frequencies; one per stimulus pair with f1 ≠ f2) into six discrete levels across participants (i.e., [< −0.18]; [-0.18 to −0.09]; [-0.09 to 0]; [0 to 0.09]; [0.09 to 0.17]; [> 0.17]). The grand average ERPs were computed separately for each level. We defined the six levels symmetrically around a SPFD of zero (corresponding to chance-level performance), and in such a way that each participant had at least one stimulus pair per level. Since SPFDs were generally small for trials with identical stimuli (i.e., f1 = f2), we used only four levels for the computation of ERPs in these trials (i.e., [< −0.09]; [-0.09 to 0]; [0 to 0.09]; [> 0.09]). Note that by binning the data into discrete levels of SPFDs, the high precision of utilizing subjective measures for a single-trial analysis is lost for the visualization of the ERPs.

Source reconstruction: The cortical sources of the observed modulations on the scalp-level were localized using the 3D source reconstruction routines provided by SPM12 (Friston et al., 2006). Based on the individually recorded electrode positions for each participant, a forward model was constructed using an 8196-point cortical mesh of distributed dipoles perpendicular to the cortical surface of a template brain (see Friston et al., 2006). The lead field of the forward model was computed using the three-shell Boundary Elements Method (BEM) EEG head model available in SPM12. Multiple sparse priors (Friston et al., 2008) under group constraints (Litvak and Friston, 2008) were applied to invert the forward model. For model inversion, we used a representative time interval (i.e., −200 to 1500 ms relative to f2) of ERPs that were computed separately for each level of SPFDs (see ERPs above) drawing on all trials including those with identical stimuli (i.e., f1 = f2). The results of the inversion were summarized in six corresponding 3D images (i.e., one for each level of SPFDs) that reflected source activity averaged over a time window of interest. In particular, summary images were computed for an early (250–500 ms) and a late (500–800 ms) time window capturing the two effects observed at the scalp level (i.e., modulation by signed evidence and strength of evidence, respectively). For each time window, contrasting the 3D images within each participant analogously to the sensor space analysis served as an estimate for subject-specific source locations of both effects. The results of conventional group-level statistical analyses of these source images (see Litvak et al., 2011) are displayed at a significance level of $p < 0.001$ (uncorrected). Anatomical references for source estimates were established on the basis of the SPM

anatomy toolbox (Eickhoff et al., 2005) where possible.

Single-trial correlation of CPP and upper beta band amplitude: In order to explore the relationship between the CPP and premotor choice-specific upper beta band amplitude (see Herding et al., 2016, 2017; Ludwig et al., 2018), single-trial correlations between these two measures were computed. Notably, only for experiments 1 and 2; as these experiments required immediate responses, and hence, a direct transformation of evidence into a motor response. For each participant, the magnitude of the CPP in every trial was specified by a single value for the early and for the late effect, respectively. In particular, the single-trial EEG signal from electrode CPz was averaged over a brief time period during which a modulation of the CPP by the signed values of SPFDs (i.e., 250–500 ms) or by its absolute values (i.e., 500–800 ms) was observed. Additionally, a measure of the upper beta band amplitude in electrodes over premotor areas was computed for each trial. Using response-locked time-frequency representations of the single-trial data (reported in Herding et al., 2016, 2017), average beta band amplitude was computed over a time-frequency cluster that exhibited a significant modulation by participants' choices (i.e., electrodes FC2, FCz, and C2; 20–30 Hz; −750 to −350 ms from responses for experiment 1, see Herding et al., 2016; electrodes FC2 and FC4, 24–32 Hz, −750 to −450 ms from responses for experiment 2, see Herding et al., 2017). We used correct and incorrect trials to compute the single-trial correlations for each participant. The correlation coefficients from both experiments were pooled ($N = 45$), and a one-sample *t*-test was computed to assess whether a consistent correlation was present across participants.

Data and code availability statement: Data processing and analysis code is available at an online repository at: https://github.com/jherding/SFC_ERP. The data is available upon request by message to the corresponding author. Use and sharing of this data must comply with the General Data Protection (GDPR) of the European Union and with ethics approval of the Free University Berlin.

3. Results

3.1. Behavioral results

Pooled over all experiments, participants made 72.5% correct choices on average. To test whether performance varied across the six experiments and across the different frequency differences, we performed a two-way repeated measures ANOVA on proportions of correct responses (PCRs) with between-subject factor 'Experiment' (experiments 1–6), and within-subject factor 'Frequency Difference' (−4, −2, 2, and 4 Hz stimulus difference). We used logit-transformed PCRs to account for non-normally distributed residuals. The analysis revealed no significant performance differences between experiments (main effect 'Experiment', $p = 0.125$; interaction 'Experiment' x 'Frequency Difference', $p = 0.182$). Within each experiment, PCRs varied significantly with the factor 'Frequency Difference' ($p < 0.001$). For further scrutinization of this effect, we computed post-hoc paired t-tests for each study separately to evaluate the influence of difficulty (±4 Hz vs. ±2 Hz differences), and sign of the frequency differences (positive vs. negative differences). As expected, a larger proportion of trials were judged correctly when the (physical) f2 − f1 frequency difference was ±4 Hz compared with trials where the difference was only ±2 Hz in all experiments (all $p < 0.001$; paired *t*-test; see difficulty effect, Table 1). In experiments 1 and 2, we additionally observed more correct responses for positive compared with negative frequency differences ($p = 0.03$, and $p = 0.002$; paired *t*-test; see sign effect, Table 1) which might be attributed to an observed response bias toward "f2 > f1" choices in these two experiments (mean criterion shifts: 0.116 and 0.126 with $p = 0.029$ and $p = 0.002$; one-sample *t*-test). Moreover, the analysis of median response times (RTs) from these two experiments revealed that participants responded faster when choosing "f2 > f1" as compared to choosing "f2 < f1" for correct choices (Experiment 1: 798 ms vs. 855 ms, $p < 0.001$, paired *t*-test, see Herding et al., 2016; Experiment 2: 548 ms vs. 599 ms, $p = 0.001$, paired *t*-test, see Herding et al., 2017) and vice versa for incorrect choices (Experiment 1: 952 ms vs. 903 ms, $p = 0.014$, paired *t*-test, see Herding et al., 2016; Experiment 2: 665 ms vs. 605 ms, $p = 0.002$, paired *t*-test, see Herding et al., 2017). RTs from Experiments 3–6 did not show any systematic variations and are difficult to interpret, because of the delayed decision reports.

3.2. Bayesian inference model yields good approximations for signed subjective evidence and experienced difficulty

As known from many 2-AFC studies that require the comparison of two sequentially presented stimuli, participants typically show a very particular choice pattern due to the contraction bias/TOE (e.g., Preuschhof et al., 2010; Ashourian et al., 2011; Karim et al., 2012; see squares in Fig. 3A and B), which we also observed in the current data. That is, for trials with f2 > f1 (i.e., f2 - f1 = +2 Hz or +4 Hz), participants performed better with increasing f1, whereas for trials with f2 < f1 (i.e., f2 - f1 = −2 Hz or −4 Hz), the opposite was true (Fig. 3A). In other words, the probability to choose f1 decreased with increasing f1 for all frequency differences (interestingly also for those trials with no frequency difference; Fig. 3B). Our previously proposed Bayesian inference model (Herding et al., 2016) can account for this choice pattern (lines in Fig. 3A and B). Moreover, with the individually estimated SPFDs (i.e., f2 – f1′) we obtained a subjective, fine-grained measure that reliably predicted participants' choices. Hence, we used the signed SPFDs as a proxy for signed subjective evidence towards a decision in this task (Fig. 3C). Computing Bayes factors (BFs) to formally assess the quality of our Bayesian model provided positive evidence ($BF > 3$) in favour of the SPFD model for 91.4% of the participants (106/116), and strong evidence ($BF > 20$) for 87.9% (102/116; Fig. 3E). Accordingly, the absolute values of the SPFDs (i.e., |f2 - f1'|) correlated significantly more with participants PCRs than the absolute values of the physical differences (i.e., |f2 – f1|), rendering SPFDs an improved predictor of subjectively experienced difficulty (paired *t*-test, $p < 0.001$; Fig. 3D).

3.3. Parietal ERP first reflects signed subjective evidence and then absolute strength of evidence

We computed a multiple regression analysis on the total EEG data

Table 1
Proportion of correct responses (PCRs) in % as a function of the physical frequency difference f2 – f1 for each experiment. Mean values ± 95% confidence interval are shown. 'Difficulty effect' compares easy (±4 Hz) and difficult (±2 Hz) trials in a paired *t*-test. 'Sign effect' compares trials with a positive (2 and 4 Hz) and negative (−4 and −2 Hz) frequency difference in a paired *t*-test. PCRs were logit-transformed before testing, due to non-normally distributed residuals.

	Frequency difference of stimuli (f2 – f1) in Hz					
	−4	−2	2	4	difficulty effect	sign effect
Exp.1	74.8 ± 6.3	63.4 ± 5.5	68.9 ± 4.0	85.0 ± 2.9	$p < 0.001$	$p = 0.030$
Exp. 2	75.9 ± 4.4	64.7 ± 3.5	70.8 ± 4.4	86.1 ± 4.3	$p < 0.001$	$p = 0.002$
Exp. 3	74.3 ± 6.1	64.2 ± 6.0	65.2 ± 5.2	78.1 ± 5.7	$p < 0.001$	$p = 0.615$
Exp. 4	77.7 ± 8.8	65.4 ± 7.7	66.2 ± 5.5	79.8 ± 6.6	$p < 0.001$	$p = 0.871$
Exp. 5	78.8 ± 5.5	66.6 ± 4.2	67.8 ± 4.5	81.1 ± 5.8	$p < 0.001$	$p = 0.388$
Exp. 6	74.2 ± 5.9	63.1 ± 4.3	66.9 ± 5.3	80.6 ± 5.0	$p < 0.001$	$p = 0.067$
pooled	**75.9 ± 2.4**	**64.5 ± 2.0**	**67.7 ± 1.9**	**81.8 ± 2.0**	**$p < 0.001$**	**$p < 0.001$**

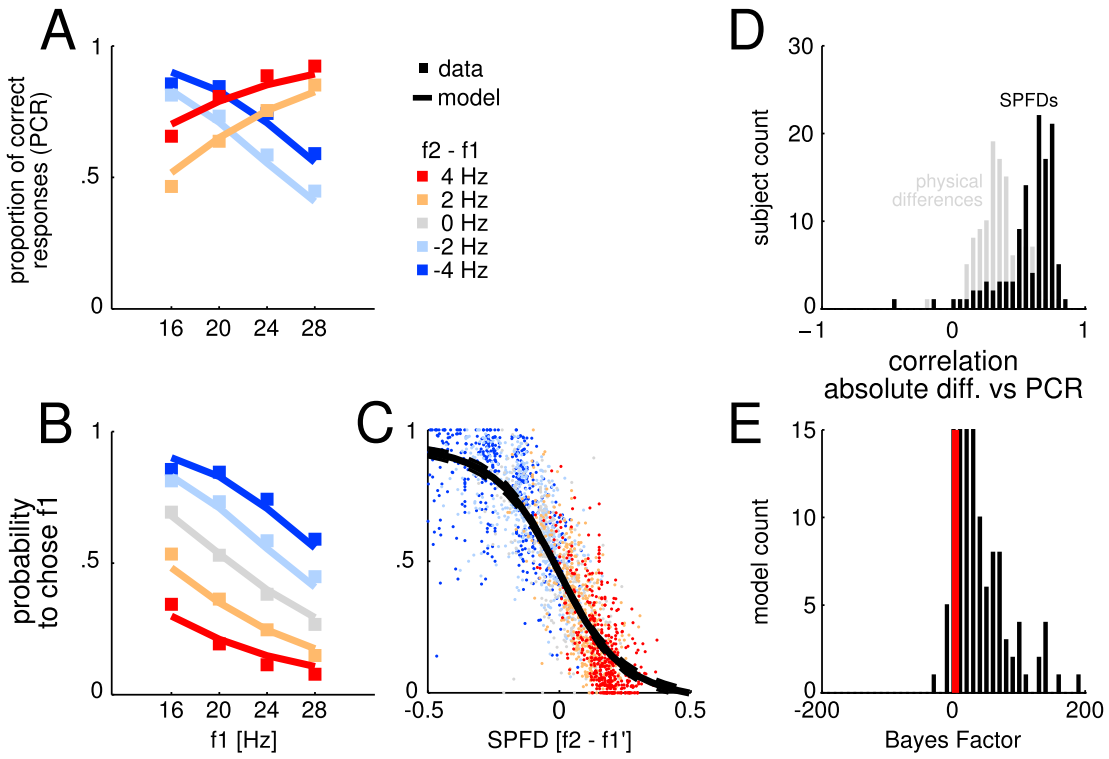

Fig. 3. Behavioral and modeling results. **A**, Grand average of observed (squares) and modeled (lines) proportions of correct responses (PCRs) plotted separately for each f1 (x-axis) and each physical stimulus difference f2 – f1 (color-code). **B**, Same as in A, but for probabilities to choose f1. Note that the blue squares/lines are identical as in A, and the red squares/lines correspond to 1-PCRs from A. **C**, Probabilities to choose f1 for each stimulus pair of each participant (dots), color-coded for physical stimulus differences (f2 – f1), and plotted against subjectively perceived frequency differences (SPFD; f2 – f1′). The solid black line represents the modeled probability to choose f1, averaged over all participants ± 95% confidence interval (dashed lines). **D**, Histogram of correlation coefficients from all participants obtained from correlating absolute physical differences (|f2-f1|) with PCRs (gray), and from correlating absolute values of SPFDs (|f2-f1'|) with PCRs (black). **E**, Histogram of Bayes factors (BFs), comparing the SPFD model with a "null" model (based on physical stimulus differences) for each participant. Red line marks threshold for positive evidence in favor of SPFD model (BF > 3).

using the signed SPFDs (i.e., f2 – f1′) as well as their absolute values (i.e., |f2 – f1'|) as single-trial regressors. This way, we could independently assess correlations of scalp potentials with signed subjective evidence and with the absolute strength of evidence. For a first analysis, we used trials in which objective sensory evidence was present (i.e., physically different stimuli with f2 ≠ f1; >80% of all trials). For correct decisions, we found that a centro-parietal positive ERP after the second stimulus was positively correlated with signed subjective evidence early on (168–709 ms, 35 electrodes, strongest in P1, Pz, CPz, CP4, CP2, and P2 with $p_{FWE} = 0.002$; Fig. 4). Later, however, the same ERP was positively correlated with the absolute strength of the evidence (273–953 ms, 33 electrodes, strongest in P1, CPz, Cz, C2, CP2, and P2 with $p_{FWE} = 0.002$; scalp topographies in Fig. 4). The overall profile of the underlying ERP strongly resembled the classic P300 or CPP (see time courses in Fig. 4). Please note that the time courses of the grand average ERPs per binned SPFD level provide a much coarser view on the underlying effects as compared to the statistical t-maps, because the fine-grained information contained in subject-specific SPFDs is lost for the ERP visualization. For incorrect decisions, the above-mentioned modulations by subjective evidence vanished (all $p_{FWE} > 0.05$), however, the overall profile of the ERP remained unchanged conforming to the shape of a typical P300/CPP. Directly comparing the modulations between correct and incorrect decisions revealed significant differences (i.e., interaction effects) in both the modulation by signed subjective evidence (326–367 ms and 418–455 ms, electrodes P1, P3, P5, PO7, and PO3 with $p_{FWE} = 0.022$ and $p_{FWE} = 0.032$) and in the modulation by absolute strength of evidence (723–750 ms, electrodes CP5, P7, PO7, O1, Iz, and O2 with $p_{FWE} = 0.028$). In sum, parietal ERPs reflected the signed and absolute subjective evidence only for correct trials – and significantly more than for incorrect trials. Hence, a faithful representation of the subjective evidence is tightly linked to correct decisions, implying the behavioral relevance of these effects.

The significant positive correlations in centro-parietal electrodes in correct trials (both with signed and absolute values of SPFDs) were accompanied by significant negative correlations in bilateral fronto-temporal electrodes for all described effects, hinting at the rough orientation of underlying dipole generators (see scalp topographies in Fig. 4). In general agreement with the scalp topographies, the reconstructed source locations suggest that the modulation by signed subjective evidence originates from left superior parietal lobule (SPL; Brodman area 7A; MNI peak coordinates: −24, −62, 54) in the posterior parietal cortex (PPC; Fig. 5). On a considerably lower significance level ($p < 0.05$; uncorrected), also the right SPL is implicated as a likely source. The modulation by the absolute strength of subjective evidence additionally suggested probable sources in bilateral inferior frontal gyrus (IFG, Brodman area 44/45, MNI peak coordinates: −54/+48, 14, 12; Fig. 5).

We challenged our findings in a series of control analyses to exclude confounding factors as the driving forces behind the observed effects. First and foremost, we examined whether the observed modulations of parietal ERPs were driven by the outermost stimulus pairs alone. That is, in the given stimulus set, some choices could have been based on exceptionally high or low f2 alone, possibly associated with qualitatively distinct percepts. We excluded these outermost stimuli from the stimulus set and repeated the multiple regression analysis on the remaining subset of data (see inset Fig. 6). Notably, the subset only included trials in which any f2 could lead to either choice (i.e., f2 alone did not predict the correct decision in these trials), resulting in substantially reduced trial numbers for the analysis (i.e., 450–700 per subject). Nevertheless, the results were

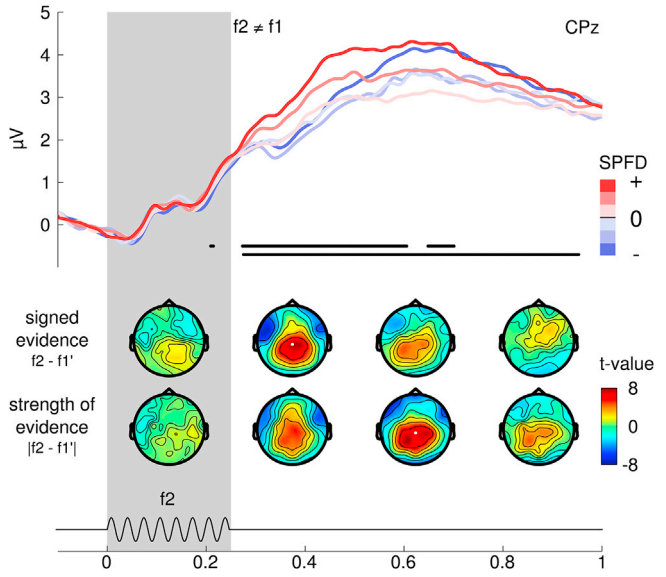

Fig. 4. CPP is first modulated by signed subjective evidence and then by the absolute value, displayed for trials with available physical evidence (i.e., f2 ≠ f1) from all experiments. **Lower**, Scalp topographies of t-values reflecting group-level statistics for modulations by signed subjective evidence (f2 - f1′) and by the absolute strength of evidence (|f2 – f1'|). Displayed topographies are averages over 250 ms windows, starting at 0 with the onset of the second stimulus. The modulation by signed subjective evidence peaks clearly earlier (250–500 ms topography) than the modulation by the absolute strength of evidence (500–750 ms topography). **Upper**, ERPs from electrode CPz (white dot in scalp topographies), are computed separately for six levels of SPFDs, and display a modulation by the signed values of the SPFDs and then by the absolute values of the SPFDs.

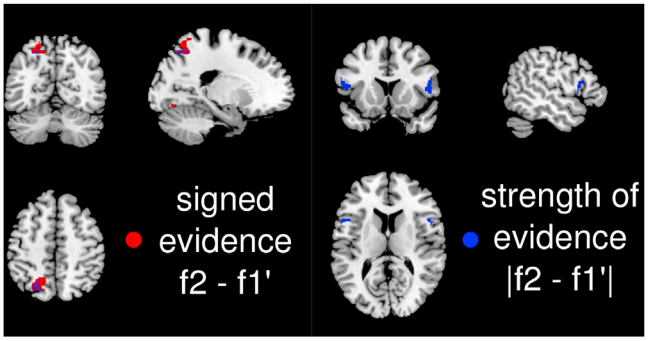

Fig. 5. Source reconstructions for the early CPP modulation by signed subjective evidence (red), and the late modulation by the absolute strength of evidence (blue).

qualitatively identical to those obtained when using the full set (compare Figs. 4 and 6). After the presentation of f2, the ERP was first modulated by the signed subjective evidence (295–578 ms, 24 electrodes, strongest in Pz, CPz, CP1, CP2, P1, and P3 with $p_{FWE} = 0.002$) and then by the absolute strength of evidence for correct decisions (486–676 ms, 14 electrodes, strongest in CPz, Pz, POz, CP1, CP2, and P1 with $p_{FWE} = 0.002$), but not for incorrect decisions (no clusters). A significant difference between correct and incorrect trials was only observed for the modulation by absolute strength of evidence (602–654 ms, electrodes P1, P3, PO7, PO3, POz, and PO4 with $p_{FWE} = 0.026$). With respect to the computation of ERPs, excluding the outermost stimuli led to fewer trials falling into the most extreme bins of SPFDs (see ERPs in Material and Methods). In particular, this concerned large negative and large positive SPFDs (dark blue and dark red in upper panel of Fig. 6) with 26 and 78 participants respectively contributing data to the grand average ERPs

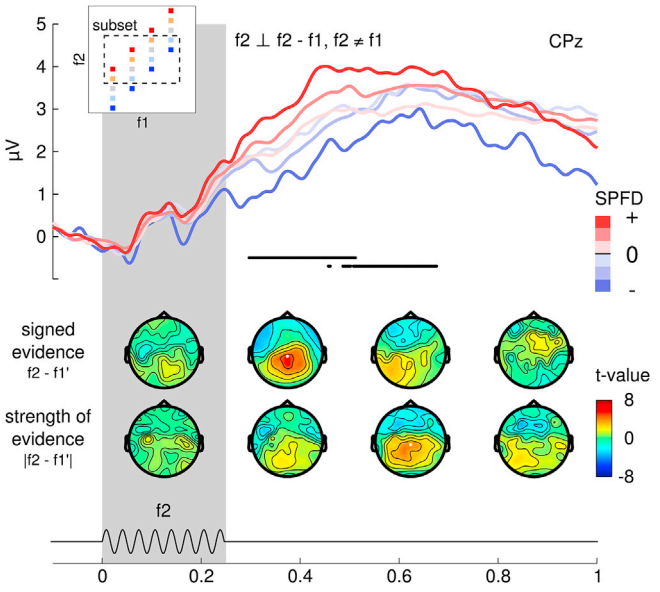

Fig. 6. CPP modulation persists when using a subset of trials in which f2 alone does not predict decision outcome (i.e., f2 and f2 – f1 are orthogonal). Inset in upper left corner highlights the stimulus pairs that were used for this analysis (see also Fig. 1). Same conventions as in Fig. 4. Note that only the grand average ERPs for the most negative and most positive level of SPFDs (dark blue and dark red) were affected by using a reduced dataset (see text for details).

(i.e., 35 trials per individual ERPs on average). For comparison, all 116 participants contributed individual ERPs (based on 67 trials on average) to the grand average for the remaining levels of SPFDs. Again, note that the computation of classic ERPs only served displaying purposes. The statistical analysis was based on single trials (i.e., not binned into discrete levels of SPFDs), and was hence unaffected by any imbalances in trial numbers per discrete levels of SPFDs. Taken together, this analysis ruled out that the outermost stimulus pairs alone accounted for the observed modulations in the EEG signal.

In a further control analysis, we focused on the observation that for some participants SPFDs were distributed asymmetrically around zero due to an overall response bias. As a consequence, the corresponding absolute values were not fully independent from the signed SPFDs. We therefore orthogonalized the absolute values with respect to the signed SPFDs by subtracting the individual mean before computing the multiple regression and again obtained qualitatively identical results (modulation by signed SPFDs: 264–537 ms, 29 electrodes, strongest in Pz, CPz, POz, CP2, CP4, and P1 with $p_{FWE} = 0.002$; modulation by absolute values of SPFDs: 279–947 ms, 32 electrodes, strongest in Pz, CPz, CP1, CP2, P1, and P3 with $p_{FWE} = 0.002$).

Next, we explored whether the EEG signal was possibly also affected after the first stimulus by the quantity that had to be kept in working memory (i.e., in analogy to the presumed subjective difference quantity on which the decision is based). That is, we studied whether we could find a parietal potential that was modulated by f1 in a similar way as the ERPs after f2 were modulated by SPFDs. We did not find any comparable effect (i.e., no cluster with comparable spatial and temporal configuration; for a similar result, see Spitzer et al., 2016).

Finally, when examining the data from each experiment (see Materials and Methods) separately, we found a highly similar pattern of modulations by subjective evidence as with the pooled data. In all experiments, the ERP was first modulated by the signed subjective evidence, and then by the strength of subjective evidence for correct (all effects with $p_{FWE} < 0.014$, except for Experiment 3, with $p_{FWE} = 0.178$ and $p_{FWE} = 0.248$ for the early and late modulation, respectively), but not for incorrect decisions (no clusters in parietal electrodes, except for Experiment 2 showing a negative modulation by the absolute strength of

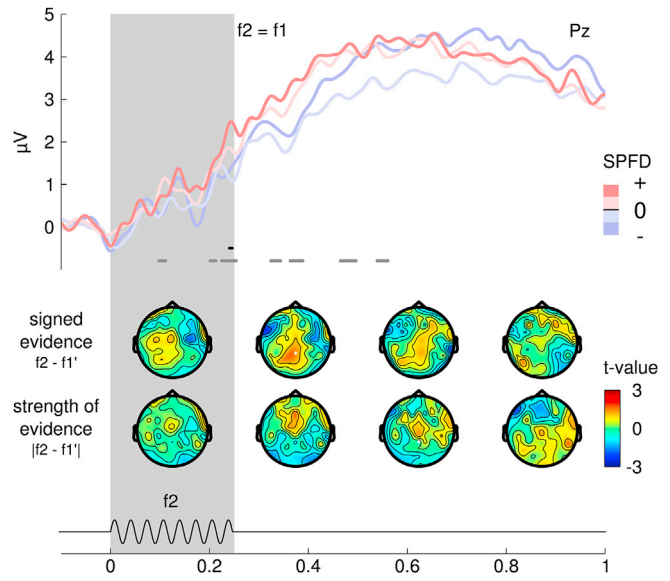

Fig. 7. CPP is modulated by signed subjective evidence even in the absence of physical evidence (i.e., f2 = f1). Same conventions as in Fig. 4. **Lower**, the modulation by subjective evidence peaks in the same time window as the modulation for trials with f2 ≠ f1 (250–500 ms topography) and displays a similar topography (see Figs. 4 and 6). A modulation by the absolute strength of evidence is not observed. Note the different scale of t-values. **Upper**, ERPs from electrode Pz (white dot in scalp topographies) are computed separately for four levels of SPFDs and display a weak modulation by the signed values of the SPFDs. Note that a considerably reduced set of trials (25% of all presented trials) and participants (Experiments 3–6; 73/116 participants) was available for this analysis.

evidence with $p_{FWE} = 0.018$). Note that for Experiments 1 and 2, RTs indicated that participants responded faster for "f2 > f1" choices as compared with "f2 < f1" choices even when considering only trials from the subset shown in Fig. 6, while for Experiments 3–6, RTs were uninformative due to the delayed response paradigm. As a consequence, the observed modulation by signed SPFDs (i.e., higher ERP for positive SPFDs as compared with negative SPFDs immediately after f2) might be explained by faster unfolding decision processes for choices of "f2 > f1". In particular, the observed effect might reflect that the 'classic' CPP, only tracking absolute strength of evidence, started earlier for positive SPFDs than for negative SPFDs. However, since we only have meaningful RT data for two out of six experiments, we can neither confirm nor rule out the possibility that variations in the onset of the CPP explain the observed modulation by signed SPFDs.

3.4. Signed subjective evidence modulates parietal ERPs even during judgements of physically identical stimulus pairs

We repeated the multiple regression analysis with signed SPFDs and their absolute values as regressors, however, this time only using trials without any physical evidence for one or the other choice. In other words, we only used trials with two identical stimuli (i.e., f1 = f2: 12 Hz vs 12 Hz, 16 Hz vs 16 Hz, 20 Hz vs 20 Hz, 24 Hz vs 24 Hz). Crucially, although the physical difference f2 – f1 is always zero for these trials, the individually estimated SPFDs yielded non-zero values for each stimulus pair. This is a direct consequence of the known biases in choice behavior that are typically observed in sequential comparison tasks (i.e., comparing f2 with mean-biased f1' instead of the physical value of f1). Based on the non-zero SPFDs, we were hence able to divide trials according to decisions that were in line with the estimated SPFDs (i.e., SPFD < 0: f1 chosen, and SPFD > 0: f2 chosen), and those that were not (i.e., SPFD < 0: f2 chosen, and SPFD > 0: f1 chosen). This way, we could divide trials into "consistent" and "inconsistent" with respect to the model outcome. Remarkably, for "consistent" decisions, we found a qualitatively similar positive correlation of ERPs with signed SPFDs as for correct trials in which physical evidence for a decision was genuinely present (236–246 ms, electrodes PO3, POz, Pz, CP6, CP4, CP2, P2, P4, P6, P8, and PO8, $p_{cluster} = 0.016$, FWE corrected, Fig. 7). For decisions identified as "inconsistent", no such correlation was found. A comparison between "consistent" and "inconsistent" trials revealed that the modulation of ERPs by signed subjective evidence was significantly different (i.e., an interaction effect) between both sets of trials (322–338 ms, electrodes P3, P5, PO3, Oz, POz, Pz, P2, P8, PO8, and PO4, $p_{cluster} = 0.044$, FWE corrected). Notably, the separation of trials into these two sets was solely based on the modeled SPFDs, and yet, we were able to observe a significant difference in the EEG signal. However, the absolute values of the SPFDs did not modulate the CPP in trials with identical stimuli. Only a more anterior cluster became statistically significant for "consistent" decisions (268–279 ms, electrodes F1, F3, Fz, F2, F4, FC2, and FCz, $p_{cluster} = 0.03$, FWE corrected), however, this effect did not differ between "consistent" and "inconsistent" trials.

3.5. Modulation by absolute strength of evidence relates to statistical decision confidence

That the CPP was correlated with the absolute values of SPFDs may suggest that this modulation could be linked with the level of confidence in a decision. To explore such potential relationship, we checked whether the late CPP conforms with the predictions of "statistical decision confidence" (see Drugowitsch, 2016; Hangya et al., 2016; Sanders et al., 2016). In this framework, confidence exhibits four key characteristics that can be tested without the need for explicit confidence ratings, simply based on statistical decision theory: (1) confidence is positively correlated with PCRs; (2) confidence increases with evidence strength for correct choices, but decreases for incorrect choices (see Fig. 2B for intuition); (3) when (almost) no evidence is available (i.e., in very hard trials), confidence exhibits the same intermediate level for correct and incorrect choices; (4) for the same strength of evidence, high-confidence trials still yield higher PCRs than low-confidence trials.

Concerning (1), as reported in our main results, we found that the late CPP was positively correlated with the absolute values of SPFDs, which in turn were highly correlated with PCRs (Fig. 3D). For (2) and (3) we extracted single-trial amplitudes of the CPP (mean amplitude between 500 and 800 ms after f2 in electrode CPz), and grouped these amplitudes according to the discrete levels of absolute SPFDs separately for correct and incorrect trials (i.e., three levels for trials with f2 ≠ f1: [0 to 0.09]; 0.09 to 0.17]; [> 0.17]; two levels for trials with f2 = f1: [0 to 0.09]; [> 0.09]). As predicted by statistical decision confidence, we found that the CPP amplitude increased with evidence strength for correct trials, and (initially) decreased for incorrect trials, in remarkable alignment with the

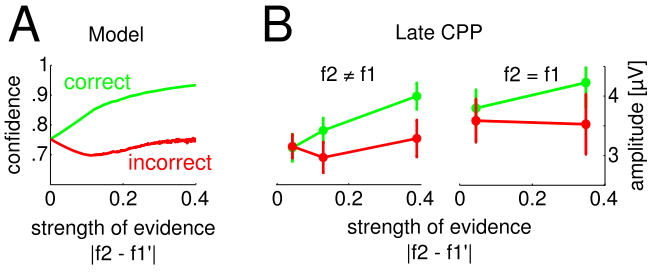

Fig. 8. Late CPP corresponds to statistical decision confidence. **A**, Average statistical decision confidence based on simulations from behavioral models of each participant. Confidence increases with evidence strength (i.e., |f2 – f1'|) for correct trials and decreases (initially) for incorrect trials. For very hard trials (|f2 – f1'| = 0), confidence is at the same intermediate level for correct and incorrect trials. **B**, Average amplitude (+/− standard error of mean) of late CPP (500–800 ms) exhibits same pattern as predicted by simulations shown in A, for trials with and without objective evidence (i.e., f2 ≠ f1 and f2 = f1, respectively).

average confidence computed from individual model fits (Fig. 8A and B). Moreover, for the most difficult trials (i.e., least evidence strength), the CPP amplitude was at the same intermediate level for correct and incorrect trials (Fig. 8B). Notably, predictions (2) and (3) were also reflected in CPP amplitudes when considering only trials with f2 = f1 (Fig. 8B, right panel). Lastly, we did a median split of our data based on CPP amplitudes to simulate a division into high- and low-confidence trials (4). We compared PCRs between high- and low-amplitude trials for each of the three levels of evidence strength (i.e., [0 to 0.09]; [0.09 to 0.17]; [> 0.17]), and found that for the intermediate and high level of evidence strength, PCRs were significantly higher in trials with a high CPP amplitude as compared to trials with a low CPP amplitude (paired *t*-test, both $p < 0.001$). Taken together, the late CPP in the present dataset fulfils all requirements posed by statistical decision theory for a measure of confidence.

3.6. Parietal ERPs correlate with upper beta band amplitude over effector-specific premotor areas

Finally, we explored whether the observed modulation of ERPs was related to choice-specific modulations of upper beta band amplitude over premotor areas previously identified in the same data (i.e., experiments 1 and 2 as reported in Herding et al., 2016, 2017). In this earlier work, beta band power was shown to be higher for "f2 > f1" choices as compared to "f2 < f1" choices, regardless of whether the choice was correct or incorrect. Indeed, we found a positive correlation between the amplitude of parietal ERPs during the early modulation by signed SPFDs and the beta band amplitude (one-sample *t*-test across single-trial correlations of participants, mean rho = 0.03, $p < 0.001$). Notably, we obtained the same positive correlation when considering data from both experiments separately (experiment 1: mean rho = 0.03, $p = 0.016$; experiment 2: mean rho = 0.02, $p = 0.002$). The late CPP (i.e., during the modulation by absolute SPFDs) was also positively correlated with single-trial beta band amplitudes (mean rho = 0.02, $p = 0.006$). However, when considering both experiments separately, only data from experiment 1 showed a significant positive correlation (mean rho = 0.02, $p = 0.02$), but not data from experiment 2 (mean rho = 0.01, $p = 0.15$). Importantly, average response times in experiments 1 and 2 (~862 ms and ~603 ms) intersected with the timing of the late CPP (i.e., 500–800 ms after f2) and thus render a causal relation between this late component and choice-specific upper beta band amplitude unlikely. For the early parietal signal (i.e., 250–500 ms after f2), on the other hand, a causal role in choice selection seems chronologically possible. However, such potential causality remains to be thoroughly investigated in future studies.

4. Discussion

In the current study, we investigated human ERP signals during the comparison of two sequentially presented vibrotactile stimuli (with frequencies f1 and f2). We pooled a sizeable amount of data (N = 116) over six different variants of this task, varying in response modality, response timing, and response mapping, whereas stimuli and comparison task remained unchanged. Despite the variations, we consistently found that ERPs after the second stimulus were first modulated by the signed subjective evidence in favour of the ensuing decision (i.e., signed SPFDs), and later by the absolute strength of evidence (i.e., absolute values of SPFDs). Notably, both modulations were only observed for correct decisions, linking a successful discrimination of f1 and f2 with a faithful representation of the perceived stimulus difference (i.e., SPFDs) in the parietal cortex. Even in the absence of any objective differences between f1 and f2 (i.e., f1 = f2), ERPs indexed the signed values of SPFDs, but not their absolute values. This observation implies that parietal signals may index endogenous evidence for subsequent decisions beyond what has been observed in the CPP (see CPP in the absence of stimuli in O'Connell et al., 2012). Accordingly, we found a correlation between the early ERP effect and choice-selective upper beta band amplitudes in effector-specific premotor areas. The late modulation by the absolute values of SPFDs on the other hand seemed to index the amount of evidence for a decision, which we related to the concept of statistical decision confidence. The putative neuronal sources of both early and late ERP modulation were located in SPL (Brodman area 7A; primarily in the left hemisphere), whereas the late modulation by absolute differences additionally exhibited likely sources in bilateral IFG (Brodman area 44/45).

Several studies of the broadband human EEG signal have shown that the CPP reflects the accumulated evidence for perceptual decisions which require the integration of noisy sensory input over time for immediate and delayed responses (e.g., O'Connell et al., 2012; Kelly and O'Connell, 2013; Philiastides et al., 2014; Twomey et al., 2016). These findings might be directly linked to seminal work on visual perceptual decision making in monkeys that implicated the PPC as a key site for evidence accumulation (see Shadlen and Kiani, 2013). A recent study showed that also in other tasks, i.e., in a classic oddball paradigm, the CPP, or rather the P300, was modulated by the evidence in favour of a successful 'deviant' detection (i.e., a modulation by the difference between 'deviant' and 'standard' stimulus; Twomey et al., 2015). The topography and evolution of the present ERP modulations match these reports. Yet, all previous studies that associated the CPP with decisional evidence, found a modulation of the CPP by the evidence *within* a single choice category, but never a modulation by evidence *across* choice alternatives (e.g., O'Connell et al., 2012; Kelly and O'Connell, 2013; Philiastides et al., 2014; Twomey et al., 2015; Twomey et al., 2016). That is, the CPP was shown to track the strength of available evidence, albeit concealing for which choice alternative. In particular, Kelly and O'Connell (2013) showed that only the proportion of coherent motion, independent of direction (i.e., leftward or rightward), modulated the CPP in an RDM task (see also Twomey et al., 2016). Moreover, Philiastides et al. (2014) were able to discriminate different levels of presented evidence based on a parietal potential (i.e., likely the CPP), no matter whether an image of a face or a car was shown. However, a classification between faces and cars was not possible. In the current study, we report for the first time that parietal signals were modulated by both the amount of evidence and the choice alternative at the same time (i.e., by signed evidence in form of SPFDs). Only later, the absolute strength of evidence alone (i.e., absolute values of SPFDs; independent of the specific choice category) was reflected by the CPP as known from previous work. We propose that the early modulation by the signed values of SPFDs indexes the evidence on which a decision was based. The late modulation by the absolute values of SPFDs, on the other hand, might refer to decision confidence. Given the observed differences in RTs between both choice alternatives (i.e., faster RTs for "f2 > f1" choices) in two out of six experiments, another possible explanation for the reported effects is that we observed the 'classic' CPP, however, with different onset times depending on the subsequent choice. In particular, this could implicate that one choice alternative ("f2 > f1") was processed faster than the other ("f2 < f1"), possibly hinting at a preferred/default choice. However, a recent EEG study of sequential comparisons in the visual domain observed RT differences in the opposite direction, but a similar modulation of the late CPP, suggesting that RTs are unlikely a determining factor for the present data (von Lautz et al., 2019).

In any case, our findings do not entirely correspond to the 'classic' CPP, as we did not observe a saturation of the CPP at a fixed threshold, but rather a modulation by subjective task difficulty (cf. Kelly and O'Connell, 2013; but see Philiastides et al., 2014, von Lautz et al., 2019). A further investigation of this relationship revealed that the modulation by absolute values of SPFDs complied in all respects with the definition of statistical decision confidence (see Drugowitsch, 2016; Hangya et al., 2016; Sanders et al., 2016). Importantly, in line with the classic definition of confidence, statistical decision confidence refers to the probability that a choice is correct (given the evidence) and was recently shown to align with human confidence judgements (Sanders et al., 2016). That is, this framework allows to infer confidence levels even in the absence of

explicit confidence ratings. That the CPP, or rather the P300, might indicate confidence has been suggested for a long time (e.g., Squires et al., 1973; Sutton et al., 1982; Curran, 2004), and has recently been reiterated. Gherman and Philiastides (2015), for instance, reported a higher amplitude of the CPP for choices that were made with high certainty as compared to choices with low certainty (see also Philiastides et al., 2014). Moreover, although the CPP has been typically reported to reach a fixed level at the time of the response report (see Kelly and O'Connell, 2015; but Philiastides et al., 2014; von Lautz et al., 2019), when considering false alarm trials, a clearly lower amplitude was observed, possibly indexing lower confidence in those trials (Fig. 2C in O'Connell et al., 2012). The lack of differences in CPP amplitudes at response time for the remaining results might be related to the task demands per se. By applying continuous task designs (e.g., O'Connell et al., 2012; Kelly and O'Connell, 2013), decision-unrelated stimulus evoked EEG signals were elegantly avoided, however, an additional level was added to the task, requiring the detection of stimuli. This might have led to a rather constant level of confidence before committing to a decision (see Discussion in Philiastides et al., 2014). The reconstructed sources of the confidence-related signal in bilateral IFG appear unusual at first glance, however, recent fMRI studies also implicated the IFG in the processing of confidence (Hebart et al., 2016; Sherman et al., 2016).

Finally, the present vibrotactile 2-AFC task has been used extensively by Romo and colleagues during electrophysiological recordings from monkeys (reviewed in Romo and deLafuente, 2013). In this research, premotor areas were identified as one of the first sites to show decision-related firing rate patterns that encoded the differences between f1 and f2 (Hernández et al., 2002, 2010; Romo et al., 2004). Furthermore, a choice-specific (i.e., binary) amplitude modulation of large-scale upper beta band oscillations (~20–30 Hz) in premotor areas was recently identified in monkey local field potentials (Haegens et al., 2011) as well as in human EEG data (Herding et al., 2016, 2017; Ludwig et al., 2018; von Lautz et al., 2019). With the current results, we might thus provide first evidence for a previously missing EEG signature that indexes the fine-grained subjective evidence in favour of the ensuing choices. Given the stimulus-locked early onset of the ERP modulation by signed evidence, the response-locked character of the beta band modulation, the conceptually reasonable gradient (i.e., choices are based on evidence), and the source locations of both findings (i.e., evidence in parietal cortex and choice in premotor cortex), we presume that parietal ERPs precede and potentially drive the beta band effect. In fact, using the data from Experiments 1 and 2 with immediate responses (i.e., with a direct translation from evidence to choices), we found a positive correlation between single-trial parietal ERP magnitude (during the early modulation by signed evidence) and the level of beta band amplitude (during the choice modulation). However, since the data could only be pooled from two out of six experiments, this possible connection between CPP and beta band amplitude deserves a more thorough investigation in future research.

To conclude, our data corroborate the notion of the CPP tracking evidence in perceptual decision making (see Kelly and O'Connell, 2015). Using a vibrotactile 2-AFC comparison task, we could show, however, that this signal may have additional features depending on task demands. Our results revealed that parietal ERPs first index signed subjective evidence, and only later the absolute strength of evidence. Provided that this observation cannot be attributed to differences in the processing time of evidence accumulation for both choice alternatives, we propose that the early modulation reflects the quantity on which a decision is based, whereas the late modulation indexes the decision confidence. In the context of the vibrotactile 2-AFC task, our findings suggest that the fine-grained signed evidence that is reflected early in the parietal cortex might index the input to more categorical choice representations, e.g., in effector-specific premotor areas (see Haegens et al., 2011; Herding et al., 2016, 2017; Ludwig et al., 2018).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, GRK1589/1 and project 409180874). We thank Ulf Toelch for useful comments on the manuscript.

References

Ashourian, P., Loewenstein, Y., 2011. Bayesian inference underlies the contraction bias in delayed comparison tasks. PLoS One 6.
Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.
Cornelissen, F.W., Peters, E.M., Palmer, J., 2002. The eyelink toolbox: eye tracking with MATLAB and the psychophysics toolbox. Behav. Res. Methods Instrum. Comput. 34, 613–617.
Curran, T., 2004. Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. Neuropsychologia 42, 1088–1106.
Daunizeau, J., Adam, V., Rigoux, L., 2014. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS Comput. Biol. 10, e1003441.
Drugowitsch, J., 2016. Becoming confident in the statistical nature of human confidence judgments. Neuron 90 (3), 425–427.
Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325–1335.
Fechner, G.T., 1860. Elemente der Psychophysik. Breitkopf & Härtel, Leipzig.
Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. Neuroimage 39, 1104–1120.
Friston, K., Henson, R., Phillips, C., Mattout, J., 2006. Bayesian estimation of evoked and induced responses. Hum. Brain Mapp. 27, 722–735.
Gherman, S., Philiastides, M.G., 2015. Neural representations of confidence emerge from the process of decision formation during perceptual choices. Neuroimage 106, 134–143.
Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu. Rev. Neurosci. 30, 535–574.
Haegens, S., Nácher, V., Hernández, A., Luna, R., Jensen, O., Romo, R., 2011. Beta oscillations in the monkey sensorimotor network reflect somatosensory decision making. Proc. Natl. Acad. Sci. 108, 10708–10713.
Haegens, S., Vergara, J., Rossi-Pool, R., Lemus, L., Romo, R., 2017. Beta oscillations reflect supramodal information during perceptual judgment. Proc. Natl. Acad. Sci. 114, 13810–13815.
Hangya, B., Sanders, J.I., Kepecs, A., 2016. A mathematical framework for statistical decision confidence. Neural Comput. 28 (9), 1840–1858.
Hebart, M.N., Schriever, Y., Donner, T.H., Haynes, J.D., 2016. The relationship between perceptual decision variables and confidence in the human brain. Cerebr. Cortex 26, 118–130.
Hellström, Å., 2003. Comparison is not just subtraction: effects of time- and space-order on subjective stimulus difference. Percept. Psychophys. 65, 1161–1177.
Hellström, Å., 1985. The time-order error and its relatives: mirrors of cognitive processes in comparing. Psychol. Bull. 97, 35–61.
Herding, J., Spitzer, B., Blankenburg, F., 2016. Upper beta band oscillations in human premotor cortex encode subjective choices in a vibrotactile comparison task. J. Cogn. Neurosci. 28, 668–679.
Herding, J., Ludwig, S., Blankenburg, F., 2017. Response-modality-specific encoding of human choices in upper beta-band oscillations during vibrotactile comparisons. Front. Hum. Neurosci. 11, 118.
Hernández, A., Nácher, V., Luna, R., Zainos, A., Lemus, L., Alvarez, M., Vázquez, Y., Camarillo, L., Romo, R., 2010. Decoding a perceptual decision process across cortex. Neuron 66, 300–314.
Hernández, A., Zainos, A., Romo, R., 2002. Temporal evolution of a decision-making process in medial premotor cortex. Neuron 33, 959–972.
Ille, N., Berg, P., Scherg, M., 2002. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. J. Clin. Neurophysiol. 19, 113–124.
Jou, J., Leka, G.E., Rogers, D.M., Matus, Y.E., 2004. Contraction bias in memorial quantifyingjudgment: does it come from a stable compressed memory representation or a dynamicadaptation process? Am. J. Psychol. 117 (4), 543–564.
Karim, M., Harris, J.A., Morley, J.W., Breakspear, M., 2012. Prior and present evidence: how prior experience interacts with present information in a perceptual decision making task. PLoS One 7.
Kelly, S.P., O'Connell, R.G., 2013. Internal and external influences on the rate of sensory evidence accumulation in the human brain. J. Neurosci. 33, 19434–19441.
Kelly, S.P., O'Connell, R.G., 2015. The neural processes underlying perceptual decision making in humans: recent progress and future directions. J. Physiol. 109, 27–37.
Lak, A., Costa, G.M., Romberg, E., Koulakov, A.A., Mainen, Z.F., Kepecs, A., 2014. Orbitofrontal cortex is required for optimal waiting based on decision confidence. Neuron 84 (1), 190–201.
Litvak, V., Friston, K., 2008. Electromagnetic source reconstruction for group studies. Neuroimage 42, 1490–1498.
Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., Friston, K., 2011. EEG and MEG data analysis in SPM8. Comput. Intell. Neurosci. 2011, 852961.
Ludwig, S., Herding, J., Blankenburg, F., 2018. Oscillatory EEG signatures of postponed somatosensory decisions. Hum. Brain Mapp. 2018, 00:1–14.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190.
O'Connell, R.G., Dockree, P.M., Kelly, S.P., 2012. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat. Neurosci. 15.
Philiastides, M.G., Heekeren, H.R., Sajda, P., 2014. Human scalp potentials reflect a mixture of decision-related signals during perceptual choices. J. Neurosci. 34, 16877–16889.
Preuschhof, C., Schubert, T., Villringer, A., Heekeren, H.R., 2010. Prior Information biases stimulus representations during vibrotactile decision making. J. Cogn. Neurosci. 22, 875–887.
Raviv, O., Lieder, I., Loewenstein, Y., Ahissar, M., 2014. Contradictory behavioral biases result from the influence of past stimuli on perception. PLoS Computational Biology 10 (12), e1003948.
Roitman, J.D., Shadlen, M.N., 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci. 22, 9475–9489.
Romo, R., de Lafuente, V., 2013. Conversion of sensory signals into perceptual decisions. Prog. Neurobiol. 103, 41–75.
Romo, R., Hernández, A., Zainos, A., 2004. Neuronal correlates of a perceptual decision in ventral premotor cortex. Neuron 41, 165–173.
Romo, R., Salinas, E., 2003. Flutter discrimination: neural codes, perception, memory and decision making. Nat. Rev. Neurosci. 4, 203–218.
Sanchez, G., 2014. Real-time Electrophysiology in Cognitive Neuroscience: towards Adaptive Paradigms to Study Perceptual Learning and Decision Making in Humans.
Sanders, J.I., Hangya, B., Kepecs, A., 2016. Signatures of a statistical computation in the human sense of confidence. Neuron 90, 499–506.
Shadlen, M.N., Kiani, R., 2013. Decision making as a window on cognition. Neuron 80, 791–806.
Sherman, M.T., Seth, A.K., Kanai, R., 2016. Predictions shape confidence in right inferior frontal gyrus. J. Neurosci. 36, 10323–10336.
Smith, P.L., Ratcliff, R., 2004. Psychology and neurobiology of simple decisions. Trends Neurosci. 27, 161–168.
Spitzer, B., Blankenburg, F., 2011. Stimulus-dependent EEG activity reflects internal updating of tactile working memory in humans. Proc. Natl. Acad. Sci. 108, 8444–8449.
Spitzer, B., Blankenburg, F., Summerfield, C., 2016. Rhythmic gain control during supramodal integration of approximate number. Neuroimage 129, 470–479.
Spitzer, B., Wacker, E., Blankenburg, F., 2010. Oscillatory correlates of vibrotactile frequency processing in human working memory. J. Neurosci. 30, 4496–4502.
Squires, K.C., Hillyard, S.A., Lindsay, P.H., 1973. Vertex potentials evoked during auditory signal detection : relation to decision criteria *. Percept. Psychophys. 14, 265–272.
Sutton, S., Rouchkin, D.S., Munson, R., Kietzman, M.L., Hammer, M., 1982. Event related potentials in a two-interval forced choice decision task. Percept. Psychophys. 32, 360–374.
Talbot, W., Darian-Smith, I., Kornhuber, H., Mountcastle, V., 1968. The sense of flutter-vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. Neurophysiology 31, 301–334.
Tobimatsu, S., Zhang, Y.M., Kato, M., 1999. Steady-state vibration somatosensory evoked potentials: physiological characteristics and tuning function. Clin. Neurophysiol. 110, 1953–1958.
Twomey, D.M., Kelly, S.P., O'Connell, R.G., 2016. Abstract and effector-selective decision signals exhibit qualitatively distinct dynamics before delayed perceptual reports. J. Neurosci. 36, 7346–7352.
Twomey, D.M., Murphy, P.R., Kelly, S.P., O'Connell, R.G., 2015. The classic P300 encodes a build-to-threshold decision variable. Eur. J. Neurosci. 42, 1636–1643.
Urai, A.E., Braun, A., Donner, T.H., 2017. Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. Nat. Commun. 8, 14637.
von Lautz, A., Herding, J., Blankenburg, F., 2019. Neuronal signatures of a random-dot motion comparison task. Neuroimage 193, 57–66.
Woodrow, H., 1935. The effect of practice upon time-order errors in the comparison of temporal intervals. Psychol. Rev. 42, 127–152.