

1 Separable neural signatures of confidence during perceptual decisions

2 Balsdon, T.^{1,2*}, Mamassian, P.^{1**}, and Wyart, V.^{2**}

3 1. Laboratoire des Systèmes Perceptifs (CNRS UMR 8248), DEC, ENS, PSL University, 75005 Paris,
4 France

5 2. Laboratoire de Neurosciences Cognitives et Computationnelles (Inserm U960), DEC, ENS, PSL
6 University, 75005, Paris, France

7 * Corresponding author

8 ** Equal contributors

9 Abstract

10 Perceptual confidence is an evaluation of the validity of perceptual decisions. While there is behavioural
11 evidence that confidence evaluation differs from perceptual decision-making, disentangling these two
12 processes remains a challenge at the neural level. Here we examined the electrical brain activity of human
13 participants in a protracted perceptual decision-making task where observers tend to commit to perceptual
14 decisions early whilst continuing to monitor sensory evidence for evaluating confidence. Premature decision
15 commitments were revealed by patterns of spectral power overlying motor cortex, followed by an
16 attenuation of the neural representation of perceptual decision evidence. A distinct neural representation
17 was associated with the computation of confidence, with sources localised in the superior parietal and
18 orbitofrontal cortices. In agreement with a dissociation between perception and confidence, these neural
19 resources were recruited even after observers committed to their perceptual decisions, and thus delineate
20 an integral neural circuit for evaluating perceptual decision confidence. [148 words]

21 Introduction

22 Whilst perception typically feels effortless and automatic, it requires probabilistic inference to resolve the
23 uncertain causes of essentially ambiguous sensory input (Helmholtz, 1856). Human observers are capable of
24 discriminating which perceptual decisions are more likely to be correct using subjective feelings of
25 confidence (Pollack and Decker, 1958). These feelings of perceptual confidence have been associated with
26 metacognitive processes (Fleming and Daw, 2017) that enable self-monitoring for learning (Veenman,
27 Wilhelm, & Beishuizen, 2004) and communication (Bahrami et al., 2012; Frith, 2012). We are only just
28 beginning to uncover the complex functional role of metacognition in human behaviour, and outline the
29 computational and neural processes that enable metacognition. The study of perceptual confidence offers
30 promising insight into metacognition, because one can use our detailed knowledge of perceptual processes
31 to isolate factors which affect the computation of perceptual confidence.

32 At the computational level, perceptual decisions are described by sequential sampling processes (Vickers,
33 1970; Ratcliff, 1978), in which noisy samples of evidence are accumulated over time, until there is sufficient

34 evidence to commit to a decision. The most relevant information for evaluating perceptual confidence is the
35 quantity and quality of evidence used to make the perceptual decision (Vickers, 1979; Kepecs et al., 2008;
36 Moreno-Bote, 2010). At the neural level, perceptual confidence could therefore follow a strictly serial circuit:
37 Relying only on information computed by perceptual processes, with any additional processes contributing
38 only to transform this information for building the confidence response required by the task. Indeed,
39 confidence (or a non-human primate proxy for confidence) can be reliably predicted from the firing rates of
40 neurons coding the perceptual decision itself (Kiani and Shadlen, 2009), suggesting that confidence may be a
41 direct by-product of perceptual processing. However, a large body of behavioural studies suggest that the
42 computation of confidence is not strictly serial. Confidence can integrate additional evidence after the
43 observer commits to their perceptual decision (Baranski and Petrusic, 1994; Pleskac and Busemeyer, 2010),
44 and while this continued evidence accumulation could incorporate only perceptual information, it implies
45 that confidence evaluation does not directly follow from perceptual decision commitment (and therefore
46 involves at least partially dissociable neural processes).

47 There is also evidence that perceptual confidence can rely on separate (non-perceptual) sources of
48 information, such as decision time (Kiani, Corthell, and Shadlen, 2014) and attentional cues (Denison et al.,
49 2018). This suggests that the processes involved in the computation of perceptual confidence may not be
50 reduced to the same processes as for the perceptual decision. Higher-order theories of metacognition
51 propose a framework in which specialised metacognitive resources could be recruited for computing
52 confidence across all forms of decision-making (a general metacognitive mechanism). Indeed, there is some
53 evidence that confidence precision is correlated across different cognitive tasks (such as memory and
54 perception; Mazancieux et al., 2018), suggesting a common source of noise affecting the computation of
55 confidence across tasks (on top of the sensory noise; Bang, Shekhar, and Rahnev, 2019; Shekhar and Rahnev,
56 2020).

57 It is reasonable to expect that a general metacognitive mechanism relies on processing in higher order brain
58 regions. Several experiments have linked modulations in confidence with activity in a variety of subregions
59 of the prefrontal cortex (including the orbitofrontal cortex, Masset et al., 2020, Lak et al., 2014; right
60 frontopolar cortex, Yokoyama et al., 2010; rostro-lateral prefrontal cortex, Fleming et al., 2012, Geurts et al.,
61 2021; inferior frontal sulcus, medial frontal sulcus and medial frontal gyrus, Cortese et al., 2016; see also
62 Vaccaro and Fleming, 2018, for a meta-analysis). Moreover, disrupting the processing in subregions of the
63 prefrontal cortex (Rounis et al., 2010; Lak et al., 2014; Fleming et al., 2014) tends to impair (though not
64 obliterate) the ability to appropriately adjust behavioural confidence responses, whilst leaving perceptual
65 decision accuracy largely unaffected (though these results can be difficult to replicate, Bor et al., 2017;
66 Lapate et al., 2020, and may not generalise to metacognition for memory; Fleming et al., 2014). A challenge
67 in this literature is in specifically relating the neural processing to the computation of confidence, as opposed
68 to transforming confidence into a behavioural response, or a downstream effect of confidence, such as the
69 positive valence (and sometimes reward expectation) accompanying correct decisions. Moreover,
70 identifying how these neural mechanisms could be separable from the underlying perceptual processes is

71 important for understanding the computational architecture of metacognition.

72 One promising avenue of research for separating the mechanisms of metacognition from perceptual
73 processes has been to utilise tasks where the observer may integrate additional evidence for confidence
74 after they have committed to their perceptual decision (Murphy et al., 2015; Fleming et al., 2018), which
75 presumably relies on processing independent of the perceptual decision. These studies show that post-
76 decisional changes in confidence magnitude correlate with signals from the posterior medial frontal cortex.
77 However, these signals could reflect processes occurring downstream of confidence, such as an emotional
78 response to the error signal, which has been shown to drive medial frontal activity more strongly than
79 decision accuracy (Gehring and Willoughby, 2002). Further research is therefore required to link neural
80 processes specifically with the computation of perceptual confidence.

81 In this experiment we aim to identify the neural processes specifically contributing to the computation of
82 confidence, in a paradigm in which these processes can be delineated from those of perceptual decision-
83 making. We exploit a protracted decision-making task in which the evidence presented to the observer can
84 be carefully controlled. On each trial, the observer is presented with a sequence of visual stimuli, oriented
85 Gabor patches, which offer a specific amount of evidence towards the perceptual decision. The orientations
86 are sampled from one of two overlapping circular Gaussian distributions, and the observer is asked to
87 categorise which distribution the orientations were sampled from. We manipulate the amount of evidence
88 presented such that the observer tends to covertly commit to their perceptual decision before evidence
89 presentation has finished, whilst continuing to monitor ongoing evidence for assessing their confidence
90 (Balsdon et al., 2020). These covert decisions are evident from behaviour and computational modelling, and
91 we show similarities between the neural processes of decision-making across conditions of immediate and
92 delayed response execution.

93 To examine the computation of confidence, we compare human behaviour to an optimal observer who
94 perfectly accumulates all the presented evidence for perceptual decisions and confidence evaluation. The
95 optimal observer must accurately encode the stimulus orientation, the decision update relevant for the
96 categorisation, and add this to the accumulated evidence for making the perceptual decision. We uncover
97 dynamic neural representations of these variables using model-based electroencephalography (EEG), and
98 examine how the precision of these representations fluctuate with behavioural precision. We find two
99 distinct representations of the accumulated evidence. The first one reflects the internal evidence used to
100 make perceptual decisions. The second representation reflects the internal evidence used to make
101 confidence evaluations (separably from the perceptual evidence), and is localised to the superior parietal
102 and orbitofrontal cortices. Whilst the perceptual representation is attenuated following covert decisions, the
103 confidence representation continues to reflect evidence accumulation. This is consistent with a neural circuit
104 that can be recruited for confidence evaluation independently of perceptual processes, providing empirical
105 evidence for the theoretical dissociation between perception and confidence.

106 Results

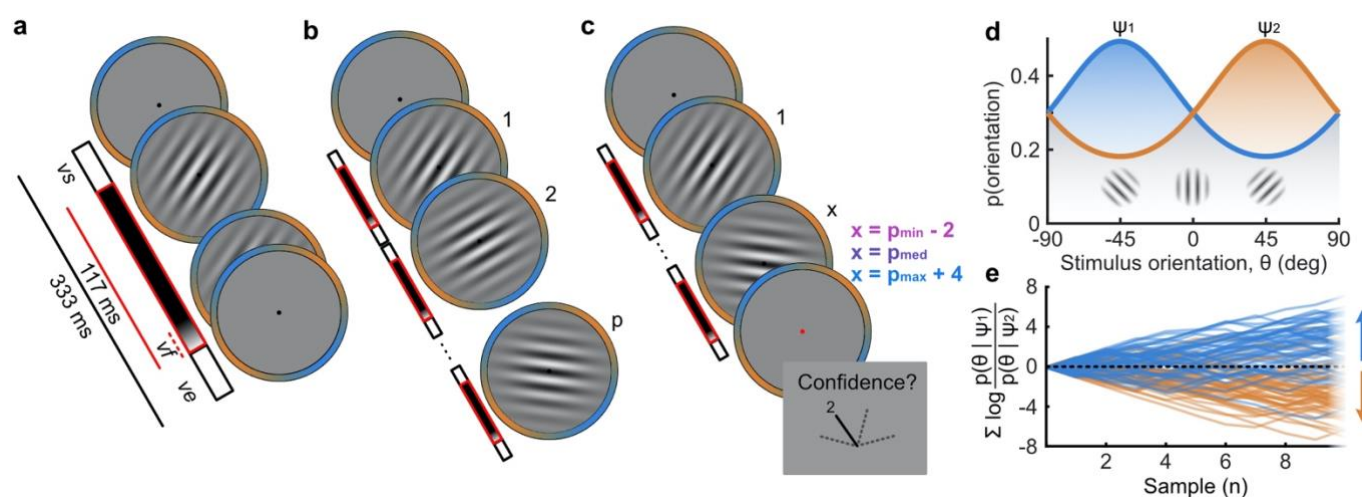
107 Preview

108 We present analyses to address two key hypotheses in this experiment: First, that observers are
109 prematurely committing to their perceptual decisions whilst continuing to monitor additional evidence for
110 evaluating their confidence. And second, that there are separable neural signatures of the evaluation of
111 confidence during perceptual decision-making. To address the first hypothesis, we use a combination of
112 behavioural analyses and computational modelling, and in addition, show that the EEG signatures of
113 response preparation are triggered from the time of decision commitment, even when this occurs seconds
114 prior to the response cue. To address the second hypothesis, we use the stimulus evoked responses in EEG to
115 trace the representation of the presented evidence throughout each trial. We show that these neural
116 representations of the optimal accumulated decision evidence are less precise when the observers'
117 behavioural responses were also less precise relative to optimal. We use this to isolate clusters of activity
118 that specifically reflect the internal evidence used for observers' confidence evaluations beyond the
119 presented evidence. We then localise the sources of this activity, and relate these processes back to
120 observers' eventual confidence ratings.

121 The computational architecture of perceptual confidence

122 Human observers ($N = 20$) performed two versions of the task whilst EEG was recorded. Across the two
123 tasks, 100 predefined sequences of oriented Gabors were repeated for each observer, with stimuli presented
124 as described in **Figure 1a**. In the Free task, the sequence continued until observers entered their perceptual
125 decision (**Figure 1b**), indicating which category (**Figure 1d**) they thought the orientations were sampled
126 from. Observers were instructed to enter their response as soon as they 'felt ready', on three repeats of each
127 predefined sequence (300 trials in total). In the Replay task (**Figure 1c**), observers were shown a specific
128 number of samples and could only enter their response after the response cue. After entering their
129 perceptual decision, they made a confidence evaluation, how confident they were that their perceptual
130 decision was correct, on a 4-point scale. Importantly, the number of samples shown in the Replay task was
131 manipulated relative to the Free task, in three intermixed conditions: in the Less condition, they were shown
132 two fewer than the minimum they had chosen to respond to over the three repeats of that predefined
133 sequence in the Free task; in the Same condition they were shown the median number of samples; and in the
134 More condition, four more than the maximum. The variability across repeats in the Free task means that in
135 the More condition, observers were shown at least four additional stimuli, but often more than that. There is
136 an optimal way to perform this task, in the sense of maximising perceptual decision accuracy across trials.
137 The optimal computation takes as decision evidence the log probability of each orientation given the
138 category distributions (**Figure 1d**) and accumulates the difference in this evidence for each category (**Figure**
139 **1e**, Drugowitsch et al., 2016). We refer to the accumulated difference in log probabilities as the optimal
140 presented evidence, L . Human observers may have a suboptimal representation of this evidence, L^* , and we
141 estimate the contribution of different types of suboptimalities (specifically, inference noise, and a temporal

142 integration bias) with the help of a computational model (full details in **Methods** and **Supplementary Note**
 143 1).



144

145 **Figure 1. Procedure.** **a)** Stimulus presentation: stimuli were presented at an average rate of 3 Hz, but with
 146 variable onset and offset ($vs \in [83, 133]$ ms, $vs_s + ve_{s-1} \geq 216$ ms; see **Methods**). Stimuli were presented
 147 within a circular annulus which acted as a colour guide for the category distributions. The colour guide and the
 148 fixation point were present throughout the trial. **b)** Free task: on each trial observers were presented with a
 149 sequence of oriented Gabors, which continued until the observer entered their response (or 40 samples were
 150 shown). 100 sequences were predefined and repeated three times. **c)** Replay task: The observer was presented
 151 with a specific number of samples and could only enter their response after the cue (fixation changing to red).
 152 The number of samples (x) was determined relative to the number the observer chose to respond to on that
 153 same sequence in the Free task (p). There were three intermixed conditions, Less ($x = p_{min} - 2$; where p_{min} is the
 154 minimum p of the three repeats), Same ($x = p_{med}$; where p_{med} is the median p) and More ($x = p_{max} + 4$; where p_{max}
 155 is the maximum p of the three repeats of that predefined sequence). **d)** Categories were defined by circular
 156 Gaussian distributions over the orientations, with means -45° (ψ_1 , blue) and 45° (ψ_2 , orange), and
 157 concentration $\kappa = 0.5$. The distributions overlapped such that an orientation of 45° was most likely drawn
 158 from the orange distribution but could also be drawn from the blue distribution with lower likelihood. **e)** The
 159 optimal observer accumulates the difference in the evidence for each category, which is defined as the log
 160 probability of the sample orientation (θ) given the distributions. The perceptual decision is determined by the
 161 sign of the accumulated evidence, where the evidence accumulated across more samples better differentiates
 162 the true categories (example evidence traces are coloured by the true category).

163 Based on our previous findings (Balsdon et al., 2020) we expected observers to prematurely commit to
 164 perceptual decisions in the More condition, whilst continuing to monitor sensory evidence for evaluating
 165 their confidence. Replicating these previous results (Balsdon et al., 2020), we found that perceptual decision
 166 sensitivity (d') was significantly decreased with just two fewer stimuli in the Less condition compared to
 167 those same (p_{min}) trials in the Free task (Wilcoxon sign rank $Z = 3.88$, $p < 0.001$, Bonferroni corrected for
 168 three comparisons, **Figure 1a**), but four additional stimuli (**Figure 1b**) in the More condition resulted in
 169 only a small but not significant increase compared to the p_{max} trials in the Free task ($Z = -1.53$, $p = 0.13$,

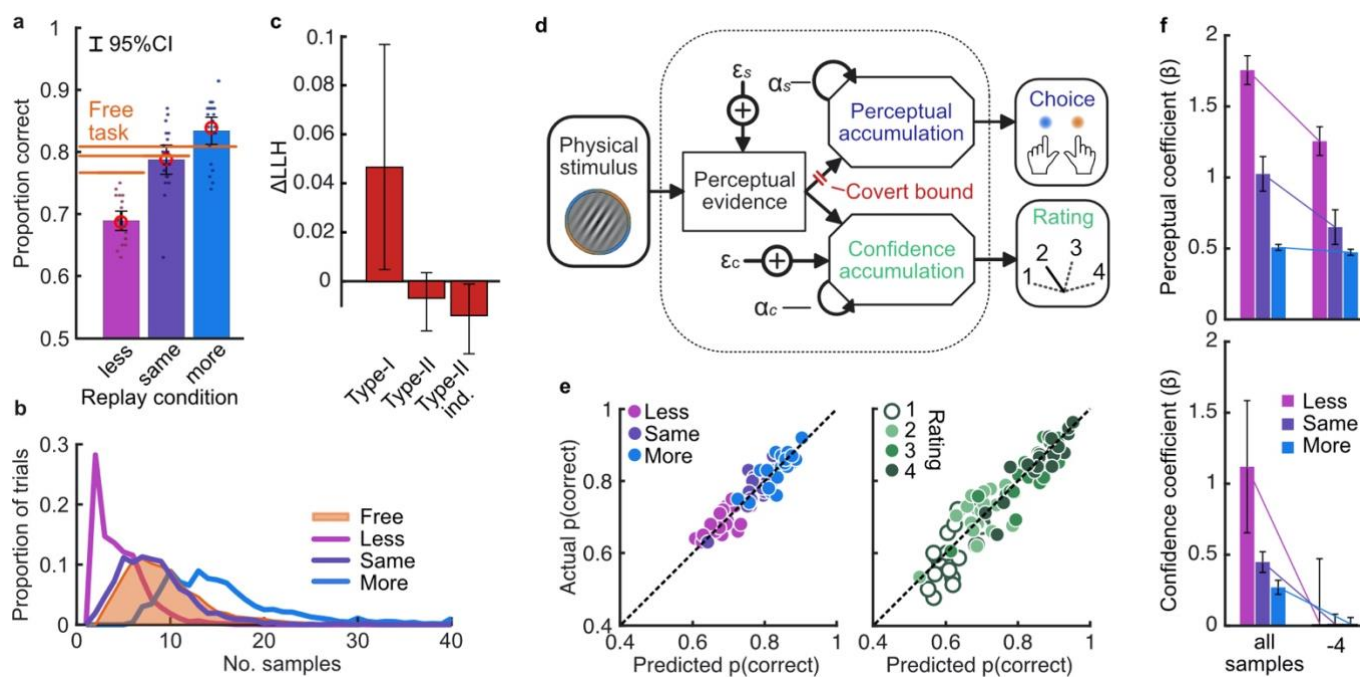
170 uncorrected). There was also no significant difference for the Same condition ($Z = 1.21$, $p = 0.23$,
171 uncorrected; **Figure 2a**).

172 This lack of substantial increase in performance in the More condition could be the result of either a
173 performance ceiling effect or a premature commitment to the perceptual decision. The former explanation
174 reflects a limitation of the perceptual evidence accumulation process, whereas the latter refers to an active
175 mechanism that ignores the final sensory evidence. We compared these two hypotheses using a
176 computational modelling approach (Balsdon et al., 2020; see **Methods**). Specifically, we compared a model
177 in which performance in the More condition is limited by the suboptimalities evident from the Same and the
178 Less conditions (inference noise, and temporal integration bias, see **Methods** and **Supplementary Note 1**),
179 to a model in which performance could be impacted by a covert bound at which point observers commit to a
180 decision irrespective of additional evidence. Cross-validated model comparison provided significant
181 evidence that observers were implementing a covert bound (mean relative increase in model log-likelihood
182 = 0.048, bootstrapped $p = 0.001$, **Figure 2c**). The winning model provided a good description of the data (red
183 open markers in **Figure 2a**, and individual participants in **Figure 2e**).

184 In contrast to what we found for the perceptual decision, there was no evidence that observers were
185 implementing a covert bound on confidence: Implementing the same bound as the perceptual decision did
186 not improve the fit (relative improvement with bound = -0.007, bootstrapped $p = 0.11$, uncorrected) and an
187 independent bound actually significantly *reduced* the fit compared to continued accumulation (relative
188 improvement = -0.014, $p = 0.022$, Bonferroni corrected for two comparisons; **Figure 2c**). We obtained
189 further distinctions between perceptual and confidence processes through computational modelling:
190 additional noise was required to explain the confidence ratings, along with a separate temporal bias. The
191 best description of both perceptual and confidence responses was provided by a partially dissociated
192 computational architecture (full details in **Supplementary Note 1**), where perceptual and confidence
193 decisions are based on the same noisy representation of the sensory evidence, but confidence accumulation
194 incurs additional noise and can continue after the completion of perceptual decision processes (**Figure 2d**,
195 and the predictions of this model for individual participants are show in **Figure 2e**). These computational
196 differences between perceptual decisions and confidence evaluations suggest deviations between the
197 internal evidence on which observers base their perceptual and confidence decisions (see **Supplementary**
198 **Note 2** for model simulations).

199 These modelling results are supported by an analysis using general linear models to examine the
200 relationship between the optimal presented evidence, L , and observers' behaviour in the perceptual decision
201 and confidence evaluation. As stated above, L is the evidence that which maximises the probability of a
202 correct response: the accumulated difference in the log probabilities of the presented orientations given the
203 category distribution (**Figure 1e**). First, we find the presented evidence accumulated over all samples does
204 explain substantial variance in observers' perceptual decisions (average $\beta = 0.77$, $t(19) = 6.48$, $p < 0.001$),
205 and confidence evaluations (with the evidence signed by the perceptual response; $\beta = 0.24$, $t(19) = 6.46$, $p <$
206 0.001). This suggests that the internal evidence that observers were using to make their responses, L^* ,

207 correlated significantly with the optimal evidence L (as has been found previously; Drugowitsch et al., 2016).
 208 Second, the total accumulated evidence in the More condition was not a significantly better predictor of the
 209 observers' perceptual decisions than the evidence up to four samples prior to the response (average
 210 difference in $\beta = 0.034$, $t(19) = 1.63$, $p = 0.12$), while for the Same and Less conditions the total accumulated
 211 evidence was a significantly better predictor (Less: $t(19) = 4.99$, $p < 0.001$; Same: $t(19) = 3.11$, $p = 0.006$;
 212 causing a significant interaction between condition and sample accumulated to $F(2,38) = 10.348$, $p = 0.001$,
 213 Bonferroni corrected for three comparisons, **Figure 2f**, top). This supports the finding from model
 214 comparison and behaviour that observers implemented a covert bound on perceptual evidence
 215 accumulation. And finally, this interaction was not present when examining how the presented evidence
 216 affected confidence evaluations ($F(2,38) = 3.124$, $p = 0.09$, uncorrected, **Figure 2f**, bottom). Rather, the
 217 accumulated evidence up to the final sample in the More condition was a significantly better predictor of
 218 confidence than the evidence accumulated to four samples from the response (average difference in $\beta =$
 219 0.26 , $t(19) = 5.33$, $p < 0.001$), supporting the prediction from the computational model analysis that
 220 observers integrated all the presented evidence for evaluating confidence.



221

222 **Figure 2. Behaviour and computational modelling. a)** Proportion correct in each condition of the Replay
 223 task, relative to the Free task (orange horizontal lines). Individual data are shown in scattered points, error
 224 bars show 95% between- (thin) and 95% within- (thick) subject confidence intervals. Open red markers show
 225 the model prediction. **b)** Distributions of the number of samples per trial in the Free task, and Replay task
 226 conditions (over all observers). **c)** Difference in log-likelihood of the models utilising a covert bound relative to
 227 the models with no covert bound. On the left, the model fitting perceptual decisions only. The middle bar shows
 228 the difference in log-likelihood of the fit to confidence ratings with identical perceptual and confidence bounds.
 229 The right bar shows the difference in log-likelihood of the fit to confidence ratings of the model with an
 230 independent bound for confidence evidence accumulation. Error bars show 95% between-subject confidence
 231 intervals. **d)** The computational architecture of perceptual and confidence decisions, based on model

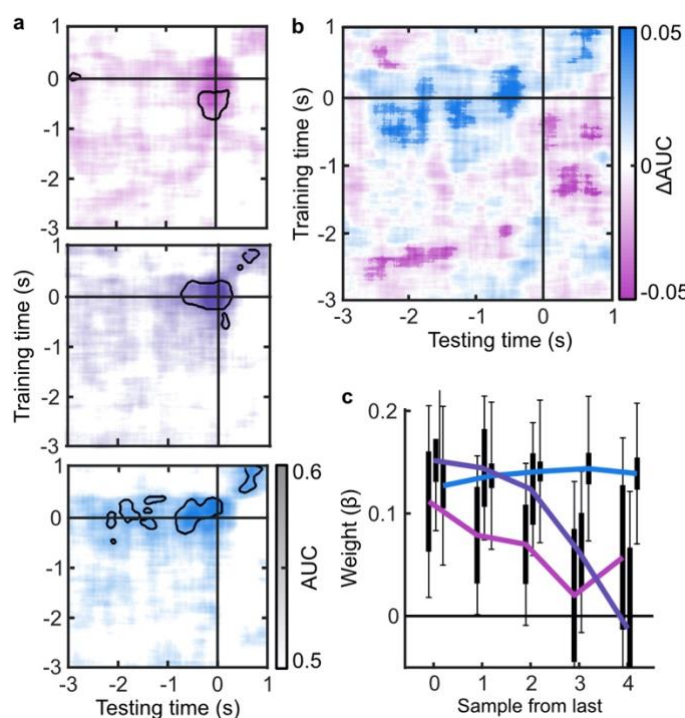
232 *comparison. Perceptual and confidence decisions accumulate the same noisy perceptual evidence, but*
233 *confidence is affected by additional noise (ϵ_c) and a separate temporal bias (α_c). This partial dissociation*
234 *allows Type-II accumulation to continue after the observer has committed to a perceptual decision. e) Predicted*
235 *proportion correct compared to actual proportion correct for each observer, based on the fitted model*
236 *parameters of the final computational model. The left panel shows proportion correct split by condition, and the*
237 *right, split by confidence rating. f) Regression coefficients from the GLM analysis showing the relationship*
238 *between the optimal evidence L , and observers' perceptual (top) and confidence (bottom) responses for trials*
239 *split by condition. The right set of bars show the same analysis but with evidence accumulated up to four*
240 *samples from the response cue.*

241 **EEG signatures of premature perceptual decision commitment**

242 The analysis of behaviour and computational modelling so far has suggested that observers were committing
243 to their perceptual decisions early in the More condition and ignoring the additional evidence for their
244 perceptual decision. We questioned the extent of this covert decision commitment, that is, whether
245 observers were going as far as to plan their motor response before the response cue. We examined the
246 neural signatures of the planning and execution of motor responses using a linear discriminant analysis of
247 the spectral power of band-limited EEG oscillations (see **Methods**). Initial analysis suggested the spectral
248 power in the 8 to 32 Hz frequency range (the 'alpha' and 'beta' bands) could be used to classify perceptual
249 decisions based on lateralised differences over motor cortex (**Supplementary Note 5**). A classifier was
250 trained to discriminate observers' perceptual decisions at each time-point in a four second window around
251 the response in the Free task (3 seconds prior to 1 second after). This classifier was then tested across time
252 in each condition of the Replay task, to trace the progression of perceptual decision-making in comparison to
253 the Free task (where decisions are directly followed by response execution). If covert decisions lead to early
254 motor response preparation, we would expect asymmetries in cross-classification performance on trials
255 where the observer was likely to have covertly committed to a decision (in the More condition) compared to
256 those trials in which they were unlikely to have committed to their decision (in the Less condition). Indeed,
257 there were opposite asymmetries in the cross-classification of the Less and the More conditions (**Figure 3a**).
258 Statistical comparison revealed substantial clusters of significant differences (**Figure 3b**): Training around -
259 0.78 to 0.44 s from the time of the response in the Free task led to significantly better accuracy testing in the
260 More condition than in the Less condition, prior to when the response was entered (for the cluster testing at
261 -2.5 to -1.6 s $Z_{ave} = 2.04$, $p_{cluster} = 0.002$; testing at -1.5 to -1 s, $Z_{ave} = 1.95$, $p_{cluster} = 0.01$; testing at -0.8 to -0.3,
262 $Z_{ave} = 2.32$, $p_{cluster} < 0.001$). This pattern of findings suggests that observers were not only committing to their
263 perceptual decision early, but already preparing their motor response.

264 As an exploratory analysis, we took the strength of the classifier prediction trained and tested at the time of
265 the response as a trial-wise measure of the decision variable used by the participant to enter a response. We
266 reasoned that the amount of evidence in favour of the decision could influence the assiduity with which
267 observers enter their response. We found that the optimal evidence L , accumulated over all samples, could
268 predict the strength of the classifier prediction at response time (mean $\beta = 0.11$, $t(19) = 3.89$, $p < 0.001$;

269 **Figure 3c**). For the Same and Less conditions, the weight on the accumulated evidence appeared to decrease
 270 as evidence was accumulated to samples further prior from the response. But, in the More condition, the
 271 evidence accumulated up to four samples prior to the response still predicted the strength of the classifier
 272 prediction ($t(19) = 3.81, p = 0.001$). This difference between conditions over samples is evidenced by a
 273 significant interaction based on a repeated measures ANOVA ($F(8,152) = 2.429, p = 0.05$, after Bonferroni
 274 correction for three comparisons). Leading up to the response, the accumulated evidence becomes
 275 increasingly predictive of the strength of the classifier prediction, except in the More condition, where this
 276 prediction is already accurate up to four samples prior to the response: After committing to a perceptual
 277 decision, the observer's perceptual response is no longer influenced by additional evidence.



278

Figure 3. EEG signatures of premature perceptual decisions. **a)** Classifier AUC training at each time-point in the Free task and testing across time in the Less (top), Same (middle), and More (bottom) conditions of the Replay task. Black contours encircle regions where the mean is 3.1 standard deviations from chance (0.5; 99% confidence). **b)** Difference in AUC between the More and Less conditions. Cluster corrected significant differences are highlighted. **c)** The relationship between the evidence accumulated up to n samples prior to the response cue and the strength of the neural signature of response execution in each condition. Error bars show 95% within- (thick) and between-subject (thin) confidence intervals.

279 Representations of decision evidence in EEG signals

280 Our main goal was to isolate the neural signatures of the computation of confidence. Observers' behaviour
 281 varied with the optimal evidence L presented to them, but the internal evidence on which they based their
 282 perceptual decisions and confidence evaluations, L^* , clearly deviated from L . In other words, the observers'
 283 behavioural performance was not optimal. To identify the neural computations underlying human
 284 behaviour, we therefore began by isolating the neural signals which correlate with L . We then isolated where
 285 and when deviations in the neural representation of L covary with deviations in L^* - the internal evidence
 286 reflected in observers' behaviour.

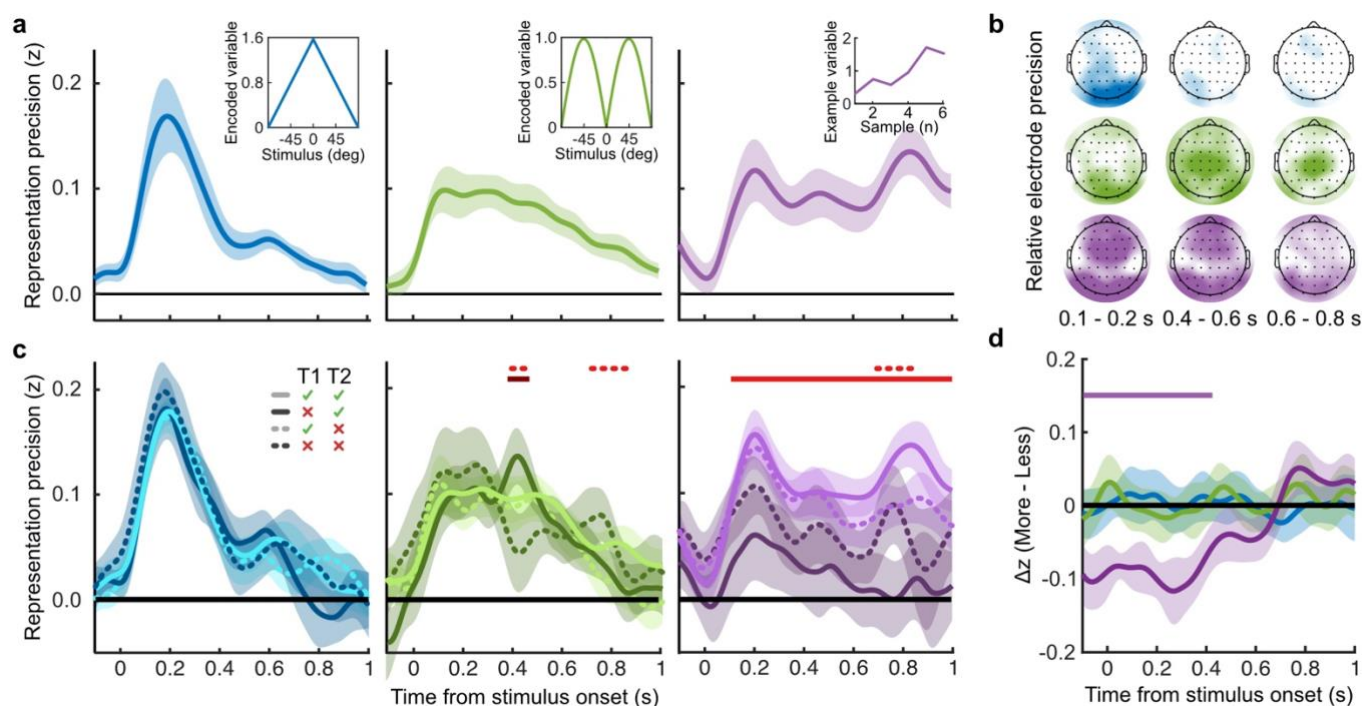
287 To perform this task the optimal observer must encode the orientation of the stimulus, estimate the decision
 288 update based on the categories, and add this to the accumulated evidence for discriminating between the
 289 categories (Wyart et al., 2012; Wyart et al., 2015). We examined the neural representation of these optimal
 290 variables using a regression analysis with the EEG signals (evoked response, bandpass filtered between 1
 291 and 8 Hz, see **Methods**). At each time point, we used the relationship between the pattern of neural activity

292 and the encoding variables on 90% of the data to predict the encoding variables on the remaining 10% of the
293 data (10-fold cross validation). The precision of the neural representation was calculated as the correlation
294 between the predicted encoding variable and actual encoding variable in the held-out data, across all 10
295 folds (see **Methods**). **Figure 4a** shows the time course of the precision of the neural representation of
296 stimulus orientation, momentary decision update, and accumulated evidence (L), locked to stimulus onset.
297 The precision of the representations of these variables showed distinct time courses and relied on distinct
298 patterns of EEG activity over scalp topography (**Figure 4b**). There was a transient representation of stimulus
299 orientation localised over occipital electrodes. The representation of the momentary decision update was
300 maintained for a longer duration, initially supported by occipital electrodes, then increasingly localised over
301 central-parietal electrodes. The representation of the accumulated evidence was sustained even longer and
302 relied on both frontal and occipital electrodes.

303 The internal evidence on which observers base their response, L^* , can differ from the optimal evidence, L .
304 When the eventual behavioural response differs from that predicted by L , L^* is likely to be more different
305 from L . A neural representation of L that reflects L^* (that is, reflecting the underlying processing responsible
306 for behaviour) should also be less precise for samples in these trials. For each variable, we estimated the
307 representation precision separately for epochs leading to behavioural responses that differed from the
308 optimal response (based on L), and responses that matched those of the optimal observer (Replay task
309 epochs only; **Figure 4c**; **Supplementary Note 3**). For perceptual decisions, the optimal observer responds
310 with the correct category. For confidence evaluations, the optimal observer gives high confidence on trials
311 with greater than the median evidence (over all trials) for their perceptual response. The precision of the
312 representation of stimulus orientation did not significantly vary based on whether behaviour matched the
313 optimal response. The representation precision of the momentary decision update showed a significant
314 effect for the perceptual decision from 380 to 468 ms ($F_{avg}(1,19) = 7.97, p_{cluster} = 0.008$) and a significant
315 interaction between perceptual and confidence responses from 396 to 468 ms ($F_{avg}(1,19) = 6.66, p_{cluster} =$
316 0.022) and from 716 to 856 ms ($F_{avg}(1,19) = 10.75, p_{cluster} < 0.001$). The largest effects were seen in the
317 representation precision of the accumulated evidence. Representation precision was significantly reduced in
318 epochs leading to non-optimal perceptual decisions from 108 ms post stimulus onset to the end of the epoch
319 ($F_{avg}(1,19) = 13.65, p_{cluster} < 0.001$). In addition, there was a significant interaction with confidence from 696
320 to 836 ms ($F_{avg}(1,19) = 8.72, p_{cluster} = 0.005$). The precision of the EEG representations therefore showed
321 distinct associations with behaviour.

322 The presence of a covert bound implies that, after the observer commits to a decision, they no longer
323 incorporate additional evidence for that decision. We should therefore see significant decreases in the
324 precision of representations that specifically contribute to perceptual evidence accumulation. Indeed, the
325 precision of the early representation of accumulated evidence was significantly attenuated for the last four
326 samples of the More condition (in which observers were likely to have already committed to a decision),
327 compared to the last four samples of the Less condition (where observers were unlikely to have committed
328 to a decision; from the start of the epoch to 424 ms, **Figure 4d**; $t_{avg}(19) = -5.19, p_{cluster} < 0.001$). These

329 differences in representation precision were not present for the encoding of stimulus orientation, nor the
 330 decision update, suggesting that these processes may reflect input to perceptual evidence accumulation, but
 331 not the accumulation process itself. As a control analysis, this decreased precision was not evident in a
 332 comparison of the first four samples (**Supplementary Note 6**), suggesting this effect on the representation
 333 of accumulated evidence is specific to those samples likely to have occurred after perceptual decision
 334 commitment, as opposed to those samples in More condition trials per se. Together, these comparisons
 335 suggest that different aspects of these evolving EEG representations of decision variables are related to the
 336 neural processes for perception and confidence.



337

338 **Figure 4. Representation of decision variables.** **a)** Representation precision (Fischer transformed correlation
 339 coefficient, z) of stimulus orientation (blue, left), momentary decision update (green, middle), and accumulated
 340 decision evidence (purple, right). The encoded variables are shown in the insets (the accumulated evidence is
 341 the cumulative sum of the momentary evidence signed by the response, only one example sequence is shown).
 342 Shaded regions show 95% between-subject confidence intervals. **b)** Relative electrode representation precision
 343 over three characteristic time windows (100 – 200 ms, left; 400 – 600 ms, middle; and 600 – 800 ms, right). **c)**
 344 Representation precision for epochs leading to optimal and suboptimal perceptual (T1) and confidence (T2)
 345 responses. Lighter lines show perceptual decisions that match the optimal response, dashed lines show
 346 suboptimal confidence ratings. Dashed red horizontal lines show significant interactions between perceptual
 347 and confidence suboptimality. The light red horizontal line shows the significant effect of suboptimal perception
 348 and the dark red horizontal line shows the significant effect of suboptimal confidence. Shaded regions show
 349 95% within-subject confidence intervals. **d)** Difference in decoding precision between the More and the Less
 350 conditions for epochs corresponding to the last four samples of the trial. The purple horizontal line shows the
 351 significant difference in decoding of accumulated evidence.

352 Neural processes for confidence

353 The analysis above shows that the EEG representation of accumulated evidence reflected greater differences
354 from the optimal presented evidence L in trials where behaviour does not match the optimal response. This
355 suggests that the corresponding neural signals reflect more closely L^* (the internal evidence actually used by
356 observers to decide) than L . To isolate the neural signals which reflect L^* , we assume that L^* approximates L
357 with normally distributed errors, and that these errors have larger variance on trials leading to responses
358 that do not match the optimal evidence L (a similar approach as in Van Bergen et al., 2015). We used
359 multivariate Bayesian scan statistics (Neill, 2011; Neill, 2019) to cluster signals in space (electrode location)
360 and time where the variance from L in the neural representation corresponded to deviations in L^* , based on
361 behaviour. The statistic tested whether the variability in the neural representation was related to L^* to a
362 greater extent than could be explained by measurement noise alone (see **Supplementary Note 7** for further
363 details). In this way, the statistic isolates signals more closely related to L^* than can be explained by L , taking
364 into account the noise affecting our measurement of these neural signals.

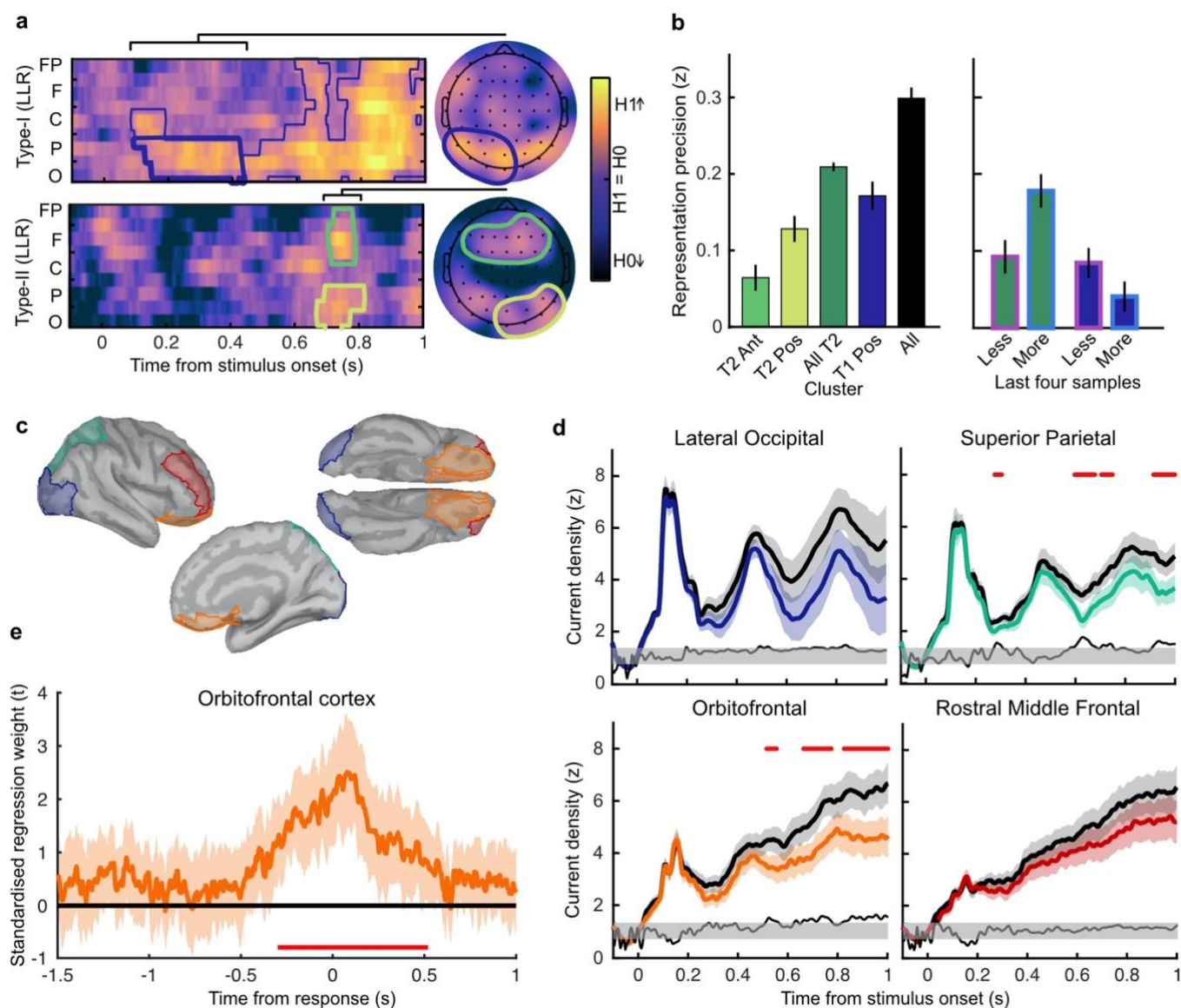
365 For perceptual decision-making, signals related to L^* were initially clustered over posterior electrodes,
366 becoming dispersed over more anterior electrodes late in the epoch (**Figure 5a**, top). For confidence, we
367 found two co-temporal clusters in posterior and anterior electrodes emerging from 668 ms to 824 ms from
368 stimulus onset (**Figure 5a**, bottom). In **Figure 5a** we highlight an early posterior cluster of signals strongly
369 related to L^* for perceptual decisions, that was not diagnostic of confidence evaluations (in fact the evidence
370 was in favour of the null hypothesis; summed log likelihood ratio = -1176). We obtained cluster-wide
371 representations of L from the signals in this early posterior cluster and the two confidence related clusters.
372 The precision of these representations is shown in **Figure 5b**, left. That the information from these clusters
373 is not redundant is evident from the fact that combining the clusters improves the representation precision
374 (**Figure 5b**). For simplicity, we combined the two confidence clusters for further analysis. Similar to the
375 previous analysis (**Figure 4d**), the representation precision of the early posterior cluster was attenuated for
376 the last four samples of the More condition. But, the representation precision of the confidence cluster was
377 maintained (a repeated measures ANOVA revealed a significant interaction between cluster and condition
378 for decoding precision in the last four samples, $F(1,19) = 32.00$, $p = 0.001$, Bonferroni corrected for three
379 comparisons). These results are consistent with dissociable stages of neural processing for confidence
380 evaluation and perceptual decision-making, and support the computational modelling in suggesting a partial
381 dissociation between the internal evidence used for making perceptual decisions and confidence
382 evaluations.

383 We used the representation from the confidence cluster as an estimate of the internal evidence on which
384 observers base their confidence ratings. We then took the difference from L in the estimate of L^* from the
385 cluster representation as an estimate of the single-sample inference error. This estimate of the single-sample
386 inference error was significantly correlated with the single-sample inference error estimated from the
387 computational model of confidence ratings ($t(19) = 5.12$, $p < 0.001$), and this correlation was significantly
388 greater than the correlation with the error estimated from the model of perceptual decisions alone ($t(19) =$

389 2.62, $p = 0.017$; see **Supplementary Note 8**). This suggests that this cluster representation is indeed
390 reflecting activity specific to the computation of confidence.

391 We asked what processes were responsible for driving variability in the internal evidence for confidence
392 beyond what could be explained by the evidence presented to the observer. We selected 'Noise Min' and
393 'Noise Max' epochs as the top and bottom quartile of epochs sorted by the estimate of the inference error
394 from the cluster representation, and examined the source-localised EEG activity across these epochs. The
395 presented sensory evidence was similar across Noise Min and Noise Max epochs (see **Supplementary Note**
396 **8**), but the additional variability in the Noise Max epochs pushes the represented evidence further from the
397 mean, and should therefore correspond to a greater absolute normalised signal. We estimated the sources of
398 activity in the Noise Min and Noise Max epochs using a template brain (see **Methods**) and tested for
399 differences in the rectified normalised current density in ROIs defined based on the previous literature
400 (**Figure 5c**; Graziano, Parra, and Sigman, 2015; Gherman and Philiastides, 2018; Herding et al., 2019, see
401 **Supplementary Note 9**). As expected, Noise Max epochs showed a greater increase in current density
402 power over time. Significant differences first emerged in the superior parietal cortex (**Figure 5d**; 276 - 304
403 ms; $t_{avg}(19) = 2.37$, $p_{cluster} = 0.016$, re-emerging at 596 - 748 ms; $t_{avg}(19) = 2.53$, $p_{cluster} = 0.016$; and 912 ms;
404 $t_{avg}(19) = 2.50$, $p_{cluster} = 0.014$), and then in the orbitofrontal cortex (OFC; 516 - 556 ms; $t_{avg}(19) = 2.30$, $p_{cluster}$
405 = 0.022, re-emerging at 660 - 772 ms; $t_{avg}(19) = 2.79$, $p_{cluster} = 0.032$, and 824 - 1000 ms; $t_{avg}(19) = 2.60$,
406 $p_{cluster} = 0.022$). No differences in the rostral middle frontal cortex nor lateral occipital cortex survived cluster
407 correction.

408 Whilst the activity localised to the superior parietal cortex reflected stimulus driven computations (the
409 consecutive peaks correspond temporally to the response to subsequent stimuli), the activity localised to the
410 orbitofrontal cortex was more indicative of an accumulation process across samples (a smoother increase in
411 signal over time). As an exploratory analysis, we tested whether the activity localised to the orbitofrontal
412 cortex could predict observers' confidence ratings, presumably by accumulating evidence for evaluating
413 confidence up to the observers' perceptual decision response. Indeed, the activity localised to the
414 orbitofrontal cortex predicted observers' confidence ratings, based on the predictions of a generalised linear
415 model with 90/10 cross validation: the standardised regression coefficients increased up to and continued
416 after the perceptual decision response (**Figure 5e**, a significant cluster was located from -300 to 520 ms
417 around the time of the response; $t_{ave}(19) = 3.46$, cluster-corrected $p < 0.001$).



418

419 **Figure 5. Clusters of behaviourally relevant representations and their sources.** **a)** Log likelihood ratio
 420 (LLR) of the data given the hypothesis that decoding precision varies with behavioural suboptimalities, against
 421 the null hypothesis that decoding precision varies only with measurement noise. Perceptual (Type-I) behaviour
 422 is shown on top and confidence (Type-II) behaviour is shown on the bottom. Clusters where the log posterior
 423 odds ratio outweighed the prior are circled, only the bold area of the perceptual cluster was further analysed.
 424 Time series (left) show the maximum LLR of electrodes laterally, with frontal polar electrodes at the top
 425 descending to occipital electrodes at the bottom. Scalp maps (right) show the summed LLR over the indicated
 426 time windows. **b)** Left: representation precision (z) training and testing on signals within the clusters. Colours
 427 correspond to the circles in a), with the dark green bar showing the combined decoding precision of the
 428 anterior and posterior confidence clusters, and the black bar showing the combined representation precision of
 429 all clusters. Right: Representation precision of the last four samples in the Less and the More conditions for the
 430 combined confidence representation and the perceptual representation. Error bars show 95% within-subject
 431 confidence intervals. **c)** ROIs (defined by mindBoggle coordinates; Klein et al., 2017): lateral occipital cortex
 432 (blue); superior parietal cortex (green); orbitofrontal cortex (orange); and rostral middle frontal cortex (red).
 433 **d)** ROI time series for Noise Max (black) and Noise Min (coloured) epochs, taking the average rectified

434 *normalised current density (z) across participants. Shaded regions show 95% within-subject confidence*
435 *intervals, red horizontal lines indicate cluster corrected significant differences. Standardised within-subject*
436 *differences are traced above the x-axis, with the shaded region marking $z = 0$ to $z = 1.96$ (95% confidence).e)*
437 *Standardised regression weight (t-statistic) of the GLM comparing observers' confidence ratings to those*
438 *predicted from the activity localised to the orbitofrontal cortex. The shaded region shows the 95% between*
439 *subject confidence interval, and the red horizontal line marks the time-window showing cluster-corrected*
440 *significant differences from 0.*

441 **Discussion**

442 We examined the dynamic neural signals associated with the accumulation of evidence for evaluating
443 confidence in perceptual decisions. Observers were required to integrate evidence over multiple samples
444 provided by a sequence of visual stimuli. When observers were unable to control the amount of evidence
445 they were exposed to, they employed a covert decision bound, committing to perceptual decisions when
446 they had enough evidence, even if stimulus presentation continued. We had previously shown evidence for
447 this premature decision commitment based on behaviour and computational modelling (Balsdon, Wyart and
448 Mamassian, 2020). We replicated these results here, and further examined the neural signatures of covert
449 decision making. We found that the distribution of spectral power associated with the preparation and
450 execution of motor responses in the Free task (where the response is entered as soon as the decision is
451 made) could be used to accurately predict responses in the More condition of the Replay task over 1 s prior
452 to when the response was entered, and with significantly greater sensitivity than in the Less condition
453 (when observers were unlikely to have committed to a decision early). This suggests that covert decisions
454 could trigger the motor preparation for pressing the response key. Moreover, the strength of the eventual
455 motor response signal could be predicted by earlier decision evidence in the More condition, as if observers
456 are maintaining some representation of the decision evidence whilst waiting to press the response key.

457 Based on the evoked representation of accumulated evidence, perceptual decision accuracy relied on a flow
458 of information processing from early occipital and parietal signals, which then spread through to anterior
459 electrodes. When observers committed to perceptual decisions prematurely, only the early part of the
460 representation of accumulated evidence was attenuated. This selective dampening of the representation of
461 accumulated evidence following premature decision commitment delineates which computations are
462 devoted solely to the perceptual decision process, and which computations reflect the input to the decision
463 process: The representations of stimulus orientation and decision update (Wyart et al., 2012; Wyart et al.,
464 2015; Weiss et al., 2021), which are necessary input for the perceptual decision, did not substantially change
465 after committing to a perceptual decision. This initial perceptual processing stage likely remained important
466 for the continued accumulation of evidence for evaluating confidence (even after the completion of
467 perceptual decision processes), though it could also be that these processes are automatically triggered by
468 stimulus onset irrespective of whether the evidence is being accumulated for decision-making.

469 Confidence should increase with increasing evidence for the perceptual decision. It is therefore unsurprising

470 that the neural correlates of confidence magnitude have found similar EEG markers as those related to the
471 accumulation of the underlying perceptual decision evidence: the P300 (Gherman and Philiastides, 2015;
472 Desender et al., 2016; Desender et al., 2019; Zakrzewski et al., 2019; Rausch et al., 2020); and Central
473 Parietal Positivity (CPP; Boldt et al., 2019; Herding et al., 2019, indeed we show a similar effect in
474 **Supplementary Note 4**). The analysis presented in this manuscript targeted confidence precision rather
475 than confidence magnitude, by assessing confidence relative to an optimal observer who gives high
476 confidence ratings on trials where the evidence in favour of the perceptual choice is greater than the median
477 across trials. We isolated part of the neural representation of accumulated evidence where imprecision
478 relative to the optimal presented evidence predicted greater deviations from optimal in the internal
479 representation of evidence used for confidence evaluation implied from behaviour. The internal evidence
480 predicted from this neural representation was also more strongly related to the evidence for confidence than
481 the evidence used for perceptual decisions based on the computational model fit to describe behaviour.

482 We analysed the sources of activity more closely representing the internal evidence on which the confidence
483 evaluation was based than the optimal presented evidence. Activity localised to the Superior Parietal and
484 Orbitofrontal cortices was found to track this internal evidence for confidence throughout decision-making.
485 This is not at odds with the previous literature: The difference in superior parietal cortex could be linked
486 with findings from electrophysiology that suggest that confidence is based on information coded in parietal
487 cortex, where the underlying perceptual decision evidence is integrated (Kiani et al., 2009; Rutishauser et al.,
488 2018; though at least a subset of these neurons reflect bounded accumulation, which is in contrast with the
489 continued confidence accumulation described in this experiment; Kiani, Hanks, and Shadlen, 2007). Early
490 electrophysiological investigation into the function of the orbitofrontal cortex revealed neural coding
491 associated with stimulus value (Thorpe, Rolls, and Maddison, 1983), which has since been linked with a
492 confidence-modulated signal of outcome-expectation (Kepecs et al., 2008; and in human fMRI; Rolls,
493 Grabenhorst, and Deco, 2010) and recently, shown to be domain-general (single OFC neurons were
494 associated with confidence in both olfactory and auditory tasks; Masset et al., 2020). The source localisation
495 analysis therefore connects previous findings, indicating confidence feeds off an evidence accumulation
496 process, culminating in higher-order brain areas that appear to function for guiding outcome-driven
497 behaviour based on decision certainty.

498 These neural signatures of confidence evidence encoding were present throughout the process of making a
499 perceptual decision. This is in line with more recent evidence suggesting that confidence could be computed
500 online, alongside perceptual evidence accumulation (Zizlsperger et al., 2014; Gherman and Philiastides,
501 2015; Balsdon et al., 2020), as opposed to assessing the evidence in favour of the perceptual decision only
502 after committing to that decision. Computational model comparison supported this interpretation, showing
503 the best description of confidence behaviour was an accumulation process that was partially dissociable
504 from perceptual evidence accumulation (**Supplementary Note 1**; replicating our previous analysis, Balsdon
505 et al., 2020). This partial dissociation mediates the ongoing debate between single-channel (for example,
506 Maniscalco and Lau, 2016) and dual-channel (for example, Charles, King, and Deheane 2014) models, as it

507 constrains confidence by perceptual suboptimalities, at the same time as allowing additional processing to
508 independently shape confidence. The combination of this partial dissociation and online monitoring could
509 allow for metacognitive control of perceptual evidence accumulation, to flexibly balance perceptual accuracy
510 against temporal efficiency, by bounding perceptual evidence accumulation according to contemporaneous
511 confidence.

512 Using this protocol, we were able to delineate two distinct representations of accumulated evidence which
513 correspond to perceptual decision-making and confidence evaluations. These neural representations were
514 partially dissociable in that the perceptual representation neglected additional evidence following
515 premature decision commitment whilst the confidence representation continued to track the updated
516 evidence independently of decision commitment. This partial dissociation validates the predictions of the
517 computational model and provides a framework for the cognitive architecture underlying the distinction
518 between perception and confidence. That the neural resources involved in the confidence representation can
519 be recruited independently of perceptual processes implies a specific neural circuit for the computation of
520 confidence, a necessary feature of a general metacognitive mechanism flexibly employed to monitor the
521 validity of any cognitive process.

522 **Methods**

523 **Participants**

524 A total of 20 participants were recruited from the local cognitive science mailing list (RISC) and by word of
525 mouth. No participant met the pre-registered
526 (https://osf.io/346pe/?view_only=ddbc092996f34438964cf45a239498bb) exclusion criteria of chance-
527 level performance or excessive EEG noise. Written informed consent was provided prior to commencing the
528 experiment. Participants were required to have normal or corrected to normal vision. Ethical approval was
529 granted by the INSERM ethics committee (ID RCB: 2017-A01778-45 Protocol C15-98).

530 **Materials**

531 Stimuli were presented on a 24" BenQ LCD monitor running at 60 Hz with resolution 1920x1080 pixels and
532 mean luminance 45 cd/m². Stimulus generation and presentation was controlled by MATLAB (Mathworks)
533 and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007), run on a Dell Precision
534 M4800 Laptop. Observers viewed the monitor from a distance of 57 cm, with their head supported by a chin
535 rest. EEG data were collected using a 64-electrode BioSemi ActiveTwo system, run on a dedicated mac laptop
536 (Apple Inc.), with a sample rate of 512 Hz. Data were recorded within a shielded room.

537 **Stimuli**

538 Stimuli were oriented Gabor patches displayed at 70% contrast, subtending 4 dva and with spatial frequency
539 2 cyc/deg. On each trial a sequence of stimuli was presented, at an average rate of 3 Hz, with the stimulus
540 presented at full 70% contrast for a variable duration between 50 and 83 ms, with a sudden onset, followed
541 by an offset ramp over two flips, where the stimulus contrast decreased by 50% and 75% before complete

542 offset. Stimulus onset timing was jittered within the stimulus presentation interval such that the timing of
543 stimulus onset was irregular but with at least 216 ms between stimuli. These timings and stimulus examples
544 are shown in **Figure 1a**.

545 On each trial the orientations of the presented Gabors were drawn from one of two circular Gaussian (Von
546 Mises) distributions centred on $\pm 45^\circ$ from vertical (henceforth referred to as the 'orange' and 'blue'
547 distributions respectively), with concentration $\kappa = 0.5$ (shown in **Figure 1d**). Stimuli were displayed within
548 an annular 'colour-guide' where the colour of the annulus corresponds to the probability of the orientation
549 under each distribution, using the red and blue RGB channels to represent the probabilities of each
550 orientation under each distribution. Stimuli were presented in the centre of the screen, with a black central
551 fixation point to guide observers' gaze.

552 **Procedure**

553 The task was a modified version of the weather prediction task (Knowlton et al., 1996; Drugowitsch et al.,
554 2016). Throughout the experiment, the observer's perceptual task was to categorise which distribution the
555 stimulus orientations were sampled from. They were instructed to press the 'd' key with their left hand (of a
556 standard querty keyboard) for the blue distribution and the 'k' key with their right hand for the orange
557 distribution. There were two variants of the task: The Free task and the Replay task. The trials were
558 composed of three repetitions of 100 predefined sequences of up to 40 samples (50 trials from each
559 distribution) for each observer (300 trials per task).

560 In the 'Free' task, observers were continually shown samples (up to 40) until they entered their response.
561 They were instructed to enter their response as soon as they 'feel ready' to make a decision, with emphasis
562 on both accuracy (they should make their decision when they feel they have a good chance of being correct)
563 and on time (they shouldn't take too long to complete each trial). A graphical description of this task is
564 shown in **Figure 1b**.

565 After completing the Free task, observers then completed the Replay task. In this task they were shown a
566 specific number of samples and could only enter their response after the sequence finished, signalled by the
567 fixation point turning red. The number of samples was determined based on the number observers chose to
568 respond to in the Free task. There were three intermixed conditions: In the Less condition observers were
569 shown two fewer samples than the minimum they had chosen to respond to on that predefined sequence in
570 the Free task; In the Same condition observers were shown the median number of samples from that
571 predefined sequence; in the More condition observers were shown four additional samples compared to the
572 maximum number they chose to respond to on that sequence in the Free task. After entering their
573 perceptual (Type-I) response, observers were cued to give a confidence rating (Type-II decision). The
574 confidence rating was given on a 4-point scale where 1 represents very low confidence that the perceptual
575 decision was correct, and 4, certainty that the perceptual decision was correct. The rating was entered by
576 pressing the 'space bar' when a presented dial reached the desired rating. The dial was composed of a black
577 line which was rotated clockwise to each of 4 equidistant angles (marked 1 - 4) around a half circle, at a rate

578 of 1.33 Hz. The dial started at a random confidence level on each trial and continued updating until a rating
579 was chosen. A graphical description of this task is shown in **Figure 1c**.

580 Prior to commencing the experimental trials, participants were given the opportunity to practice the
581 experiment and ask questions. They first performed 20 trials of a fixed number of samples with only the
582 perceptual decision, with feedback after each response as to the true category. They then practiced the
583 Replay task with the confidence rating (and an arbitrary number of samples). Finally, they practiced the Free
584 task, before commencing the experiment with the Free task.

585 **Analysis**

586 **Behaviour**

587 Perceptual (Type-I) decisions were evaluated relative to the category the orientations were actually drawn
588 from. Performance is presented as proportion correct, whilst statistical analyses were performed on
589 sensitivity (d'). Sensitivity was calculated based on the proportion of hits (responding “Category A” when
590 category A was presented) and false alarms (responding “Category A” when category B was presented).
591 Confidence was evaluated relative to an optimal observer who gives high confidence when the log-likelihood
592 of the chosen category, based on the presented orientations, is above the median across trials, and low
593 confidence on trials with less than the median log-likelihood. More broadly, confidence should increase with
594 increasing evidence in favour of the perceptual decision, see **Supplementary Note 3**. A General Linear
595 Model was used to validate the influence of the optimal presented evidence on perceptual decisions and
596 confidence evaluations. The accumulated evidence up to the final sample and four samples before the
597 response was used as a regressor for the perceptual decision assuming a binomial distribution with a probit
598 link function. A comparable analysis was performed for confidence by binarizing confidence ratings into Low
599 (ratings of 1 or 2) and High (ratings of 3 or 4) and taking the evidence signed by the perceptual decision.

600 **Computational modelling**

601 Computational modelling followed the same procedure as Balsdon, Wyart, and Mamassian (2020). The
602 model parametrically describes suboptimalities relative to the Bayesian optimal observer. The Bayesian
603 optimal observer knows the category means, $\mu_1 = -\frac{\pi}{4}$, $\mu_2 = \frac{\pi}{4}$, and the concentration, $\kappa = 0.5$, and takes the
604 probability of the orientation θ_n (at sample n) given each category ψ ($\psi = 1$ or $\psi = 2$)

$$p(\theta_n | \psi) = \frac{e^{\kappa \cos(2(\theta_n - \mu_\psi))}}{\pi I_0(\kappa)} \quad (1)$$

605 where $I_0(\cdot)$ is the modified Bessel function of order 0. The optimal observer then chooses the category
606 ψ with the greatest posterior probability over all samples for that trial, T (T varies from trial to trial). Given a
607 uniform category prior, $p(\psi) \propto \frac{1}{2}$, and perfect anticorrelation in $p(\theta_n | \psi)$ over the categories, the log
608 posterior is proportional to the sum of the difference in the log-likelihood for each category ($\ell_n = \ell_{n,1} -$
609 $\ell_{n,2}$)

$$L = \sum_{n=1}^T \ell_n \quad (2)$$

610 where:

$$\ell_{n,\psi} = \log p(\theta_n | \psi) = \kappa \cos(2(\theta_n - \mu_\psi)) + \text{const.} \quad (3)$$

611 Such that the Bayesian optimal decision is 1 if $z > 0$ and 2 if $z \leq 0$.

612 The suboptimal observer suffers inaccuracies in the representation of each evidence sample, captured by
613 additive independent identically distributed (i.i.d) noise, ε_n . The noise is Gaussian distributed with zero
614 mean, and the degree of variability parameterised by σ , the standard deviation

$$\varepsilon_n \sim N(0, \sigma^2) \quad (4)$$

615 The evidence over samples is also imperfectly accumulated, incurring primacy or recency biases
616 parameterised by α , the weight on the current accumulated evidence compared to the new sample ($\alpha > 1$
617 creates a primacy effect). By the end of the trial, the weight on each sample n is equal to

$$v_n = \alpha^{T-n} \quad (5)$$

618 where T is the eventual total samples on that trial and $n \in [1, T]$.

619 In the Free task the observer responds when accumulated evidence reaches a bound, Λ . The optimal
620 observer sets a constant bound on proportion correct over sequence length, which is an exponential function
621 on the average evidence over the samples accumulated. The human observer can set the scale, b , and the rate
622 of decline, λ , of the bound suboptimally, resulting in

$$\Lambda_{n+} = n \times \left(a + b e^{-\frac{n}{\lambda}} \right) \quad (6)$$

623 for the positive decision bound (the negative bound, $\Lambda_{n-} = -\Lambda_{n+}$). The likelihood $f(n)$ of responding at
624 sample n was estimated by computing the frequencies, over 1000 samples from ε_n (Monte Carlo simulation),
625 of first times where the following inequality is verified

$$\left| \sum_{n=1}^N (\ell_n + \varepsilon_n) \cdot v_n \right| > \Lambda_n \quad (7)$$

626 The response time, relative to reaching the decision bound, is delayed by non-decision time for executing the
627 motor response, which is described by a Gaussian distribution of mean, μ_U , and variance, σ_U^2 .

628 **Model fitting**

629 Parameters were optimised to minimise the negative log-likelihood of the observer making response r on
630 sample n on each trial for each participant using Bayesian Adaptive Direct Search (Acerbi and Ma, 2017). The
631 log-likelihoods were estimated using Monte Carlo Simulation, with the sensitivity of this approach being
632 addressed in previous work (Balsdon et al., 2020). The full model was simplified using a knock-out
633 procedure based on Bayesian Model Selection (Rigoux et al., 2014) to fix the bias (exceedance probability =

634 0.93) and lapse (exceedance probability >0.99) parameters (not described above, see **Supplementary Note**
635 **1**).

636 In the Replay task, confidence ratings were fit using the same model described above, but with additional
637 criteria determining confidence ratings, described by three bounds on the confidence evidence,
638 parameterised in the same manner as the decision bound. These models were then used to simulate the
639 internal evidence of each observer from sample to sample, and the error compared to the optimal evidence
640 (uncorrupted by suboptimalities, see **Supplementary Note 2**).

641 **EEG pre-processing**

642 EEG data were pre-processed using the PREP processing pipeline (Bigdely-Shamlo, et al., 2015),
643 implemented in EEGLab (v2019.0, Delorme & Makeig, 2004) in MATLAB (R2019a, Mathworks). This includes
644 line noise removal (notch filter at 50 Hz and harmonics) and re-referencing (robust average re-reference on
645 data detrended at 1 Hz). The data were then filtered to frequencies between 0.5 and 80 Hz, and down-
646 sampled to 256 Hz. Large epochs were taken locked to each stimulus (-500 to 1500 ms) and each response (-
647 5000 to 1500 ms). Independent Components Analysis was used to remove artefacts caused by blinks and
648 excessive muscle movement identified using labels with a probability greater than 0.35 from the ICLabel
649 project classifier (Swartz Centre for Computational Neuroscience).

650 **Response classification analysis**

651 The power spectrum across frequency tapers from 1 to 64 Hz with 25% spectral smoothing was resolved
652 using wavelet convolution implemented in FieldTrip (Oostenveld et al., 2011). The epochs were then clipped
653 at -3 to 1 s around the time of entering the perceptual response. Linear discriminant analysis was
654 performed to classify perceptual responses, using 10-fold cross validation, separately on each taper at each
655 time-point. An analysis of the frequencies contributing to accurate classification at the time of the response
656 revealed significant contributions from 8 to 26 Hz (**Supplementary Note 4**). We therefore continued by
657 using the power averaged across these frequency bands to train and test the classifier. Classifier accuracy
658 was assessed using the area under the receiver operating characteristic curve (AUC). At the single-trial level,
659 the probability of the response based on the classifier was computed from the relative normalised Euclidean
660 distance of the trial features from the response category means in classifier decision space.

661 **Encoding Variable Regression**

662 We used a linear regression analysis to examine the EEG correlates of different aspects of the decision
663 evidence (encoding variables) in epochs locked to stimulus onset. Regularised ridge regression (ridge $\lambda = 1$)
664 was used to predict the encoding variables based on EEG data, over 10-fold cross validation. The precision of
665 the representation of each encoding variable was computed within each observer by taking the Fisher
666 transform of the correlation coefficient (Pearson's r) between the encoded variable and predicted variable.
667 To maximise representation precision, the data were bandpass filtered (1 – 8 Hz) and decomposed into real
668 and imaginary parts using a Hilbert Transform (**Supplementary Note 5**). For each time point, the data from
669 all electrodes were used to predict the encoded variable. The temporal generalisation of decoding weights
670 was examined by training at one time point and testing at another. The contribution of information from

671 signals at each electrode was examined by training and testing on the signals at each electrode at each time
672 point (further details in **Supplementary Note 5**).

673 Behaviourally relevant signals were isolated by comparing representation precision at each time point and
674 electrode for epochs leading to optimal perceptual and confidence responses, compared to responses that
675 did not match the optimal observer. Cluster modelling was used to isolate contiguous signals where the log
676 posterior odds were in favour of the alternative hypothesis that the representation systematically deviated
677 further from the optimal presented evidence than what could be explained by measurement noise alone
678 (**Supplementary Note 6**). New regression weights were then calculated on signals from the entire cluster
679 and representation errors calculated as the difference of the predicted variable from the expected value
680 given the representation.

681 **Source Localisation**

682 Identifying the clusters of signals associated with confidence processes offers relatively poor spatial and
683 temporal (given the bandpass filter; de Cheveigné, and Nelken, 2019) resolution for identifying the source of
684 confidence computations. Source localisation was therefore performed, using Brainstorm (Tadel et al.,
685 2011). The forward model was computed using OpenMEEG (Gramfort et al., 2010; Kybic et al., 2005) and the
686 ICBM152 anatomy (Fonov et al., 2011; 2009). Two conditions were compared, Noise Min and Noise Max,
687 which corresponded to quartiles of epochs sorted by representation error in the confidence clusters (see
688 **Supplementary Note 7** for more details). Cortical current source density was estimated from the average
689 epochs using orientation-constrained minimum norm imaging (Baillet, Mosher, and Leahy, 2001). ROIs in
690 the Lateral Occipital, Superior Parietal, Rostral Middle Frontal (including dlPFC), Medial Orbitofrontal, and
691 rostral Anterior Cingulate Cortex, were defined using MindBoggle coordinates (Klein et al., 2017). Statistical
692 comparisons were performed on the bilateral ROI time series (using cluster correction and a minimum
693 duration of 20 ms), computed over separate conditions on rectified normalised subject averages (low-pass
694 filtered at 40 Hz).

695 To predict confidence magnitude from the activity localised to the orbitofrontal cortex, we recovered to
696 current density from 20 subregions (approximately equal parcellations) of the orbitofrontal cortex in epochs
697 locked to the time of the response. A general linear model (assuming a normal distribution with identity
698 link) was used to predict the observers' confidence ratings on held-out data (90/10 cross-fold) from the
699 neural activity at each time-point leading to the response. The prediction was quantified as the standardised
700 regression weight from a new general linear model comparing the predicted and actual confidence ratings
701 across all folds.

702 References

- 703 Acerbi, L., & Ma, W. J. Practical Bayesian optimization for model fitting with Bayesian adaptive direct search.
704 In *Advances in Neural Information Processing Systems*, December 2017; 1836-1846
- 705 Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. What failure in collective decision-making
706 tells us about metacognition. *Philosophical Transactions of the Royal Society B: Biological*
707 *Sciences*, 2012; **367**(1594), 1350-1365.
- 708 Baillet, S., Mosher, J. C., & Leahy, R. M. Electromagnetic brain mapping. *IEEE Signal Processing*
709 *Magazine*, 2001; **18**(6), 14-30.
- 710 Balsdon, T., Wyart, V., & Mamassian, P. Confidence controls perceptual evidence accumulation. *Nature*
711 *Communications*, 2020; **11**(1), 1-11
- 712 Bang, J. W., Shekhar, M., & Rahnev, D. Sensory noise increases metacognitive efficiency. *Journal of*
713 *Experimental Psychology: General*, 2019; **148**(3), 437.
- 714 Baranski, J. V., & Petrusic, W. M. The calibration and resolution of confidence in perceptual
715 judgments. *Perception & Psychophysics*, 1994; **55**(4), 412-428.
- 716 Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., & Robbins, K. A. The PREP pipeline: standardized
717 preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, 2015; **9**, 16
- 718 Boldt, A., Schiffer, A. M., Waszak, F., & Yeung, N. Confidence predictions affect performance confidence and
719 neural preparation in perceptual decision making. *Scientific Reports*, 2019; **9**(1), 1-17.
- 720 Ruby, E., Maniscalco, B., & Peters, M. A. On a 'failed' attempt to manipulate visual metacognition with
721 transcranial magnetic stimulation to prefrontal cortex. *Consciousness and cognition*, 2018; **62**, 34-41.
- 722 Brainard, D. H. The psychophysics toolbox. *Spatial Vision*, 1997; **10**(4), 433-436.
- 723 Charles, L., King, J. R., & Dehaene, S. Decoding the dynamics of action, intention, and error detection for
724 conscious and subliminal stimuli. *Journal of Neuroscience*, 2014; **34**(4), 1158-1170.
- 725 Cortese, A., Amano, K., Koizumi, A., Kawato, M., & Lau, H. Multivoxel neurofeedback selectively modulates
726 confidence without changing perceptual performance. *Nature communications*, 2016; **7**(1), 1-18.
- 727 de Cheveigné, A., & Nelken, I. Filters: when, why, and how (not) to use them. *Neuron*, 2019; **102**(2), 280-293.
- 728 Delorme, A., & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including
729 independent component analysis. *Journal of Neuroscience Methods*, 2004; **134**(1), 9-21.
- 730 Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. Humans incorporate attention-dependent uncertainty
731 into perceptual decisions and confidence. *Proceedings of the National Academy of Sciences*, 2018
732 **115**(43), 11090-11095.
- 733 Desender, K., Van Opstal, F., Hughes, G., & Van den Bussche, E. The temporal dynamics of metacognition:
734 Dissociating task-related activity from later metacognitive processes. *Neuropsychologia*, 2016; **82**, 54-
735 64.
- 736 Desender, K., Murphy, P., Boldt, A., Verguts, T., & Yeung, N. A postdecisional neural marker of confidence
737 predicts Information-Seeking in Decision-Making. *Journal of Neuroscience*, 2019; **39**(17), 3309-3319.
- 738 Drugowitsch, J., Wyart, V., Devauchelle, A. D., & Koechlin, E. Computational precision of mental inference as
739 critical source of human choice suboptimality. *Neuron*, 2016; **926**, 1398-1411

- 740 Fleming, S. M., Huijgen, J., & Dolan, R. J. Prefrontal contributions to metacognition in perceptual decision
741 making. *Journal of Neuroscience*, 2012; **32**(18), 6117-6125.
- 742 Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. Domain-specific impairment in metacognitive accuracy
743 following anterior prefrontal lesions. *Brain*, 2014; **137**(10), 2811-2822.
- 744 Fleming, S. M., & Daw, N. D. Self-evaluation of decision-making: A general Bayesian framework for
745 metacognitive computation. *Psychological Review*, 2017; **124**(1), 91.
- 746 Fleming, S. M., Van Der Putten, E. J., & Daw, N. D. Neural mediators of changes of mind about perceptual
747 decisions. *Nature neuroscience*, 2018; **21**(4), 617-624.
- 748 Fonov VS, Evans AC, McKinstry RC, Almlí CR, Collins DL. Unbiased nonlinear average age-appropriate brain
749 templates from birth to adulthood. *NeuroImage*, 2009; **47**, S102.
- 750 Fonov, V., Evans, A. C., Botteron, K., Almlí, C. R., McKinstry, R. C., Collins, D. L., & Brain Development
751 Cooperative Group. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, 2011;
752 **54**(1), 313-327.
- 753 Frith, C. D. The role of metacognition in human social interactions. *Philosophical Transactions of the Royal
754 Society B: Biological Sciences*, 2012; **367**(1599), 2213-2223.
- 755 Gherman, S., & Philiastides, M. G. Human VMPFC encodes early signatures of confidence in perceptual
756 decisions. *eLife*, 2018; **7**, e38293.
- 757 Gherman, S., & Philiastides, M. G. Neural representations of confidence emerge from the process of decision
758 formation during perceptual choices. *NeuroImage*, 2015; **106**, 134-143.
- 759 Gramfort, A., Papadopoulos, T., Olivi, E., & Clerc, M. OpenMEEG: opensource software for quasistatic
760 bioelectromagnetics. *Biomedical Engineering Online*, 2010; **9**(1), 45.
- 761 Graziano, M., Parra, L. C., & Sigman, M. Neural correlates of perceived confidence in a partial report
762 paradigm. *Journal of Cognitive Neuroscience*, 2015; **27**(6), 1090-1103.
- 763 Geurts, L. S., Cooke, J. R., van Bergen, R. S., & Jehee, J. F. Subjective confidence reflects representation of
764 Bayesian probability in cortex. 2021; *bioRxiv*.
- 765 Helmholtz, H.L.F.v. *Treatise on Physiological Optics*, Thoemmes Press 1856.
- 766 Herding, J., Ludwig, S., von Lutz, A., Spitzer, B., & Blankenburg, F. Centro-parietal EEG potentials index
767 subjective evidence and confidence during perceptual decision making. *NeuroImage*, 2019; **201**,
768 116011.
- 769 Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. Neural correlates, computation and behavioural impact
770 of decision confidence. *Nature*, 2008; **455**, 227-231.
- 771 Kiani, R., & Shadlen, M. N. Representation of confidence associated with a decision by neurons in the parietal
772 cortex. *Science*, 2009; **324**, 759-764.
- 773 Kiani, R., Corthell, L., & Shadlen, M. N. Choice certainty is informed by both evidence and decision
774 time. *Neuron*, 2014; **84**(6), 1329-1342.
- 775 Kiani, R., Hanks, T. D., & Shadlen, M. N. Bounded integration in parietal cortex underlies decisions even when
776 viewing duration is dictated by the environment. *Journal of Neuroscience*, 2008; **28**(12), 3017-3029.
- 777 Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., ... & Keshavan, A. Mindboggling morphometry
778 of human brains. *PLoS Computational Biology*, 2017; **13**(2), e1005350.

- 779 Kleiner, M., Brainard, D., & Pelli, D. What's new in Psychtoolbox-3? 2007.
- 780 Knowlton, B. J., Mangels, J. A., & Squire, L. R. A neostriatal habit learning system in humans. *Science*, 1996;
781 **273**(5280), 1399-1402
- 782 Kybic, J., Clerc, M., Abboud, T., Faugeras, O., Keriven, R., & Papadopoulos, T. A common formalism for the
783 integral formulations of the forward EEG problem. *IEEE Transactions on Medical Imaging*, 2005; **24**(1),
784 12-28.
- 785 Lak, A., Costa, G. M., Romberg, E., Koulakov, A. A., Mainen, Z. F., & Kepecs, A. Orbitofrontal cortex is required
786 for optimal waiting based on decision confidence. *Neuron*, 2014; **84**(1), 190-201.
- 787 Lapate, R. C., Samaha, J., Rokers, B., Postle, B. R., & Davidson, R. J. Perceptual metacognition of human faces is
788 causally supported by function of the lateral prefrontal cortex. *Communications biology*, 2020; **3**(1), 1-
789 10.
- 790 Maniscalco, B., & Lau, H. The signal processing architecture underlying subjective reports of sensory
791 awareness. *Neuroscience of Consciousness*, 2016; **1**.
- 792 Masset, P., Ott, T., Lak, A., Hirokawa, J., & Kepecs, A. Behavior- and modality-general representation of
793 confidence in orbitofrontal cortex. *Cell*, 2020; **182**(1), 112-126.
- 794 Mazancieux, A., Fleming, S., Souchay, C., & Moulin, C. Retrospective confidence judgments across tasks:
795 domain-general processes underlying metacognitive accuracy. *BioRxiv* 2018.
- 796 Moreno-Bote, R. Decision confidence and uncertainty in diffusion models with partially correlated neuronal
797 integrators. *Neural Computation*, 2010; **22**, 1786-1811.
- 798 Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. Neural evidence accumulation persists after choice
799 to inform metacognitive judgments. *Elife*, 2015; **4**, e11946.
- 800 Neill, D. B. Fast Bayesian scan statistics for multivariate event detection and visualization. *Statistics in*
801 *Medicine*, 2011; **30**(5), 455-469.
- 802 Neill, D. B. Bayesian Scan Statistics. In: Glaz J., Koutras M. (eds) *Handbook of Scan Statistics*. Springer, New
803 York, NY. 2019.
- 804 Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. FieldTrip: open source software for advanced analysis of
805 MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011.
- 806 Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial*
807 *Vision*, 1997; **10**, 437-442.
- 808 Pleskac, T. J., & Busemeyer, J. R. Two-stage dynamic signal detection: a theory of choice, decision time, and
809 confidence. *Psychological Review*, 2010; **117**(3), 864.
- 810 Pollack, I., & Decker, L. R. Confidence ratings, message reception, and the receiver operating
811 characteristic. *The Journal of the Acoustical Society of America*, 1958; **30**(4), 286-292.
- 812 Ratcliff, R. A theory of memory retrieval. *Psychological Review*, 1987; **85**(2), 59.
- 813 Rausch, M., Zehetleitner, M., Steinhauser, M., & Maier, M. E. Cognitive modelling reveals distinct
814 electrophysiological markers of decision confidence and error monitoring. *NeuroImage*, 2020; **218**,
815 116963.
- 816 Rigoux, L., Stephan, K.E., Friston, K.J. & Daunizeau, J. Bayesian Model Selection for Group Studies Revisited.
817 *NeuroImage* 2014; **84**, 971-85.

- 818 Rolls, E. T., Grabenhorst, F., & Deco, G. Choice, difficulty, and confidence in the brain. *NeuroImage*, 2010;
819 **53**(2), 694-706.
- 820 Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. Theta-burst transcranial magnetic
821 stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive*
822 *neuroscience*, 2010; **1**(3), 165-175.
- 823 Rutishauser, U., Aflalo, T., Rosario, E. R., Pouratian, N., & Andersen, R. A. Single-neuron representation of
824 memory strength and recognition confidence in left human posterior parietal cortex. *Neuron*, 2018;
825 **97**(1), 209-220.
- 826 Shekhar, M., & Rahnev, D. Sources of Metacognitive Inefficiency. *Trends in Cognitive Sciences*.
- 827 Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. Brainstorm: a user-friendly application for
828 MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011.
- 829 Thorpe, S. J., Rolls, E. T., & Maddison, S. The orbitofrontal cortex: neuronal activity in the behaving
830 monkey. *Experimental Brain Research*, 1983; **49**(1), 93-115.
- 831 Vaccaro, A. G., & Fleming, S. M. Thinking about thinking: A coordinate-based meta-analysis of neuroimaging
832 studies of metacognitive judgements. *Brain and neuroscience advances*, 2018; **2**, 2398212818810591.
- 833 Veenman, M. V., Wilhelm, P., & Beishuizen, J. J. The relation between intellectual and metacognitive skills
834 from a developmental perspective. *Learning and Instruction*, 2004; **14**(1), 89-109.
- 835 Vickers, D. Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 1970; **13**(1),
836 37-58.
- 837 Vickers, D. *Decision processes in visual perception*. New York, NY: Academic Press. 1979.
- 838 Weiss, A., Chambon, V., Lee, J. K., Drugowitsch, J., & Wyart, V. Interacting with volatile environments
839 stabilizes hidden-state inference and its brain signatures. *Nature communications*, 2021; **12**(1), 1-17.
- 840 Wyart, V., De Gardelle, V., Scholl, J., & Summerfield, C. Rhythmic fluctuations in evidence accumulation during
841 decision making in the human brain. *Neuron*, 2012; **76**(4), 847-858.
- 842 Wyart, V., Myers, N. E., & Summerfield, C. Neural mechanisms of human perceptual choice under focused and
843 divided attention. *Journal of Neuroscience*, 2015; **35**(8), 3485-3498.
- 844 Yokoyama, O., Miura, N., Watanabe, J., Takemoto, A., Uchida, S., Sugiura, M., ... & Nakamura, K. Right
845 frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-
846 term recognition memory performance. *Neuroscience research*, 2010; **68**(3), 199-206.
- 847 Zakrzewski, A. C., Wisniewski, M. G., Iyer, N., & Simpson, B. D. Confidence tracks sensory-and decision-related
848 ERP dynamics during auditory detection. *Brain and Cognition*, 2019; **129**, 49-58
- 849 Zizlsperger, L., Sauvigny, T., Händel, B., & Haarmeier, T. Cortical representations of confidence in a visual
850 perceptual decision. *Nature Communications*, 2014; **5**(1), 1-13.
- 851