



HAL
open science

Behavioural, modeling, and electrophysiological evidence for supramodality in human metacognition

Nathan Faivre, Elisa Filevich, Guillermo Solovey, Simone Kühn, Olaf Blanke

► **To cite this version:**

Nathan Faivre, Elisa Filevich, Guillermo Solovey, Simone Kühn, Olaf Blanke. Behavioural, modeling, and electrophysiological evidence for supramodality in human metacognition. *Journal of Neuroscience*, 2017, 10.1523/JNEUROSCI.0322-17.2017 . hal-01668131

HAL Id: hal-01668131

<https://hal.science/hal-01668131>

Submitted on 19 Dec 2017

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

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

Research Articles: Behavioral/Cognitive

Behavioural, modeling, and electrophysiological evidence for supramodality in human metacognition

Nathan Faivre^{1,2,3}, Elisa Filevich^{4,5,6}, Guillermo Solovey⁷, Simone Kühn^{4,8} and Olaf Blanke^{1,2,9}

¹Laboratory of Cognitive Neuroscience, Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Geneva, Switzerland

²Center for Neuroprosthetics, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Geneva, Switzerland

³Centre d'Economie de la Sorbonne, CNRS UMR 8174, Paris, France

⁴Department of Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

⁵Department of Psychology, Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

⁶Bernstein Center for Computational Neuroscience Berlin, Philippstr. 13 Haus 6, 10115 Berlin, Germany

⁷Instituto de Cálculo, FCEyN, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina

⁸Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistrasse 52 20246 Hamburg, Germany

⁹Department of Neurology, University Hospital Geneva, Geneva, Switzerland

DOI: 10.1523/JNEUROSCI.0322-17.2017

Received: 3 February 2017

Revised: 4 August 2017

Accepted: 6 August 2017

Published: 15 September 2017

Author Contributions: N.F., and E.F., developed the study concept and contributed to the study design. Testing and data collection were performed by N.F. N.F., and E.F., performed the data analysis. E.F., and G.S., performed modeling work. N.F., and E.F., drafted the paper; all authors provided critical revisions and approved the final version of the paper for submission.

Conflict of Interest: The authors declare no competing financial interests.

NF was an EPFL fellow co-funded by Marie Skłodowska-Curie. O.B. is supported by the Bertarelli Foundation, the Swiss National Science Foundation, and the European Science Foundation. We thank Shruti Nanivadekar and Michael Stettler for their help during data collection.

Corresponding author: Nathan Faivre, Centre d'Economie de la Sorbonne, 106-112 Boulevard de l'Hopital, 75647 Paris cedex 13, France, Phone: +33 1 44 07 82 27, nathanfaivre@gmail.com

Cite as: J. Neurosci ; 10.1523/JNEUROSCI.0322-17.2017

Alerts: Sign up at www.jneurosci.org/cgi/alerts to receive customized email alerts when the fully formatted version of this article is published.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

Copyright © 2017 the authors

1 **Behavioural, modeling, and electrophysiological evidence**

2 **for supramodality in human metacognition**

3
4 Abbreviated title: Metacognition across senses and combination of senses

5
6 **Authors:** Nathan Faivre^{1,2,3,*}, Elisa Filevich^{4,5,6}, Guillermo Solovey⁷, Simone Kühn^{4,8}, Olaf Blanke^{1,2,9}

7 **Affiliations:**

8 1 Laboratory of Cognitive Neuroscience, Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Geneva,
9 Switzerland

10 2 Center for Neuroprosthetics, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Geneva, Switzerland

11 3 Centre d'Economie de la Sorbonne, CNRS UMR 8174, Paris, France

12 4 Department of Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

13 5 Department of Psychology, Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

14 6 Bernstein Center for Computational Neuroscience Berlin, Philippstr. 13 Haus 6, 10115 Berlin, Germany

15 7 Instituto de Cálculo, FCEyN, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina

16 8 Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistrasse 52 20246 Hamburg, Germany

17 9 Department of Neurology, University Hospital Geneva, Geneva, Switzerland

18
19 *** Corresponding author:**

20 Nathan Faivre, Centre d'Economie de la Sorbonne, 106-112 Boulevard de l'Hopital, 75647 Paris cedex 13, France

21 Phone: +33 1 44 07 82 27

22 nathanfaivre@gmail.com

23
24 Number of pages: 48

25 Number of figures: 7

26 Number of words: 150 (abstract), 680 (introduction), 1591 (discussion)

27 **Keywords:** metacognition, confidence, supramodality, audiovisual, EEG, signal detection theory

28
29 **Acknowledgements**

30 NF was an EPFL fellow co-funded by Marie Skłodowska-Curie. O.B. is supported by the Bertarelli Foundation,
31 the Swiss National Science Foundation, and the European Science Foundation. We thank Shruti Nanivadekar
32 and Michael Stettler for their help during data collection.

33
34 **Author Contributions**

35 N.F., and E.F., developed the study concept and contributed to the study design. Testing and data collection were
36 performed by N.F. N.F., and E.F., performed the data analysis. E.F., and G.S., performed modeling work. N.F.,
37 and E.F., drafted the paper; all authors provided critical revisions and approved the final version of the paper for
38 submission.

39
40 **Complementary results, raw behavioral data, and modeling scripts are available upon request.**

41

42 **Abstract (150)**

43 Human metacognition, or the capacity to introspect on one's own mental states, has been mostly
44 characterized through confidence reports in visual tasks. A pressing question is to what extent results
45 from visual studies generalize to other domains. Answering this question allows determining whether
46 metacognition operates through shared, supramodal mechanisms, or through idiosyncratic, modality-
47 specific mechanisms. Here, we report three new lines of evidence for decisional and post-decisional
48 mechanisms arguing for the supramodality of metacognition. First, metacognitive efficiency correlated
49 between auditory, tactile, visual, and audiovisual tasks. Second, confidence in an audiovisual task was
50 best modeled using supramodal formats based on integrated representations of auditory and visual
51 signals. Third, confidence in correct responses involved similar electrophysiological markers for visual
52 and audiovisual tasks that are associated with motor preparation preceding the perceptual judgment.
53 We conclude that the supramodality of metacognition relies on supramodal confidence estimates and
54 decisional signals that are shared across sensory modalities.

55 **Significance statement (118 words)**

56 Metacognitive monitoring is the capacity to access, report and regulate one's own mental states. In
57 perception, this allows rating our confidence in what we have seen, heard or touched. While
58 metacognitive monitoring can operate on different cognitive domains, we ignore whether it involves a
59 single supramodal mechanism common to multiple cognitive domains, or modality-specific
60 mechanisms idiosyncratic to each domain. Here, we bring evidence in favor of the supramodality
61 hypothesis by showing that participants with high metacognitive performance in one modality are
62 likely to perform well in other modalities. Based on computational modeling and electrophysiology, we
63 propose that supramodality can be explained by the existence of supramodal confidence estimates, and
64 by the influence of decisional cues on confidence estimates.

65

66

67 **Introduction (680 words)**

68 Humans have the capacity to access and report the contents of their own mental states including
69 percepts, emotions, and memories. In neuroscience, the reflexive nature of cognition is now the object
70 of research under the broad scope of the term metacognition (Koriat, 2006; Fleming et al., 2012). A
71 widely used method to study metacognition is to have observers do a challenging task (“first-order
72 task”), followed by a confidence judgment regarding their own task performance (“second-order task”,
73 Figure 1 left panel). In this operationalization, metacognitive accuracy can be quantified as the
74 correspondence between subjective confidence judgments and objective task performance. While some
75 progress has been made regarding the statistical analysis of confidence judgments (Galvin et al., 2003;
76 Maniscalco and Lau, 2012; Barrett et al., 2013), and more evidence has been gathered regarding the
77 brain areas involved in metacognitive monitoring (Grimaldi et al., 2015), the core properties and
78 underlying mechanisms of metacognition remain largely unknown. One of the central questions is
79 whether, and to what extent, metacognitive monitoring should be considered supramodal: is the
80 computation of confidence fully independent of the perceptual signal (i.e., supramodality), or does it
81 also involve signal-specific components? According to the supramodality hypothesis, metacognition
82 would have a quasi-homuncular status, the monitoring of all perceptual processes being operated
83 through a single shared mechanism. Instead, modality-specific metacognition would involve a
84 distributed network of monitoring processes that are specific for each sensory modality. The
85 involvement of supramodal, prefrontal brain regions during confidence judgments first suggested that
86 metacognition is partly governed by supramodal rules (Fleming et al., 2010; Yokoyama et al., 2010;
87 Rahnev et al., 2015). At the behavioural level, this is supported by the fact that metacognitive
88 performance (Song et al., 2011), and confidence estimates (de Gardelle and Mamassian, 2014; Rahnev
89 et al., 2015) correlate across subjects between two different visual tasks, as well as between a visual

90 and an auditory task (De Gardelle et al., 2016). However, the supramodality of metacognition is
91 challenged by the report of weak or null correlations between metacognitive performance across
92 different tasks involving vision, audition, and memory (Ais et al., 2016). Beyond sensory modalities,
93 metacognitive judgments across cognitive domains were shown to involve distinct brain regions
94 notably frontal areas for perception and precuneus for memory (McCurdy et al., 2013). Supporting this
95 view, patients with lesions to the anterior prefrontal cortex were shown to have a selective deficit in
96 metacognition for visual perception, but not memory (Fleming et al., 2014). This anatomo-functional
97 distinction across cognitive domains is further supported by the fact that meditation training improves
98 metacognition for memory, but not for vision (Baird et al., 2014). Compared to previous work, the
99 present study sheds new light on the issue of supramodality by comparing metacognitive monitoring of
100 stimuli from distinct sensory modalities, but during closely-matched first-order tasks. At the behavioral
101 level, we first investigated the commonalities and specificities of metacognition across sensory
102 domains including touch, a sensory modality that has been neglected so far. Namely, we examined
103 correlations between metacognitive performance during a visual, auditory, and tactile discrimination
104 task (Experiment 1). Next, extending our paradigm to conditions of audiovisual stimulation, we
105 quantified for the first time the links between unimodal and multimodal metacognition (Deroy et al.,
106 2016), and assessed through computational modeling how multimodal confidence estimates are built
107 (Experiment 2). This allowed us to assess if metacognition is supramodal because of a generic format
108 of confidence. Finally, we investigated the neural mechanisms of unimodal and multimodal
109 metacognition and repeated Experiment 2 while recording 64-channel electroencephalography (EEG,
110 Experiment 3). This allowed us to identify neural markers with high temporal resolution, focusing on
111 those preceding the response in the first-order task (ERPs, alpha suppression) to assess if metacognition
112 is supramodal because of the presence of decisional cues. The present data reveal #1 correlations in
113 metacognitive behavioral efficiencies across different unimodal and bimodal perception, #2

114 computational evidence for integrative, supramodal representations during audiovisual confidence
115 estimates, and #3 the presence of similar neural markers of supramodal metacognition preceding the
116 first-order task. Altogether, these behavioural, computational, and neural findings provide non-
117 mutually exclusive mechanisms explaining the supramodality of metacognition during human
118 perception.

119

120 [Figure 1 Here]

121 **Methods**

122 *Participants*

123 A total of 50 participants (Experiment 1: 15 including 8 females, mean age = 23.2 years, SD = 8.3
124 years; Experiment 2: 15 including 5 females, mean age = 21.3 years, SD = 2.6 years; Experiment 3: 20
125 including 6 females, mean age = 24.6 years, SD = 4.3 years) from the student population at the Swiss
126 Federal Institute of Technology (EPFL) took part in this study, in exchange for monetary compensation
127 (20 CHF per hour). All participants were right-handed, had normal hearing and normal or corrected-to-
128 normal vision, and no psychiatric or neurological history. They were naive to the purpose of the study
129 and gave informed consent, in accordance with institutional guidelines and the Declaration of Helsinki.
130 The data from two participants were not analyzed (one in Experiment 1 due to a technical issue with
131 the tactile device, and one from Experiment 2 as the participant could not perform the auditory task).

132 *Stimuli*

133 All stimuli were prepared and presented using the Psychophysics toolbox (Pelli, 1997; Brainard, 1997,
134 Kleiner et al., 2007; RRID:SCR_002881) in Matlab (Mathworks; RRID:SCR_001622). Auditory
135 stimuli consisted of either a 1100 Hz sinusoidal (high pitch “beep” sound) or 200 Hz sawtooth function

136 (low pitch “buzz” sound), played through headphones in stereo for 250 ms with a sampling rate of
137 44100 Hz. The loudness of one of the two stimuli was manipulated to control for task performance,
138 while the other stimulus remained constant. In phase 1, the initial inter-ear intensity difference was
139 50%, and increased (decreased) by 1% after each incorrect (two correct) answers. The initial difference
140 and step size were adapted based on individual performance. The initial difference in phase 2 was
141 based on the results from phase 1, and the step size remained constant. In the auditory condition of
142 Experiments 2, both sounds were played simultaneously in both ears, and were distinguished by their
143 timbre. When necessary, a correction of hearing imbalance was performed prior to the experiment to
144 avoid response biases.

145 Tactile stimuli were delivered to the palmar side of each wrist by a custom-made vibratory device,
146 using coin permanent-magnetic motors (9000 rpm maximal rotation speed, 9.8 N bracket deflection
147 strength, 55 Hz maximal vibration frequency, 22 m/s² acceleration, 30 ms delay after current onset)
148 controlled by a Leonardo Arduino board through pulse width modulation. Task difficulty was
149 determined by the difference in current sent to each motor, while one motor always received the same
150 current. In phase 1, the initial inter-wrist difference was 40%, and increased (decreased) by 2% after
151 each incorrect (two correct) answers. The initial difference and step size were adapted individually,
152 based on performance. A correction of tactile imbalance due to a difference of pressure between the
153 vibrator and the wrist was performed prior to the experiment to avoid response biases. The initial
154 difference in phase 2 was determined by the final difference from phase 1. The step size for the
155 stimulus staircase remained constant for both phases.

156 Visual stimuli consisted in pairs of two 5° x 5° Gabor patches (5 cycles/°, 11° center-to-center
157 distance). When only one pair of visual stimuli was presented (visual condition of Experiment 1,
158 audiovisual condition of Experiments 2 and 3), it was vertically centered on the screen. When two pairs

159 were presented (visual condition of Experiment 2 and 3), each pair was presented 5.5° above or below
160 the vertical center of the screen. Visual contrast of one Gabor of the pair was manipulated, while the
161 other always remained at constant contrast. The staircase procedure started with a difference of contrast
162 between Gabor patches of 40%, and an increment (decrement) of 2.5% after one incorrect (two correct)
163 answers.

164

165

166 *General procedure*

167 All three experiments were divided into two main phases. The first phase aimed at defining the
168 participant's threshold during a perceptual task using a 1-up/2-down staircase procedure (Levitt, 1971).
169 In Experiment 1, participants indicated which of two stimuli presented to the right or left ear (auditory
170 condition), wrist (tactile condition), or visual field (visual condition) was the most salient. Saliency
171 corresponded respectively to auditory loudness, tactile force, and visual contrast (see below for details).
172 In Experiment 2, participants indicated whether the two most salient stimuli among two simultaneous
173 pairs were presented to the same or different ear (auditory condition), visual field (visual condition), or
174 whether the side of the most salient auditory stimulus corresponded to the side of the most salient
175 visual one (audiovisual condition). Stimuli were presented simultaneously for 250 ms. All staircases
176 included a total of 80 trials and lasted approximately 5 min. All thresholds were defined as the average
177 stimulus intensity during the last 25 trials of the staircase procedure. All staircases were visually
178 inspected, and restarted in case no convergence occurred by the end of the 80 trials (*i.e.*, succession of
179 multiple up/down reversals). The initial stimulation parameters in the audiovisual condition of
180 Experiments 2 and 3 were determined by a unimodal staircase procedure, applied successively to the
181 auditory and visual condition.

182 In the second phase, participants did the same perceptual task, with an initial stimulus intensity given
183 by the final value in the staircase conducted in phase 1. As in phase 1, stimuli in phase 2 were
184 controlled with 1-up/2-down staircase procedure to keep task performance around 71% throughout,
185 thus accounting for training or fatigue effects. This ensured a constant level of task difficulty, which is
186 crucial to precisely quantify metacognitive accuracy across conditions, and is a standard approach (e.g.,
187 Fleming et al, 2010, McCurdy et al 2013; Ais et al., 2016). Immediately after providing their response
188 on the perceptual task, participants reported their confidence on their preceding response on a visual
189 analog scale using a mouse with their right hand. The left and right end of the scale were labeled “Very
190 unsure” and “Very sure” respectively, and participants were asked to report their confidence as
191 precisely as possible, trying to use the whole scale range. A cursor slid over the analog scale
192 automatically following mouse movements, and participants clicked the left mouse button to indicate
193 their confidence. Participants could click the right-button instead to indicate when they had made a
194 trivial mistake (e.g., pressed the wrong button, obvious lapses of attention), which allowed us to
195 exclude these trials from the analysis. During a training phase of 10 trials, the cursor changed color
196 after participants clicked to provide their answer to the perceptual task. The cursor turned green
197 following correct responses, and red following incorrect responses. No feedback was provided after the
198 training phase. In the audiovisual condition of Experiments 2 and 3, auditory and visual stimuli
199 intensities were yoked, so that a correct (incorrect) answer on the bimodal stimulus led to an increase
200 (decrease) in the stimulus intensity in both modalities. Each condition included a total of 400 trials,
201 divided into 5 blocks. Trials were interspaced with a random interval lasting between 0.5 and 1.5 s
202 drawn from a uniform distribution. The three conditions (two in Experiment 3) were run successively
203 in a counterbalanced order. One entire experimental session lasted approximately 3 hours.

204 *Behavioural analysis*

205 The first 50 trials of each condition were excluded from analysis as they contained large variations of
206 perceptual signal. Only trials with reaction times below 3 s for the type 1 task and type 2 task were kept
207 (corresponding to an exclusion of 22.2% of trials in Experiment 1 and 12.6% in Experiment 2). In
208 Experiment 3, we used a more lenient superior cutoff of 5 s, resulting in 3.7 % excluded trials, as many
209 trials had to be removed due to artifacts in the EEG signal. Meta-d' (Maniscalco and Lau, 2012) was
210 computed with Matlab (Mathworks; RRID:SCR_001622), with confidence binned into 6 quantiles per
211 participant and per condition. All other behavioural analyses were performed with R (2016;
212 RRID:SCR_001905), using notably type 3 analyses of variance with Greenhouse-Geisser correction
213 (afex package: Singmann, Bolker, and Westfall, 2015), and null effect estimates using Bayes factors
214 with a Cauchy prior of medium width (scale = 0.71; BayesFactor package: Morey, Rouder, and Jamil,
215 2015). Correlations in metacognitive efficiencies across senses were quantified by R^2 , adjusted for the
216 number of dependent variables relative to the number of data points. The overlap between confidence
217 and reaction times probability density functions after correct and incorrect responses was estimated as
218 the area defined by the x-axis and the lower of the two densities at each point in x (Overlap package:
219 Meredith and Ridout, 2016). The package ggplot2 (Wickham, 2009; RRID:SCR_014601) was used for
220 graphical representations.

221 *Preprocessing of EEG data*

222 Continuous EEG was acquired at 1024 Hz with a 64-channels Biosemi ActiveTwo system referenced
223 to the common mode sense-driven right leg ground (CMS-DRL). Signal preprocessing was performed
224 using custom Matlab (Mathworks; RRID:SCR_001622) scripts using functions from the EEGLAB (v
225 13.5.4, Delorme and Makeig, 2004; RRID:SCR_007292), Adjust (Mognon, Jovicich, Bruzzone, and
226 Buiatti, 2011; RRID:SCR_009526) and Sasica toolboxes (Chaumon, Bishop, and Busch, 2015). The
227 signal was first down-sampled to 512 Hz and band-pass filtered between 1 and 45 Hz (Hamming

228 windowed-sinc finite impulse response filter). Following visual inspection, artifact-contaminated
229 electrodes were removed for each participant, corresponding to 3.4% of total data. Epoching was
230 performed at type 1 response onset. For each epoch, the signal from each electrode was centered to
231 zero and average-referenced. Following visual inspection and rejection of epochs containing artifactual
232 signal (3.9% of total data, SD = 2.2%), independent component analysis (Makeig, Bell., Jung, and
233 Sejnowski, 1996) was applied to individual data sets, followed by a semi-automatic detection of
234 artifactual components based on measures of autocorrelation, correlation with vertical and horizontal
235 EOG electrodes, focal channel topography, and generic discontinuity (Chaumon et al., 2015).
236 Automatic detection was validated by visually inspecting the first 15 component scalp map and power
237 spectra. After artifacts rejection, epochs with amplitude changes of ± 100 μ V DC-offset were excluded
238 (2.9 % of epochs, SD = 3.1%), and the artifact-contaminated electrodes were interpolated using
239 spherical splines (Perrin, Pernier, Bertrand, & Echallier, 1989).

240 *Statistical analyses of EEG data*

241 Following preprocessing, analyses were performed using custom Matlab scripts using functions from
242 the EEGLAB (Delorme and Makeig, 2004; RRID:SCR_007292) and Fieldtrip toolboxes (Oostenveld et
243 al., 2011; RRID:SCR_004849). Event-related potentials were centered on zero. Time-frequency
244 analysis was performed using Morlet wavelets (3 cycles) focusing on the 8-12 Hz band. Voltage
245 amplitude and alpha power were averaged within 50 ms time windows, and analyzed with linear mixed
246 effects models (lme4 and lmerTest packages: Bates et al., 2014; Kuznetsova et al., 2014). This method
247 allowed analyzing single trial data, with no averaging across condition or participants, and no
248 discretization of confidence ratings (Bagiella, Sloan, & Heitjan, 2000). Models were performed on
249 each latency and electrode for individual trials, including raw confidence rating and condition (i.e.,
250 visual vs. audiovisual) as fixed effects, and random intercepts for subjects. Random slopes could not be

251 included in the models as they induced convergence failures (i.e., we used parsimonious instead of
 252 maximal models, see Bates et al., 2015). Significance of fixed effects was estimated using
 253 Satterthwaite's approximation for degrees of freedom of F statistics. Statistical significance for ERPs
 254 and alpha power within the region of interest was assessed after correction for false-discovery rate.
 255 Topographic analyses were exploratory, and significance was considered for $p < 0.001$ without
 256 correcting for multiple comparisons.

257

258 *Signal-detection theory (SDT) models of behavior*

259 The models assume that on each trial two internal signals are generated, $\{X_1, X_2\}$ and then combined
 260 into a bivariate normal. Since X_1 and X_2 are independent, the covariance matrix is diagonal. The
 261 marginal distributions of the bivariate normal corresponded to one of the stimuli pairs in each
 262 condition. Each pair can be described as *R* (or *L*) if the strongest stimulus in the pair is the right (or left)
 263 one. The bivariate distribution was parametrically defined with an arbitrary mean with $|\mu| = (1,1)$ ($\mu = 1$
 264 in cases of *R* stimuli and $\mu = -1$ in cases of *L* stimuli) and two standard deviations σ_1, σ_2 . Thus, the four
 265 probability densities can be expressed as a function of the internal signal strength X and its distribution
 266 parameters μ and σ .

$$P(X_1|L) = \frac{1}{2\pi\sigma_1} e^{\left(\frac{-(X_1+\mu_1)^2}{2\sigma_1^2}\right)}; \quad P(X_1|R) = \frac{1}{2\pi\sigma_1} e^{\left(\frac{-(X_1-\mu_1)^2}{2\sigma_1^2}\right)}$$

$$P(X_2|L) = \frac{1}{2\pi\sigma_2} e^{\left(\frac{-(X_2+\mu_2)^2}{2\sigma_2^2}\right)}; \quad P(X_2|R) = \frac{1}{2\pi\sigma_2} e^{\left(\frac{-(X_2-\mu_2)^2}{2\sigma_2^2}\right)}$$

267 For each set of four stimuli presented in every trial of experiment 2, congruent pairs correspond to
 268 either *LL* or *RR* stimuli, whereas incongruent correspond to *LR* or *RL* stimuli.

269

270 *Decision rule - type-1 task*

271 In the model, the type-1 congruency decision depends on the log-likelihood ratio:

$$272 \quad d = \log \left(\frac{P(\text{congruent})}{P(\text{incongruent})} \right) = \log \left(\frac{P(LL \text{ or } RR)}{P(LR \text{ or } RL)} \right) = \log \left(\frac{P(LL|X_1, X_2) + P(RR|X_1, X_2)}{P(LR|X_1, X_2) + P(RL|X_1, X_2)} \right)$$

273 Applying Bayes' rule and given that X_1, X_2 are independent:

$$d = \log \left(\frac{P(X_1|L) \cdot P(X_2|L) \cdot P^2(L) + P(X_1|R) \cdot P(X_2|R) \cdot P^2(R)}{P(X_1|L) \cdot P(X_2|R) \cdot P(L) \cdot P(R) + P(X_1|R) \cdot P(X_2|L) \cdot P(R) \cdot P(L)} \right)$$

274 And assuming equal priors $P(R) = P(L)$:

$$d = \log \left(\frac{P(X_1|L) \cdot P(X_2|L) + P(X_1|R) \cdot P(X_2|R)}{P(X_1|L) \cdot P(X_2|R) + P(X_1|R) \cdot P(X_2|L)} \right)$$

275 If $d > 0$ the response given is “congruent”, whereas if $d < 0$ the response is “incongruent”. The values

276 of (X_1, X_2) corresponding to $d = 0$, where the congruent and incongruent stimuli are equally likely,

277 should satisfy the relation:

$$278 \quad P(X_1|L) \cdot P(X_2|L) + P(X_1|R) \cdot P(X_2|R) = P(X_1|L) \cdot P(X_2|R) + P(X_1|R) \cdot P(X_2|L)$$

279 The solution to this relation should then satisfy:

$$P(X_1|L) \cdot [P(X_2|L) - P(X_2|R)] = P(X_1|R) \cdot [P(X_2|L) - P(X_2|R)]$$

280 And, trivially for an ideal observer, possible solutions for the type-1 decision are given by:

$$281 \quad \{X_1 = 0, X_2 \in \mathfrak{R}\} \text{ and } \{X_2 = 0, X_1 \in \mathfrak{R}\}$$

282 Therefore, the internal response space (X_1, X_2) is divided in four quadrants such that an ideal observer

283 will respond “congruent” if X_1 and X_2 are both greater than zero or both lower than zero. If X_1 and X_2

284 have different signs, the response will be incongruent.

285

286 *Confidence judgment - type 2 task*

287 All models assume that confidence in each trial is proportional to the likelihood of having given a
 288 correct answer:

$$289 \text{ conf} \propto P(\text{correct} | \text{decision}, X_1, X_2)$$

290 If a response is “congruent”, a participant’s confidence in that response is then:

$$\text{Conf}(X_1, X_2) \propto P(RR|X_1, X_2) + P(LL|X_1, X_2)$$

291 The values of confidence in this case correspond to the top-right and bottom-left quadrants in the 2-
 292 dimensional SDT model. The two remaining quadrants correspond to trials where the response was
 293 “incongruent” and are symmetrical to the former, relative to the decision axes.

294 Again applying Bayes’ rule, and assuming that confidence in the unimodal condition is calculated on
 295 the basis of the joint distribution and hence $P(X_1, X_2|RR) = P(X_1|R) \cdot P(X_2|R)$, it follows that:

$$\text{Conf}(X_1, X_2) \propto \frac{P(X_1|R) \cdot P(X_2|R) \cdot P^2(R) + P(X_1|L) \cdot P(X_2|L) \cdot P^2(L)}{P(X_1, X_2)}$$

296 Assuming equal priors $P(L) = P(R)$ and given that $P(X_i) = P(X_i|R) + P(X_i|L)$ and $P(X_2) = P(X_2|R) +$
 297 $P(X_2|L)$ the expression above can be rewritten as:

$$\text{conf} \propto \frac{1}{1 + \frac{P(X_1|R)}{P(X_1|L)} + \frac{P(X_2|R)}{P(X_2|L)} + \frac{P(X_1|R) \cdot P(X_2|R)}{P(X_1|L) \cdot P(X_2|L)}} + \frac{1}{1 + \frac{P(X_1|L)}{P(X_1|R)} + \frac{P(X_2|L)}{P(X_2|R)} + \frac{P(X_1|L) \cdot P(X_2|L)}{P(X_1|R) \cdot P(X_2|R)}};$$

298 Assuming bivariate normal distributions of the internal signals (as detailed above) and after
 299 simplification it can be shown that:

$$\text{conf} \propto \frac{1}{1 + e^{d_1} + e^{d_2} + e^{d_1+d_2}} + \frac{1}{1 + e^{-d_1} + e^{-d_2} + e^{-d_1-d_2}};$$

$$d_1 = \left| \frac{2 \cdot X_1 \cdot \mu_1}{\sigma_1^2} \right|; d_2 = \left| \frac{2 \cdot X_2 \cdot \mu_2}{\sigma_2^2} \right|$$

300

301 *Modeling strategy*

302 The modeling included two phases. In the first phase, we obtained the parameter values that best
303 explained the unimodal data. In the second phase, behavioural data in the bimodal condition were
304 predicted by combining the parameter values obtained in phase 1 according to different models. The
305 predictions of these models were compared using Bayes Information Criterion (BIC) and relative BIC
306 weights (see below).

307

308 *Phase 1 - fits to the unimodal conditions*

309 The behavioral data for each participant were summarized in 8 different categories: those trials in
310 which confidence was higher/lower than the median confidence value for each participant, for
311 correct/incorrect type 1 response, for congruent/incongruent stimuli (*i.e.*, 2 confidence bins x 2
312 accuracies x 2 conditions). We summarize these data in the vector containing the number of trials for
313 each category n_{obs} . In the context of SDT, two parameters are enough to fully determine the expected
314 probability densities p_{exp} of these 8 response types: the internal noise (σ) and confidence criterion (c).
315 We defined the best fitting model parameters as those that maximized the likelihood of the observed
316 data. More specifically, we randomly sampled the parameter space using a simulated annealing
317 procedure (using the custom function *anneal*, that implements the method presented by Kirkpatrick et
318 al (1983)) and used maximum likelihood to obtain two parameter values. The best fit (the point of
319 maximum likelihood) was defined as the set of values for p_{exp} that minimized the negative log-
320 likelihood nL of the data:

$$nL = -\log(P(n_{obs}|N, p_{exp}))$$

321 Where P is the multinomial distribution with parameters $n_{obs} = (n^1, \dots, n^8)$, $N = \Sigma(n^1, \dots, n^8)$ and $P_{exp} =$
 322 (P^1, \dots, P^8) , with the superindices corresponding to each of the 8 possible categories:

$$P(n_{obs}|N, P_{exp}) = \begin{cases} \frac{N!}{n_{obs}^1! \cdot \dots \cdot n_{obs}^8!} \cdot p_{exp}^1 \cdot \dots \cdot p_{exp}^8, & \text{when } \sum_{i=1}^8 n_{obs}^i = N \\ 0 & \text{otherwise} \end{cases}$$

323 In the unimodal conditions, σ_1 and σ_2 correspond to each of the stimuli pairs of the same modality and
 324 were therefore constrained to be equal. The parameter c determined the type-2 criterion above which a
 325 decision was associated with high confidence ratings.

326

327 The model relied on three assumptions: first, it assumed equal priors for all possible stimuli. Second,
 328 type-1 decisions were assumed to be unbiased and optimal. Third, as noted above, confidence was
 329 defined as proportional to the likelihood of having given a correct answer, given the type-1 decision
 330 and the internal signal for each stimuli pair. We argue that the second assumption of equality for σ_1 and
 331 σ_2 is a reasonable one in the unimodal visual case, where the two stimuli pairs differed only on their
 332 vertical position (but did not differ in their distance from the vertical midline). This assumption
 333 however is less clearly valid in the unimodal auditory condition, where the two pairs of stimuli were
 334 different (a sinewave ‘beep’ vs. a sawtooth ‘buzz’). To estimate the model fits in the unimodal
 335 condition, R^2 values for the correlation between observed and modeled response rates pooled for all
 336 participants were obtained. Notably, the model was flexible enough to fit the different behavioral
 337 patterns of most participants, and the model fits obtained for the unimodal auditory condition were
 338 comparable to those in the unimodal visual condition (see Results).

339

340 *Phase 2 - predictions of the bimodal condition*

341 Once the σ and c parameters were estimated from the unimodal data for each participant, they were
 342 then combined under different models to estimate the predictions of the data in the audiovisual
 343 condition. Note that with this procedure, and unlike the fits to the unimodal conditions, the data used to
 344 estimate the model parameters were different from those on which the model fits were compared. In the
 345 bimodal condition, and in contrast to the unimodal ones, σ_1 and σ_2 corresponded to the internal noise
 346 for the visual and auditory signal respectively, and were allowed to vary independently. Here X_1, X_2 are
 347 the internal responses generated by each pair of stimuli of the visual and auditory modality respectively.
 348 Because confidence was binned into ‘High’ and ‘Low’ based on individual median splits, the criterion
 349 value was a critical factor determining model fits. Models were grouped into three families to compare
 350 them systematically. The family of *integrative models* echoes the single-modality model and represents
 351 the highest degree of integration: here, confidence is computed on the basis of the joint distribution of
 352 the auditory and visual modalities (Figure 4a). Within this family, the *average model* considers one
 353 value of σ for each modality and takes a criterion resulting from the mean of the two modalities
 354 estimated. The derivation and expression of confidence in the integrative models is equal to that of the
 355 unimodal model, described in detail above.

356 The family of *comparative models* (Figure 4b) assumes that confidence can only be computed
 357 separately for each modality and combined into a single summary measure in a second step. Within this
 358 family, the *minimum-confidence model* takes the minimum of the two independent confidence
 359 estimates as a summary statistic. Following a very similar derivation as for the integrative models, here
 360 confidence can be expressed as:

$$361 \text{ conf} \propto P(\text{correct} | X_1, X_2) \propto \min(P(R|X_1), P(R|X_2)) = \min\left(\frac{1}{1+e^{-d_1}}, \frac{1}{1+e^{-d_2}}\right)$$

362 Finally, the family of *single-modality models* (Figure 4c), assumes that confidence varies with the
 363 internal signal strength of a single modality and therefore supposes no integration of information at the
 364 second-order level. Within this family, the *maximum efficiency model* computes confidence on the
 365 basis of the modality with the best metacognitive efficiency alone.

$$366 \quad \text{conf} \propto P(\text{correct} | X_1) \propto P(R | X_1) = \frac{1}{1 + e^{-d_1}}$$

367 where modality 1 had the best metacognitive efficiency for this participant.

368

369 *Model fits*

370 Single-modality models were assessed by calculating the percentage of variance explained for the data
 371 from the unimodal conditions. First, the *nlme* package in R (Pinheiro and Bates, 2010) was used to
 372 estimate the predictive power of the models while allowing for random intercepts for each participant.
 373 Then, goodness-of-fit was estimated with R^2 using the *piecewiseSEM* package (Lefcheck, 2016).
 374 Bayesian information criterion (BIC) values were then calculated to compare the different models
 375 while accounting for differences in their number of parameters. BIC weights for the model fits to the
 376 bimodal condition were estimated following Burnham and Anderson (2002) and as in Solovey et al.
 377 (2015). By definition, the BIC weight for model i can be expressed as:

$$BIC_w(\text{model } i) = \frac{e^{-\frac{1}{2}(BIC_i - BIC_{min})}}{\sum_{n=1}^3 e^{-\frac{1}{2}(BIC_n - BIC_{min})}}$$

378 where BIC_k is the BIC for model k and BIC_{min} is the lowest BIC corresponding to the best model out of
 379 those considered.

380

381 **Results**

382 *Experiment 1*

383 We first aimed at comparing metacognitive performance across the visual, auditory, and tactile
384 modalities. Participants were presented with a pair of simultaneous stimuli at a right and left location,
385 and asked to indicate which of the two stimuli had the highest intensity (Figure 1, right panel). In this
386 way, the first-order task consisted in a 2-alternative forced choice on visual, auditory, or tactile
387 intensity (i.e., respectively contrast, loudness, or force). After each choice, participants reported their
388 confidence on their previous response (second-order task) (Figure 1, left panel). The main goal of this
389 experiment was to test the hypothesis that metacognitive efficiency would correlate positively between
390 sensory modalities, suggesting a common underlying mechanism. We first report general results of
391 type-1 and type-2 performances and then turn to the central question of correlations between sensory
392 modalities.

393 We aimed to equate first-order performance in the three modalities using a 1-up/2-down staircase
394 procedure (Levitt, 1971). Although this approach prevented large inter-individual variations, some
395 small but significant differences in d' (i.e., first-order sensitivity) across modalities subsisted, as
396 revealed by a one-way ANOVA [$F(1.92,26.90) = 8.76, p < 0.001, \eta_p^2 = 0.38$] (Figure 2a). First-order
397 sensitivity was lower in the auditory condition [mean $d' = 1.20 \pm 0.05$ (95% CI)] as compared to the
398 tactile [mean $d' = 1.37 \pm 0.07, p = 0.002$] and visual conditions [mean $d' = 1.33 \pm 0.07, p = 0.004$]
399 (Figure 2a). No effect of condition on response criterion was found [$F(1.96,27.47) = 0.30, p = 0.74, \eta_p^2$
400 $= 0.02$]. The difference in first-order sensitivity is likely due to the difficulty of setting perceptual
401 thresholds with adaptive staircase procedures. Importantly however, it did not prevent us from
402 comparing metacognitive performance across senses, as the metrics of metacognitive performance we
403 used are independent of first-order sensitivity. As reported previously (Ais et al., 2016), average
404 confidence ratings correlated between the auditory and visual conditions [adjusted $R^2 = 0.26, p = 0.03$],

405 between the tactile and visual conditions [adjusted $R^2 = 0.55$, $p = 0.001$], and between the auditory and
406 tactile conditions [adjusted $R^2 = 0.51$, $p = 0.002$]. Metacognitive sensitivity was estimated with meta- d' ,
407 a response-bias free measure of how well confidence estimates track performance on the first-order
408 task (Maniscalco and Lau, 2012). A one-way ANOVA on meta- d' revealed a main effect of condition
409 [F(1.93,25.60) = 5.92, $p = 0.009$, $\eta_p^2 = 0.30$] (Figure 2b). To further explore this main effect and rule
410 out the possibility that it stemmed from differences at the first-order level, we normalized
411 metacognitive sensitivity by first-order sensitivity (i.e., meta- d'/d'), to obtain a pure index of
412 metacognitive performance called metacognitive efficiency. Only a trend for a main effect of condition
413 was found [F(1.76,24.61) = 3.16, $p = 0.07$, $\eta_p^2 = 0.18$] (Figure 2c), revealing higher metacognitive
414 efficiency in the visual [mean ratio = 0.78 ± 0.13] vs. auditory domain [mean meta- d'/d' ratio = $0.61 \pm$
415 0.15 ; paired t-test: $p = 0.049$]. The difference in metacognitive efficiency between the visual and the
416 tactile conditions [mean ratio = 0.70 ± 0.10] did not reach significance [paired t-test: $p = 0.16$, Bayes
417 Factor = 0.65].

418 We then turned to our main experimental question. We found positive correlations between
419 metacognitive efficiency in the visual and tactile conditions [adjusted $R^2 = 0.21$, $p = 0.047$] (Figure 2e),
420 and in the auditory and tactile conditions [adjusted $R^2 = 0.24$, $p = 0.038$] (Figure 2f). (The data were
421 inconclusive regarding the correlation between the visual and auditory condition [adjusted $R^2 = 0.07$, p
422 = 0.17, Bayes Factor = 0.86] (Figure 2d). These results reveal shared variance between auditory, tactile,
423 and visual metacognition, in line with the supramodality hypothesis. Moreover, the absence of any
424 correlation between first-order sensitivity and metacognitive efficiency in any of the conditions [all
425 adjusted $R^2 < 0$; all p -values > 0.19], rules out the possibility that such supramodality during the
426 second-order task was confounded with first-order performance. Finally, no effect of condition on type

427 1 reaction times [$F(1.78,24.96) = 0.28, p = 0.73, \eta^2 p = 0.02$] or type 2 reaction times [$F(1.77,24.84) =$
428 $1.77, p = 0.39, \eta^2 p = 0.06$] was found.

429

430 [Figure 2 Here]

431 *Experiment 2*

432 Experiment 1 revealed correlational evidence for the supramodality of perceptual metacognition across
433 three modalities. A previous study (McCurdy et al., 2013), however, dissociated brain activity related
434 to metacognitive accuracy in vision versus memory, despite clear correlations at the behavioural level.
435 Thus, correlations between modalities are compelling, but not sufficient to support the supramodality
436 hypothesis. We therefore put the evidence of experiment 1 to a stricter test in Experiment 2, by
437 comparing metacognitive efficiency for unimodal vs. bimodal, audiovisual stimuli. We reasoned that if
438 metacognitive monitoring operates independently from the nature of sensory signals from which
439 confidence is inferred, confidence estimates should be as accurate when made on unimodal or bimodal
440 signals. In contrast, if metacognition operated separately in each sensory modality, one would expect
441 that metacognitive efficiency for bimodal stimuli would only be as high as the minimal metacognitive
442 efficiency for unimodal stimuli. Beyond these comparisons, the supramodality hypothesis also implies
443 the existence of correlations between unimodal and bimodal metacognitive efficiencies. Participants
444 performed three different perceptual tasks, all consisting in a congruency judgment between two pairs
445 of stimuli (Figure 1, right panel). In the unimodal visual condition, participants indicated whether the
446 Gabor patches with the strongest contrast within each pair were situated on the same or different side of
447 the screen. In the unimodal auditory condition, they indicated whether the loudest sounds of each pair
448 were played in the same ear or in two different ears. In the bimodal audiovisual condition, participants
449 indicated whether the side corresponding to the most contrasted Gabor patch of the visual pair

450 corresponded with the side of the loudest sound of the auditory pair. Importantly, congruency
451 judgments required that participants responded on the basis of the two presented modalities. The
452 staircase procedure minimized variations in first-order sensitivity, such that sensitivity in the auditory
453 [mean $d' = 1.31 \pm 0.12$], audiovisual [mean $d' = 1.38 \pm 0.12$], and visual conditions [mean $d' = 1.25 \pm$
454 0.11] were similar (Figure 3a, $F(1.75,22.80) = 2.12$, $p = 0.15$, $\eta_p^2 = 0.14$). No evidence of multisensory
455 integration was found at the first-order level, as the perceptual thresholds determined by the staircase
456 procedure were not lower in the bimodal vs. unimodal condition [$p = 0.17$]. This is likely due to the
457 task at hand involving a congruency judgment. As in Experiment 1, no effect of condition on response
458 criterion was found [$F(1.87,24.27) = 2.12$, $p = 0.14$, $\eta_p^2 = 0.14$]. No effect of condition on average
459 confidence was found [$F(1.76,24.64) = 0.91$, $p = 0.40$, $\eta_p^2 = 0.06$], and average confidence ratings
460 correlated between the auditory and audiovisual conditions [adjusted $R^2 = 0.56$, $p = 0.001$], between the
461 visual and audiovisual conditions [adjusted $R^2 = 0.38$, $p = 0.01$], and a trend was found between the
462 auditory and visual conditions [adjusted $R^2 = 0.12$, $p = 0.11$]. A significant main effect of condition on
463 type 1 reaction times [$F(1.66,21.53) = 18.05$, $p < 0.001$, $\eta_p^2 = 0.58$] revealed faster responses in the
464 visual [$1.30 \text{ s} \pm 0.10 \text{ s}$] compared to the auditory [$1.47 \text{ s} \pm 0.13 \text{ s}$] and audiovisual task [$1.68 \text{ s} \pm 0.11 \text{ s}$].
465 No difference was found for type 2 reaction times [$F(1.82,23.62) = 1.69$, $p = 0.21$, $\eta_p^2 = 0.11$]. A main
466 effect of condition for both metacognitive sensitivity [meta- d' : $F(1.98,25.79) = 4.67$, $p = 0.02$, $\eta_p^2 =$
467 0.26], and metacognitive efficiency [ratio meta- d'/d' : $F(1.95,25.40) = 6.63$, $p = 0.005$, $\eta_p^2 = 0.34$]
468 (Figure 3b and 3c, respectively). Pairwise comparisons revealed higher metacognitive efficiency in the
469 visual [mean ratio = 0.94 ± 0.19] vs. auditory [mean meta- d'/d' ratio = 0.65 ± 0.17 ; paired t-test: $p =$
470 0.005] and audiovisual domains [mean meta- d'/d' ratio = 0.70 ± 0.15 ; paired t-test: $p = 0.02$]. As
471 auditory and audiovisual metacognitive efficiencies were not different [$p = 0.5$, Bayes Factor = 0.38],
472 the differences in metacognitive efficiency are likely to stem from differences between auditory and
473 visual metacognition, as found in Experiment 1. Thus, the fact that metacognitive efficiency is similar

474 in the audiovisual and auditory tasks implies that the resolution of confidence estimates in the bimodal
475 condition is as good as that in the more difficult unimodal condition (in this case, auditory), despite it
476 requiring the analysis of two sources of information.

477 Crucially, we found correlations between metacognitive efficiency in the auditory and visual conditions
478 [adjusted $R^2 = 0.24$, $p = 0.043$] (Figure 3d); more importantly, we also found correlations between
479 metacognitive efficiency in the auditory and audiovisual conditions [adjusted $R^2 = 0.23$, $p = 0.046$]
480 (Figure 3e) and a trend between metacognitive efficiency in the visual and audiovisual conditions
481 [adjusted $R^2 = 0.15$, $p = 0.097$] (Figure 3f). This contrasted with no correlations between first-order
482 sensitivity and metacognitive efficiency in any of the conditions [all $R^2 < 0.06$; all $p > 0.19$] except in
483 the visual condition, where high d' was predictive of low meta- d'/d' values [$R^2 = 0.39$, $p = 0.01$]. The
484 absence of such correlations in most conditions makes it unlikely that relations in metacognitive
485 efficiency were driven by similarities in terms of first-order performance. In addition to the equivalence
486 between the resolution of unimodal and bimodal confidence estimates, the correlations in
487 metacognitive efficiency between unimodal and bimodal conditions suggest that metacognitive
488 monitoring for unimodal vs. bimodal signals involves shared mechanisms (i.e. supramodality).

489 [Figure 3 Here]

490 *Computational models of confidence estimates for bimodal signals*

491 Using the data from experiment 2, we next sought to reveal potential mechanisms underlying the
492 computation of confidence in the bimodal condition. For this, we first modeled the proportion of trials
493 corresponding to high vs. low confidence in correct vs. incorrect type 1 responses, in the unimodal
494 auditory and unimodal visual conditions separately. Each condition was represented by a 2-dimensional
495 signal detection theory (SDT) model with standard assumptions and only 2 free parameters per
496 participant, namely internal noise σ and confidence criterion c (see Figure 4 and Methods). This simple

497 model accounted for more than half the total variance in participants' proportion of responses both in
498 the unimodal visual [$R^2 = 0.68$] and unimodal auditory conditions [$R^2 = 0.57$]. We then combined the
499 fitted parameter values under different rules to estimate and compare their fits to the audiovisual data.
500 All models assume that the visual and auditory stimuli did not interact, which is supported by the fact
501 that perceptual thresholds determined by the staircase procedure were not lower in the bimodal vs.
502 unimodal conditions (see SI). Note that with this procedure, and unlike the fits to the unimodal
503 conditions, the data used to estimate the model parameters were different from those on which the
504 model fits were compared. We evaluated different models systematically by grouping them into three
505 families, varying in degree of supramodality. We present here the best model from each family (figure
506 4), and all computed models in SI. The *integrative model* echoes the unimodal models and represents
507 the highest degree of integration: here, confidence is computed on the basis of the joint distribution of
508 the auditory and visual modalities. The *comparative model* assumes that confidence is computed
509 separately for each modality and in a second step combined into a single summary measure (in
510 particular, the minimum of the two estimates, see Methods for other measures). The *single-modality*
511 *model* assumes that confidence varies with the internal signal strength of a single modality and
512 therefore supposes no integration of information at the second-order level. We compared these different
513 models by calculating their respective BIC weights (BICw: Burnham and Anderson, 2002; Solovey et
514 al., 2015), which quantify the relative evidence in favor of a model in relation to all other models
515 considered.

516 By examining individual BICw in a ternary plot (Figure 4d), we found that the best model for most
517 participants was either the *integrative* or the *comparative model*, whereas the BICw for the *single-*
518 *domain model* was equal to 0. Yet, we note that the *single-modality* model is also plausible, as it does
519 predict the responses of four participants better than any of the other two models. The reason why our

520 models could not clearly distinguish between the integrative model and the comparative model may be
521 due to the fact that differences in intensity between the left and right stimuli of the auditory and visual
522 pairs were yoked: the staircase procedure we used controlled both pairs simultaneously, increasing
523 (decreasing) the difference between the left and right stimuli in both modalities after an incorrect (two
524 correct) response. As a result, we sampled values from a single diagonal in the space of stimulus
525 intensities, which limits the modeling results. In future studies, non-yoked stimuli pairs could be used
526 —albeit at the cost of a longer experimental session— to explore wider sections of the landscape of
527 confidence as a function of internal signal to better test the likelihood of the models studied here.

528 Taken together these computational results suggest that most participants computed confidence in the
529 bimodal task by using information from the two modalities under a supramodal format that is
530 independent of the sensory modality, in agreement with the first mechanism for supramodal
531 metacognition we introduced. We conclude that the confidence reports for audiovisual signals arise
532 either from the joint distribution of the auditory and visual signals (*integrative model*), or are computed
533 separately for distinct modalities, and then combined into a single supramodal summary statistic
534 (*comparative model*). These two models indicate that metacognition may be supramodal because
535 monitoring operates on supramodal confidence estimates, computed with an identical format or neural
536 code across different tasks or sensory modalities. We later refer to this as the first mechanism for
537 supramodal metacognition. In addition, metacognition may be supramodal in case a non-perceptual
538 signal drives the computation of confidence estimates (mechanism 2). Among them, likely candidates
539 are decisional cues such as reaction times during the first-order task, as they are present no matter the
540 sensory modality at play, and are thought to play an important role for confidence estimates (Yeung
541 and Summerfield, 2012). We next sought to assess if metacognition was supramodal due to the
542 influence of decisional cues that are shared between sensory modalities (mechanism 2).

543

544 [Figure 4 Here]

545

546 Our modeling results suggest that confidence estimates are encoded in a supramodal format,
547 compatible with the supramodality hypothesis for metacognition. Notably however, apparent
548 supramodality in metacognition could arise in case non-perceptual signals are taken as inputs for the
549 computation of confidence. In models implying a decisional locus for metacognition (Yeung and
550 Summerfield, 2012), stimulus-independent cues such as reaction times during the first-order task take
551 part in the computation of confidence estimates. This is empirically supported by a recent study
552 showing that confidence in correct responses is decreased in case response-specific representations
553 encoded in the premotor cortex are disrupted by transcranial magnetic stimulation (Fleming et al.,
554 2015). In the present study, decisional parameters were shared across sensory modalities, since
555 participants used a keyboard with their left hand to perform the first-order task for all tasks. To extend
556 our modeling results and assess whether supramodality in metacognition also involves a decisional
557 locus (mechanism 2 discussed above), we examined how participants used their reaction times to infer
558 confidence in different conditions. Specifically, we quantified the overlap of first-order reaction times
559 distributions corresponding to correct vs. incorrect responses, as a summary statistic representing how
560 reaction times differ between correct and incorrect trials. We measured how reaction time overlap
561 correlated with the overlap of confidence ratings after correct vs. incorrect first-order responses, which
562 is a summary statistic analogous to ROC-based methods typically used to quantify metacognitive
563 sensitivity with discrete confidence scales (Fleming and Lau, 2014). If confidence involves a
564 decisional-locus, one would expect a correlation between confidence overlap and reaction time overlap,
565 so that participants with the smallest confidence overlap (i.e., highest metacognitive sensitivity) are the
566 ones with the smallest reaction times overlap (i.e., distinct reaction times in correct vs. incorrect

567 responses). Interestingly in Experiment 1, the correlation strength mirrored the difference in
568 metacognitive efficiency we found between sensory modalities: higher correlations were found in the
569 visual domain (adjusted $R^2 = 0.54$, $p = 0.002$; average metacognitive efficiency = 0.78 ± 0.13),
570 compared to the tactile (adjusted $R^2 = 0.26$, $p = 0.03$; average metacognitive efficiency = 0.70 ± 0.10)
571 and auditory domains (adjusted $R^2 = -0.06$, $p = 0.70$; average metacognitive efficiency = 0.61 ± 0.15).
572 This suggests that decisional parameters such as reaction times in correct vs. incorrect trials may
573 inform metacognitive monitoring, and may be used differently depending on the sensory modality with
574 a bigger role in visual than in tactile and auditory tasks. These results are in line with second-order
575 models of confidence estimation (Fleming & Daw 2017), and support empirical results showing better
576 metacognitive performance when confidence is reported after vs. before the first-order task (Siedlecka
577 et al., 2016), or better metacognitive performance for informative vs. non-informative action during the
578 first-order task (Kvam et al., 2015). Importantly, although such correlations between reaction time
579 overlap and confidence overlap would be expected in experiments containing a mixture of very easy
580 and very difficult trials, the correlations in the visual and tactile modalities reported above persisted
581 even after the variance of perceptual evidence was taken into account using multiple regressions. This
582 result rules out the possibility that these correlations are explained by variance in task difficulty. This
583 pattern of results was not found in Experiment 2 (i.e. no correlation between reaction times and
584 confidence overlaps; all $R^2 < 0.16$, all $p > 0.1$), but replicated in Experiment 3 as further detailed below.

585

586 *Experiment 3*

587 The aim of experiment 3 was three-fold. First and foremost, we sought for the first time to document
588 the potential common and distinct neural mechanisms underlying unimodal and bimodal metacognition.
589 Following the link between reaction times and metacognitive efficiency uncovered in Experiment 1, we

590 expected to find supramodal neural markers of metacognition preceding the first-order task, as
591 quantified by the amplitude of event-related potentials (ERPs) as well as in alpha suppression over the
592 sensorimotor cortex prior to key press (Pfurtscheller and Lopes Da Silva, 1999). Second, we aimed at
593 replicating the behavioural results from Experiment 2, especially the correlation between visual and
594 audiovisual metacognitive efficiency. Third, we aimed at estimating the correlations between
595 confidence and reaction times overlap on a new group of participants. Therefore, we tested participants
596 on these two conditions only.

597 *Behavioural data*

598 The staircase procedure minimized variations in first-order sensitivity [$t(17) = 0.3$, $p = 0.76$, $d = 0.07$],
599 such that sensitivity in the audiovisual [mean $d' = 1.15 \pm 0.07$] and visual conditions [mean $d' = 1.17 \pm$
600 0.05] were similar. Contrary to what was found in Experiments 1 and 2, response criterion varied
601 across conditions [$t(17) = 4.33$, $p < 0.001$, $d = 0.63$], with a tendency to respond “congruent” more
602 pronounced in the audiovisual [mean criterion = 0.27 ± 0.12] vs. visual condition [mean criterion = -
603 0.02 ± 0.15]. This effect was unexpected but did not preclude from running subsequent analyses
604 dealing with metacognitive sensitivity. We found no effect of condition on average confidence [$t(17) =$
605 0.56 , $p = 0.14$, $d = 0.08$]. Average confidence ratings correlated between the visual and audiovisual
606 conditions [adjusted $R^2 = 0.65$, $p < 0.001$]. No difference in metacognitive sensitivity was found
607 between conditions [$t(17) = 0.78$, $p = 0.44$, $d = 0.09$] or efficiency [$t(17) = 0.78$, $p = 0.44$, $d = 0.08$].
608 Crucially, we replicated our main results from Experiment 2, as we found a positive significant
609 correlation between relative metacognitive accuracy in the audiovisual and visual conditions [adjusted
610 $R^2 = 0.47$, $p < 0.001$], and no correlation between first-order sensitivity and metacognitive efficiency in
611 either condition [both $R^2 < 0.01$; both p -values > 0.3] (Figure 5). Regarding the decisional locus of
612 metacognition, Experiment 3 confirmed the results of Experiment 1: reaction time and confidence

613 overlaps correlated more in the visual condition (adjusted $R^2 = 0.41$, $p = 0.003$), than in the audiovisual
614 condition (adjusted $R^2 = -0.05$, $p = 0.70$), suggesting that decisional parameters such as reaction times
615 may inform metacognitive monitoring, although differently between the visual and audiovisual
616 conditions. Altogether, these behavioral results from three experiments with different subject samples
617 confirm the existence of shared variance in metacognitive efficiency between unimodal and bimodal
618 conditions, and do not support major group differences between them. Further, they support the role of
619 decisional factors such as reaction times estimates, as predicted when considering a decisional locus for
620 metacognition.

621 [Figure 5 Here]

622 *EEG data*

623 Next, we explored the neural bases of visual and audiovisual metacognition, focusing on the decisional
624 locus of confidence by measuring ERPs locked to the type 1 response. This response-locked analysis
625 took into account the differences in type 1 reaction times between the visual and audiovisual tasks (562
626 ms shorter in the visual condition on average: $t(17) = 6.30$, $p < 0.001$). Since we showed that decisional
627 parameters such as reaction times inform metacognitive monitoring, this analysis was carried out on a
628 set of scalp electrodes over the right sensorimotor cortex that included the left hand representation with
629 which participants performed the first-order task (see Boldt and Yeung, 2015 for findings showing that
630 parietal scalp regions also correlate with confidence prior to response). Incorrect type 1 responses were
631 not analyzed as the lower-bound of the confidence scale we used corresponded to a “pure guess”, and
632 therefore did not allow disentangling detected vs. undetected errors. For each trial, we extracted the
633 ERP amplitude time-locked to the onset of correct type 1 responses, averaged within 50 ms time
634 windows. For each time window and each electrode, we assessed how ERP amplitude changed as a
635 function of confidence using linear mixed models with condition as a fixed effect (visual vs.

636 audiovisual) and random intercepts for subjects (see Methods for details). This analysis allowed us to
637 assess where and when ERP amplitudes associated with the type-1 response was predictive of
638 confidence ratings given during the type-2 response. Main effects correspond to similar modulations of
639 ERP amplitudes by confidence in the visual and audiovisual condition (i.e., supramodality hypothesis),
640 while interaction effects correspond to different amplitude modulations in the visual vs. audiovisual
641 conditions. A first main effect of confidence was found early before the type 1 response, underlying a
642 negative relationship between ERP amplitude and confidence (-600 to -550 ms; $p < 0.05$, FDR-
643 corrected, see figure 6a, left panel, showing the grand average between the visual and audiovisual
644 condition). A second main effect of confidence peaked at -300 ms (-400 to -100 ms; $p < 0.05$, FDR-
645 corrected), so that trials with high confidence reached maximal amplitude 300 ms before key press.
646 These two effects are characterized by an inversion of polarity from an early-negative to a late-positive
647 relationship, which has been linked to selective response activation processes (i.e., lateralized readiness
648 potentials, see Eimer and Coles (2003) for review, and Buján et al. (2009) for previous results in
649 metamemory). Thus, the present data show that sensorimotor ERP also contribute to metacognition as
650 they showed a relationship with confidence both in the audiovisual and visual conditions. Of note,
651 confidence modulated the amplitude and not the onset latency of the ERP, which suggests that the
652 timing of response selection itself does not depend on confidence. We complemented this ROI analysis
653 by exploring the relation between confidence and ERP amplitude for all recorded electrodes (figure 6a,
654 right panel). This revealed that the later effect 300 ms before key press was centered on centro-parietal
655 regions (i.e., including our region of interest; $p < 0.001$) as well as more frontal electrodes, potentially
656 in line with several fMRI studies reporting the role of the prefrontal cortex for metacognition (Fleming
657 et al., 2010; Yokoyama et al., 2010; McCurdy et al., 2013, see Grimaldi et al., 2015 for a review). The
658 linear mixed model analysis also revealed significant interactions, indicating that the modulation of
659 ERP amplitude as a function of confidence was significantly stronger in the visual condition, with

660 again one early (-750 to -600 ms) and late component (-350 to -150 ms; Figure 6b, left panel).
661 Topographical analysis of these interactions implicated frontal and parieto-occipital electrodes. These
662 results at the neural level are consistent with our behavioural data, since we found that reaction times
663 have more influence on the computation of confidence in the visual compared to the audiovisual
664 condition.

665 [Figure 6 Here]

666 Complementary to ERP amplitude, we also analyzed oscillatory alpha power (i.e, pre-movement
667 related desynchronization) as a signature of motor preparation (Pfurtscheller and Lopes Da Silva, 1999).
668 Results of the linear mixed model analysis revealed a sustained main effect of confidence starting 300
669 ms before key press and continuing until 200 ms after the type 1 response ($p < 0.05$ FDR-corrected),
670 showing a negative relationship between confidence and alpha power (i.e., alpha suppression, figure 7a,
671 left panel). Note that, opposite to what we found in the amplitude domain, the main effect of
672 confidence on alpha power was found even after a first-order response was provided. Likewise, the
673 topographical analysis revealed a different anatomical localization than the effect we found in the
674 amplitude domain, with more posterior, parieto-occipital electrodes involved. This suggests that alpha
675 suppression prior to type 1 response varies as a function of confidence non-differentially in both the
676 audiovisual and visual conditions. The linear mixed model analysis also revealed a main effect of
677 condition, with higher alpha power in the visual vs. audiovisual condition (figure 7b, left panel). This
678 could be related to the fact that the audiovisual task was judged more demanding by participants, as
679 reflected by their longer type 1 reaction times. Finally, significant interactions between confidence and
680 condition were found, with topographical locations predominantly within frontal electrodes. Taken
681 together, the main effects of confidence on voltage amplitude and alpha power reveal some of the
682 markers validating the supramodality hypothesis at a decisional locus. These are likely to be part of a
683 bigger set of neural mechanisms, operating at a decisional, but also post-decisional locus that was not
684 explored here (Pleskac et al., 2010). The existence of significant interactions reveals that some domain-
685 specific mechanisms are also at play during metacognition, which accounts for the unexplained
686 variance when correlating metacognitive efficiencies across modalities at the behavioral level.

687 [Figure 7 Here]

688

689 **Discussion (1591 words)**

690 Is perceptual metacognition supramodal, with a common mechanism for distinct sensory modalities, or
691 is it modality-specific, with idiosyncratic mechanisms for each sensory modality? As of today, this
692 issue remains unsettled because the vast majority of experiments on metacognitive perception only
693 involved the visual modality (but see Ais et al., 2016; De Gardelle et al., 2016). In vision, Song and
694 colleagues (2011) found that about half of the variance in metacognitive sensitivity during a contrast
695 discrimination task was explained by metacognitive sensitivity in an orientation discrimination task,
696 suggesting some level of generality within vision. Likewise, roughly a quarter of the variance in
697 metacognitive sensitivity during a contrast discrimination task was explained by metacognitive
698 sensitivity during a memory task involving words presented visually (McCurdy et al., 2013). Here, we
699 extend these studies by assessing the generality of metacognition across three sensory modalities as
700 well as conjunctions of two sensory modalities. In Experiment 1 we tested participants in three
701 different conditions, which respectively required discriminating the side on which visual, auditory or
702 tactile stimuli were most salient. We found positive correlations between metacognitive efficiency
703 across sensory modalities, and ruled out the possibility that these correlations stemmed from
704 differences in first-order performances (Maniscalco and Lau, 2012). These results extend previous
705 reports (Ais et al., 2016; De Gardelle et al., 2016) showing similarities between auditory and visual
706 metacognition to auditory, tactile, and visual laterality discrimination tasks, and therefore support the
707 existence of a common mechanism underlying metacognitive judgments in three distinct sensory
708 modalities.

709 In Experiment 2, we further extended these results to a different task and also generalized them to
710 bimodal stimuli (Deroy et al., 2016). First, using a first-order task that required congruency rather than
711 laterality judgments, we found again that metacognitive efficiency for auditory stimuli correlated with

712 metacognitive efficiency for visual stimuli. Second, we designed a new condition in which participants
713 had to perform congruency judgments on bimodal, audiovisual, signals, which required the information
714 from both modalities to be taken into account. Three further observations from these conditions support
715 the notion of supramodality in perceptual metacognition. First, we observed that metacognitive
716 efficiency in the audiovisual condition was indistinguishable from that in the unimodal auditory
717 condition, suggesting that the computation of joint confidence is not only possible but can also occur at
718 no behavioral additional cost. These results confirm and extend those of Experiment 1 in a different
719 task and with different participants, and further suggest that performing confidence estimates during a
720 bimodal task was not more difficult than doing so during the hardest unimodal task (in this case,
721 auditory), despite it requiring the computation of confidence across two perceptual domains. We take
722 this as evidence in support of supramodality in perceptual metacognition. Second, we found a positive
723 and significant correlation in metacognitive efficiency between the auditory and audiovisual conditions,
724 and a trend between the visual and audiovisual conditions, later replicated in Experiment 3. As in
725 Experiment 1, these results cannot be explained by confounding correlations with first-order
726 performance. We take this as another indication that common mechanisms underlie confidence
727 computations for perceptual tasks on unimodal and bimodal stimuli. While the reported correlations
728 involved a rather low number of participants and were arguably sensitive to outliers (McCurdy et al.,
729 2013), we note that they were replicated several times, under different conditions and tasks in different
730 groups of participants, which is likely in less than 1% of cases under the null hypothesis (binomial test).
731 In addition, qualitatively similar correlations were obtained when metacognitive performance was
732 quantified by the area under the type 2 receiving operative curve, and by the slope of a logistic
733 regression between type-1 accuracy and confidence.

734 The next piece of evidence we brought in favor of supramodal metacognition goes beyond correlational
735 evidence, and provides new insights regarding the mechanisms involved in confidence estimates when
736 the signal extends across two sensory modalities. Using a modeling approach, we found that data in the
737 audiovisual condition could be predicted by models that computed confidence with a supramodal
738 format, either based on the joint information from a bimodal audiovisual (integrative model)
739 representation, or on the comparison between unimodal visual and auditory representations
740 (comparative model). Although these two models have distinct properties, they both involve
741 supramodal confidence estimates with identical neural codes across different sensory modalities. Thus,
742 although we could not distinguish which of the two models was most representative of behavioral data
743 at the group level, they both bring evidence in favor of the first mechanism we introduced, according to
744 which metacognition is supramodal because monitoring operates on supramodal confidence estimates.

745 Finally, we assessed in Experiment 3 whether supramodal metacognition could arise due to the second
746 mechanism we introduced, according to which supramodality is driven by the influence of non-
747 perceptual, decisional signals during the computation of confidence estimates. For this purpose, we
748 replicated correlations in metacognitive efficiency between the visual and audiovisual conditions, while
749 examining the neural mechanisms of visual and audiovisual metacognition preceding the perceptual
750 judgment (i.e., at a decisional level). In a response-locked analysis with confidence and condition as
751 within-subject factors, we found that confidence preceding the type 1 response was reflected in ERP
752 amplitude and alpha power (main effect), within a region of interest that included the parietal and
753 sensorimotor cortex corresponding to the hand used for the type 1 task, as well as more frontal sites.
754 Before discussing the main effects of confidence, we note that the analysis also revealed interactions
755 between confidence and condition, revealing that idiosyncratic mechanisms are also at play during the
756 metacognitive monitoring of visual vs. audiovisual signals, and that modulations of ERP and alpha

757 power as a function of confidence were overall greater in the visual vs. audiovisual condition.
758 Regarding the main effects, we found an inversion ERP polarity over left sensorimotor regions,
759 suggesting a link between confidence and selective response activation, so that trials with high
760 confidence in a correct response were associated with stronger motor preparation (Eimer and Coles,
761 2003; Buján et al., 2009). Regarding oscillatory power, we found relative alpha desynchronization in
762 occipito-parietal regions, which has been shown to reflect the level of cortical activity, and is held to
763 correlate with processing enhancement (Pfurtscheller, 1992). At the cognitive level, alpha suppression
764 is thought to instantiate attentional gating, so that distracting information is suppressed (Pfurtscheller
765 and Lopes Da Silva, 1999; Foxe and Snyder, 2011; Klimesch, 2012). Indeed, higher alpha power has
766 been shown in cortical areas responsible for processing potentially distracting information, both in the
767 visual and audiovisual modalities (Foxe et al., 1998). More recently, pre-stimulus alpha power over
768 sensorimotor areas was found to be negatively correlated with confidence (Baumgarten et al., 2016;
769 Samaha et al., 2016), or attentional ratings during tactile discrimination (Whitmarsh et al., 2016).
770 Although these effects are usually observed prior to the onset of an anticipated stimulus, we observed
771 them prior to the type 1 response, suggesting that low confidence in correct responses could be due to
772 the effect of inattention to common properties of first-order task execution such as motor preparation or
773 reaction time (stimulus locked-analyses that are not reported here revealed no effect of confidence prior
774 to stimulus onset). This is compatible with a recent study showing that transcranial magnetic
775 stimulation over the premotor cortex before or after a visual first-order task disrupts subsequent
776 confidence judgments (Fleming et al., 2015).

777 The finding of lower alpha power with confidence in correct responses is compatible with the
778 observation that participants with more distinct reaction times between correct and incorrect responses
779 had better metacognitive efficiency, as revealed by the correlation between confidence and reaction

780 times overlaps following correct vs. incorrect responses. Thus, attention to motor task execution may
781 feed into the computation of confidence estimates, in a way that is independent of the sensory modality
782 involved, thereby providing a potential decisional mechanism for supramodal metacognition. In
783 experiment 1, we also found that confidence and reaction times overlap were more correlated in the
784 visual condition compared to the tactile, auditory, or audiovisual conditions. Based on these results, we
785 speculate that decisional parameters in link with processes related to movement preparation inform
786 metacognitive monitoring. Our EEG results and the correlations between reaction time and confidence
787 overlaps suggest that decisional parameters may have a stronger weight in the visual than in the other
788 modalities, which could explain the relative superiority of visual metacognition over other senses. We
789 argue that this decisional mechanism in metacognition is compatible with the supramodality hypothesis,
790 in addition to the supramodal computation of confidence supported by our behavioral and modeling
791 results. Of note, our analysis focusing on the alpha band to uncover the role of decisional cues on
792 confidence estimates is not exhaustive, and other frequencies might contribute to confidence estimates
793 equally between sensory domains (e.g., theta band, see Wokke et al., 2017).

794 Altogether, our results highlight two non-mutually exclusive mechanisms for the finding of correlated
795 metacognitive efficiencies across auditory, tactile, visual and audiovisual domains. First, our modeling
796 work showed that confidence estimates during an audiovisual congruency task have a supramodal
797 format, following computations on the joint distribution or on the comparisons of the auditory and
798 visual signals. Thus, metacognition may be supramodal because of supramodal formats of confidence
799 estimates. Second, our electrophysiological results revealed that increased confidence in a visual or
800 audiovisual task coincided with the amplitude of ERP and decreased alpha power prior to type 1
801 response, suggesting that decisional cues may be a determinant of metacognitive monitoring. Thus,
802 metacognition may be supramodal not only because confidence estimates are supramodal by nature, but

803 also because they may be informed by decisional and movement preparatory signals that are shared
804 across modalities.

805

806 **References**

- 807 Ais J, Zylberberg A, Barttfeld P, Sigman M (2016) Individual consistency in the accuracy and
808 distribution of confidence judgments. *Cognition* 146:377–386.
- 809 Bagiella, E., Sloan, R. P., & Heitjan, D. F. (2000). Mixed-effects models in psychophysiology.
810 *Psychophysiology*, 37(1), 13-20.
- 811 Baird B, Mrazek MD, Phillips DT, Schooler JW (2014) Domain-specific enhancement of
812 metacognitive ability following meditation training. *Journal of Experimental Psychology: General*
813 143:1972–1979.
- 814 Barrett AB, Dienes Z, Seth AK (2013) Measures of metacognition on signal-detection theoretic models.
815 *Psychological methods* 18:535–552.
- 816 Bates D, Maechler M, Bolker B, Walker S (2014) lme4: linear mixed-effects models using S4 classes.
817 R package version 1.1-6. R.
- 818 Bates DM, Kliegl R, Vasishth S, Baayen H (2015) Parsimonious mixed models. arXiv preprint
819 arXiv:150604967:1–27.
- 820 Baumgarten TJ, Schnitzler A, Lange J (2016) Prestimulus Alpha Power Influences Tactile Temporal
821 Perceptual Discrimination and Confidence in Decisions. *Cerebral Cortex* 26:891–903.
- 822 Boldt A, Yeung N (2015) Shared neural markers of decision confidence and error detection. *The*
823 *Journal of neuroscience : the official journal of the Society for Neuroscience* 35:3478–3484.
- 824 Brainard DH (1997) The Psychophysics Toolbox. *Spatial Vision* 10:433–436.
- 825 Buján A, Lindín M, Díaz F (2009) Movement related cortical potentials in a face naming task:
826 Influence of the tip-of-the-tongue state. *International Journal of Psychophysiology* 72:235–245.
- 827 Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-
828 theoretic approach.
- 829 Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of
830 independent components of the electroencephalogram for artifact correction. *Journal of*
831 *Neuroscience Methods*. doi:10.1016/j.jneumeth.2015.02.025
- 832 De Gardelle V, Le Corre F, Mamassian P (2016) Confidence as a common currency between vision
833 and audition. *PLoS ONE* 11.
- 834 De Gardelle, V., Mamassian, P. (2014). Does Confidence Use a Common Currency Across Two
835 Visual Tasks? *Psychological Science*, 25 (6) 1286-1288 DOI: 10.1177/0956797614528956.
- 836 Delorme A, Makeig S (2004) EEGLAB: An open source toolbox for analysis of single-trial EEG
837 dynamics including independent component analysis. *Journal of Neuroscience Methods* 134:9–21.

- 838 Deroy O, Spence C, Noppeney U (2016) Metacognition in Multisensory Perception. *Trends in*
839 *Cognitive Sciences* 20:736–747.
- 840 Eimer M, Coles MGH (2003) The lateralized readiness potential as an on-line measure of central
841 response activation processes. *Behavior Research Methods, Instruments, & Computers* 30:146–
842 156.
- 843 Fleming S, Dolan R, Frith C (2012) Metacognition: computation, biology and function. *Philosophical*
844 *Transactions of the Royal Society B: Biological Sciences* 367:1280–1286.
- 845 Fleming, S.M., Ryu, J., Golfinos, J.G. & Blackmon, K.E. (2014) Domain-specific impairment in
846 metacognitive accuracy following anterior prefrontal lesions. *Brain*, 137 (10): 2811–2822
- 847 Fleming SM, Lau HC (2014) How to measure metacognition. *Frontiers in Human Neuroscience* 8:443.
- 848 Fleming SM, Maniscalco B, Ko Y, Amendi N, Ro T, Lau H (2015) Action-specific disruption of
849 perceptual confidence. *Psychological Science* 26:89–98.
- 850 Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G (2010) Relating introspective accuracy to individual
851 differences in brain structure. *Science (New York, NY)* 329:1541–1543.
- 852 Foxe JJ, Simpson G V, Ahlfors SP (1998) Parieto-occipital approximately 10 Hz activity reflects
853 anticipatory state of visual attention mechanisms. *Neuroreport* 9:3929–3933.
- 854 Foxe JJ, Snyder AC (2011) The role of alpha-band brain oscillations as a sensory suppression
855 mechanism during selective attention. *Frontiers in Psychology* 2.
- 856 Galvin SJ, Podd J V, Drga V, Whitmore J (2003) Type 2 tasks in the theory of signal detectability:
857 discrimination between correct and incorrect decisions. *Psychonomic bulletin & review* 10:843–
858 876.
- 859 Grimaldi P, Lau H, Basso MA (2015) There are things that we know that we know, and there are things
860 that we do not know we do not know: Confidence in decision-making. *Neuroscience and*
861 *Biobehavioral Reviews* 55:88–97.
- 862 Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*,
863 220(4598), 671–680.
- 864 Kleiner M, et al. (2007) What’s new in Psychtoolbox-3? *Perception* 36:S14.
- 865 Klimesch W (2012) Alpha-band oscillations, attention, and controlled access to stored information.
866 *Trends in Cognitive Sciences* 16:606–617.
- 867 Koriat A (2006) Metacognition and consciousness. In: *The Cambridge Handbook of Consciousness*, pp
868 289–326.

- 869 Kuznetsova A, Brockhoff PB, Christensen HB (2014) lmerTest: Tests for random and fixed effects for
870 linear mixed effect models (lmer objects of lme4 package). R package version:R package version
871 2.0–6.
- 872 Kvam P. D., Pleskac T. J., Yu S., & Busemeyer J. R. (2015). Interference effects of choice on
873 confidence: Quantum characteristics of evidence accumulation. PNAS Proceedings of the
874 National Academy of Sciences of the United States of America, 112, 10645–10650.
875 10.1073/pnas.1500688112
- 876 Lefcheck JS (2016) piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution,
877 and systematics. *Methods in Ecology and Evolution* 7:573–579.
- 878 Levitt H (1971) Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical*
879 *society of America*:467–477.
- 880 Makeig, S., J. Bell., A., Jung, T.-P., & Sejnowski, T. J. (1996). Independent Component Analysis of
881 Electroencephalographic Data. In *Advances in Neural Information Processing Systems* (Vol. 8, pp.
882 145–151). doi:10.1109/ICOSP.2002.1180091
- 883 Maniscalco B, Lau H (2012) A signal detection theoretic approach for estimating metacognitive
884 sensitivity from confidence ratings. *Consciousness and cognition* 21:422–430.
- 885 McCurdy LY, Maniscalco B, Metcalfe J, Liu KY, de Lange FP, Lau H (2013) Anatomical coupling
886 between distinct metacognitive systems for memory and visual perception. *The Journal of*
887 *neuroscience : the official journal of the Society for Neuroscience* 33:1897–1906.
- 888 Meredith, M., & Ridout, M. (2016). overlap: Estimates of Coefficient of Overlapping for Animal
889 Activity Patterns.
- 890 Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact
891 detector based on the joint use of spatial and temporal features. *Psychophysiology*, 48(2), 229–240.
- 892 Morey, R., Rouder, J., & Jamil, T. (2015). Package “BayesFactor.”
- 893 Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R² from generalized
894 linear mixed-effects models. *Methods in Ecology and Evolution* 4:133–142.
- 895 Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: Open source software for advanced
896 analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and*
897 *neuroscience* 2011:156869.
- 898 Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into
899 movies. *Spatial Vision* 10(4):437–442.

- 900 Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and
901 current density mapping. *Electroencephalography and Clinical Neurophysiology*, 72(2), 184–187.
902 doi:10.1016/0013-4694(89)90180-6
- 903 Pfurtscheller G (1992) Event-related synchronization (ERS): an electrophysiological correlate of
904 cortical areas at rest. *Electroencephalography and Clinical Neurophysiology* 83:62–69.
- 905 Pfurtscheller G, Lopes Da Silva FH (1999) Event-related EEG/MEG synchronization and
906 desynchronization: Basic principles. *Clinical Neurophysiology* 110:1842–1857.
- 907 Pinheiro JC, Bates DM (2010) The NLME package. October.
- 908 Pleskac TJ, Busemeyer JR, others (2010) Two-stage dynamic signal detection: a theory of choice,
909 decision time, and confidence. *Psychological review* 117:864.
- 910 Pouget A, Drugowitsch J, Kepecs A (2016) Confidence and certainty: distinct probabilistic quantities
911 for different goals. *Nature Neuroscience* 19:366–374.
- 912 Rahnev, D., Koizumi, A., McCurdy, L. Y., D’Esposito, M., & Lau, H. (2015). Confidence leak in
913 perceptual decision making. *Psychological science*, 0956797615595037.
- 914 Samaha J, Iemi L, Postle B (2016) Prestimulus alpha-band power biases visual discrimination
915 confidence, but not accuracy. *bioRxiv*.
- 916 Siedlecka M., Paulewicz B., & Wierchoń M. (2016). But I was so sure! Metacognitive judgments are
917 less accurate given prospectively than retrospectively. *Frontiers in Psychology*, 7, 218.
- 918 Singmann, H., Bolker, B., & Westfall, J. (2015). afex: Analysis of Factorial Experiments.
- 919 Solovey G, Graney GG, Lau H (2015) A decisional account of subjective inflation of visual perception
920 at the periphery. *Attention, perception & psychophysics* 77:258–271.
- 921 Song C, Kanai R, Fleming SM, Weil RS, Schwarzkopf DS, Rees G (2011) Relating inter-individual
922 differences in metacognitive performance on different perceptual tasks. *Consciousness and*
923 *cognition* 20:1787–1792.
- 924 Whitmarsh S, Oostenveld R, Almeida R, Lundqvist D (2016) Metacognition of attention during tactile
925 discrimination. *Neuroimage*.
- 926 Wokke, M. E., Cleeremans, A., & Ridderinkhof, K. R. (2017). Sure I’m Sure: Prefrontal Oscillations
927 Support Metacognitive Monitoring of Decision Making. *Journal of Neuroscience*, 37(4), 781-789.
- 928 Yeung N, Summerfield C (2012) Metacognition in human decision-making: confidence and error
929 monitoring. *Philosophical transactions of the Royal Society of London Series B, Biological*
930 *sciences* 367:1310–1321.

931 Yokoyama O, Miura N, Watanabe J, Takemoto A, Uchida S, Sugiura M, Horie K, Sato S, Kawashima
 932 R, Nakamura K (2010) Right frontopolar cortex activity correlates with reliability of retrospective
 933 rating of confidence in short-term recognition memory performance. *Neuroscience Research*
 934 68:199–206.

935

936 Legends

937

938 Figure 1: **Experimental procedure.** Participants had to perform a perceptual task on a stimulus (first-order task), and then
 939 indicate their confidence in their response by placing a cursor on a visual analog scale (second-order task). The types of
 940 stimuli and first-order task varied across conditions and experiments, as represented schematically on the right panel. In
 941 Experiment 1, a pair of two images, sounds, or tactile vibrations was presented on each trial. The stimuli of each pair were
 942 lateralized and differed in intensity (here high intensity is depicted in red, low intensity in pink). The first-order task was to
 943 indicate whether the most intense stimulus was located on the right (as depicted here) or left side. In Experiment 2, either
 944 two pairs of two images (unimodal visual condition), two sounds (unimodal auditory condition), or one pair of two images
 945 with one pair of two sounds (bimodal audiovisual condition) were presented on each trial. The first-order task was to
 946 indicate whether the most intense stimulus of each pair were both on the same side (congruent trial), or each on a different
 947 side (incongruent trial, as depicted here). Experiment 3 was a replication of Experiment 2 including EEG recordings,
 948 focusing on the unimodal visual condition and the bimodal audiovisual condition. The order of conditions within each
 949 experiment was counterbalanced across participants.

950 Figure 2: **Upper row:** Violin plots representing first-order sensitivity (**a:** d'), metacognitive sensitivity (**b:** meta- d'), and
 951 metacognitive efficiency (**c:** meta- d'/d') in the auditory (A, in red), tactile (T, in green), and visual modalities (V, in blue).
 952 Full dots represent individual data points. Empty circles represent average estimates. Error bars represent the standard
 953 deviation. The results show that independently of first-order performance, metacognitive efficiency is higher in vision
 954 compared to audition. **Lower row:** correlations between individual metacognitive efficiencies in the visual and auditory
 955 conditions (**2d**), visual and tactile conditions (**2e**), and tactile and auditory conditions (**2f**). The results show that
 956 metacognitive efficiency correlates across sensory modalities, providing evidence in favor of the supramodality hypothesis.
 957 *** $p < 0.001$, ** $p < 0.01$, $\cdot p < 0.1$.

958 Figure 3: **Upper row:** Violin plots representing first-order sensitivity (**3a:** d'), metacognitive sensitivity (**3b:** meta- d'), and
 959 metacognitive efficiency (**3c:** meta- d'/d') in the auditory (A, in red), audiovisual (AV, in green), and visual modalities (V, in
 960 blue). Full dots represent individual data points. Empty circles represent average estimates. Error bars represent the standard
 961 deviation. The results show that independently of first-order performance, metacognitive efficiency is better for visual
 962 stimuli vs. auditory or audiovisual stimuli, but not poorer for audiovisual vs. auditory stimuli. **Lower row:** correlations
 963 between individual metacognitive efficiencies in the visual and auditory conditions (**3d**), audiovisual and auditory
 964 conditions (**3e**), and audiovisual and visual conditions (**3f**). The results show that metacognitive efficiency correlates
 965 between unimodal and bimodal perceptual tasks, in favor of the supramodality hypothesis. ** $p < 0.01$, * $p < 0.05$.

966 Figure 4: **Top row:** Parameters estimation in the unimodal visual and unimodal auditory conditions. In the middle panel,
 967 circles represent the partially overlapping bivariate internal signal distributions for each of the stimulus combinations,
 968 represented at a fixed density contour. The top right quadrant corresponds to congruent stimuli, where the stimuli in each
 969 pair were stronger on the right side. The colours represent the predicted confidence, normalized to the interval [0,1] for
 970 every combination of internal signal strength for each stimulus pair (X_1, X_2). Parameters for internal noise (σ) and criterion
 971 (c) were defined for each participant based on the fitting of response rates (“congruent”/“incongruent” and “sure”/“unsure”
 972 based on a median split of confidence ratings) in the unimodal visual (left panel) and auditory (right panel) conditions. The
 973 thick black and gray lines correspond respectively to observed responses in congruent and incongruent trials for a
 974 representative participant. The red lines represent the response rates predicted by the model with fitted parameters. **Middle**
 975 **row:** Model predictions. Modeling of bimodal data based on the combination of c_A, c_V and σ_A, σ_V according to integrative

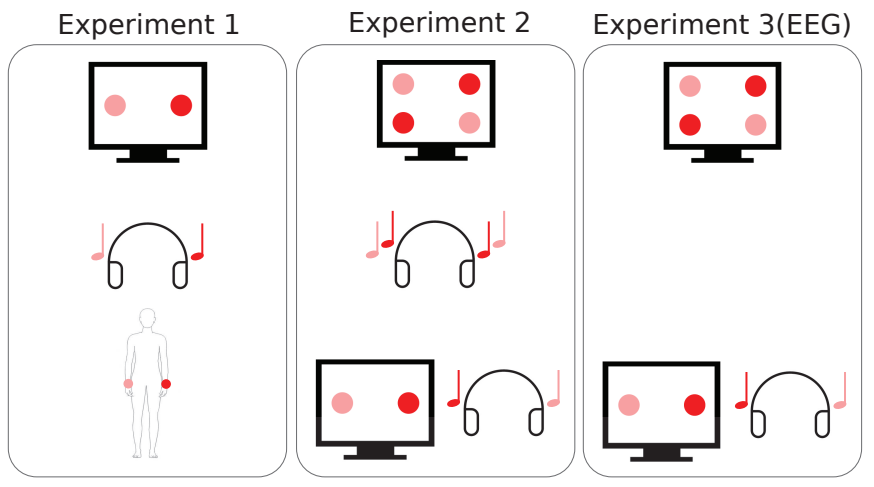
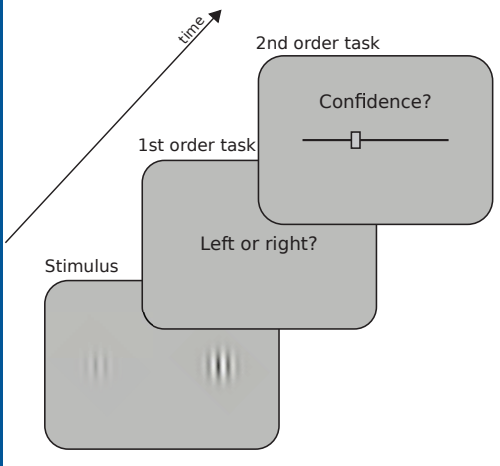
976 (A, in blue), comparative (B, in red), and single-modality rules (C, in green). Note that for models A and B, confidence
 977 increases with increasing internal signal level in both modalities, whereas in the single-modality model C, confidence
 978 depends on the signal strength of only one modality. **Lower row:** Model comparison for the audiovisual condition. Left
 979 panel: fit of response rates in the audiovisual condition for a representative participant according to model A (blue), B (red)
 980 and C (green). Right panel: Individual BIC weights for the three model fits. The arrows show how to read the plot from an
 981 arbitrary data point in the diagram, indicated with a red triangle. Consider that the sum of the BICw for all models A, B and
 982 C amounts to 1 for each participant. To estimate the relative BICw of each model for any given participant, take the lines
 983 parallel to the vertex labeled 1 for that model. The intersection between the line parallel to the vertex and the triangle edge
 984 corresponding to the model indicates the BICw.

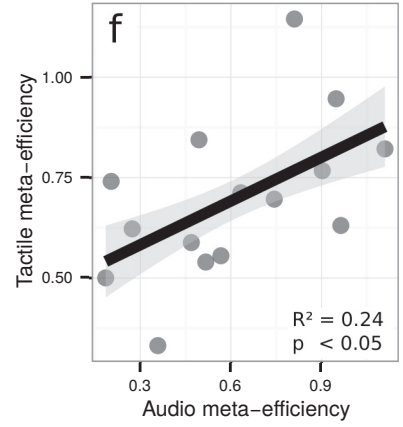
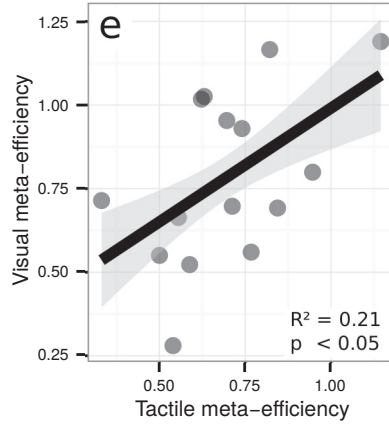
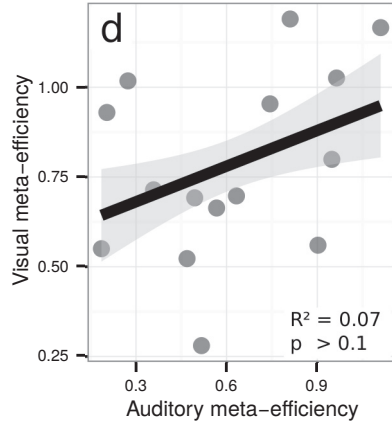
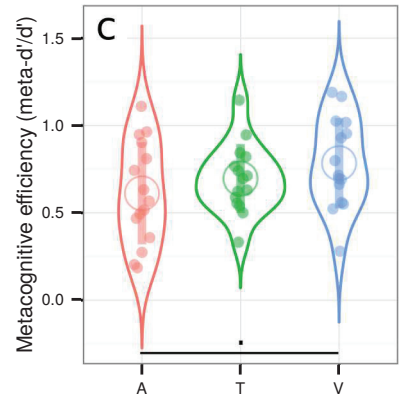
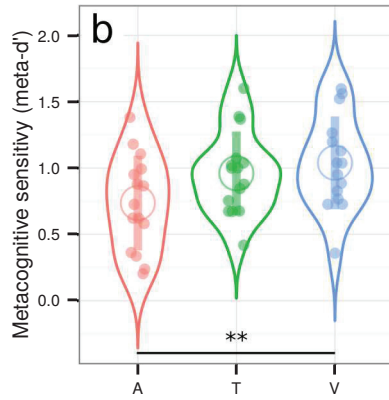
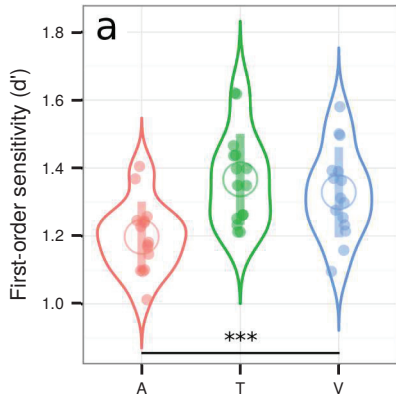
985 **Figure 5:** Violin plots representing first-order sensitivity (**5a: d'**), metacognitive sensitivity (**5b: meta- d'**), and
 986 metacognitive efficiency (**5c: meta- d'/d'**) in the audiovisual (AV, in green), and visual conditions (V, in blue). Full dots
 987 represent individual data points. Empty circles represent average estimates. Error bars represent the standard deviation. The
 988 results show no difference between visual and audiovisual metacognitive efficiency. **5d:** correlation between individual
 989 metacognitive efficiencies in the audiovisual and visual conditions (5d).

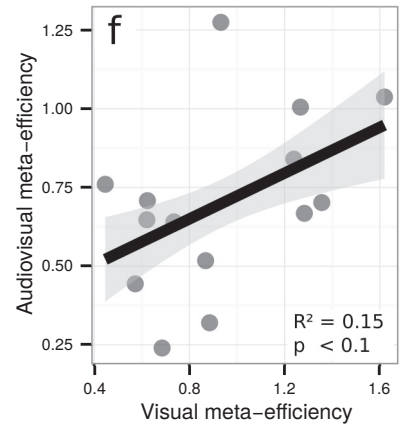
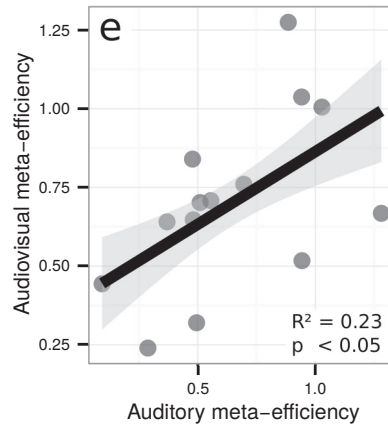
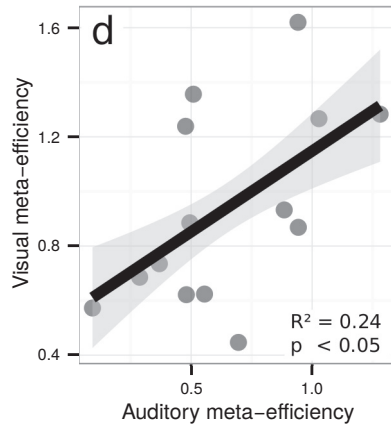
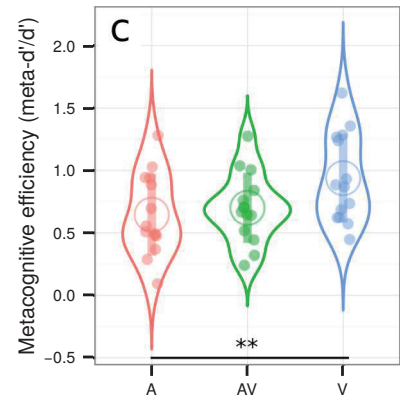
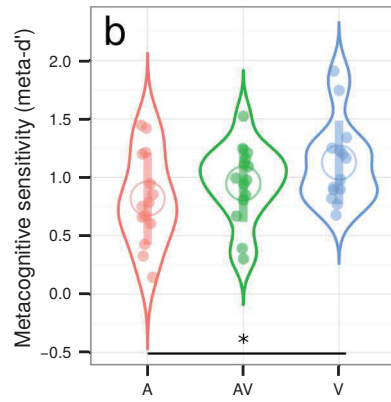
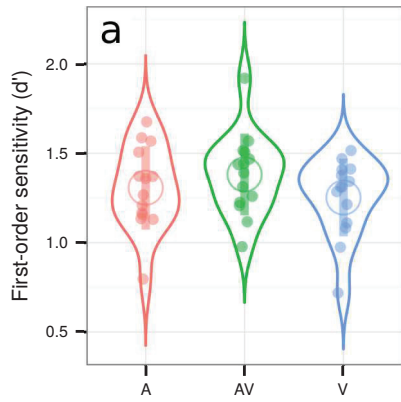
990 **Figure 6. Voltage amplitude time-locked to correct type 1 responses as a function of confidence. a. Left panel:** time
 991 course of the main effect of confidence within a pre-defined ROI. Although raw confidence ratings were used for the
 992 statistical analysis, they are depicted here as binned into four quartiles, from quartile 1 corresponding to trials with the 25%
 993 lowest confidence ratings (light pink), to quartile 4 corresponding to trials with the 25% highest confidence ratings (dark
 994 red). The size of each circle along the amplitude line is proportional to the corresponding F-value from mixed model
 995 analyses within 50 ms windows. **Right panel:** same analysis as shown in (a) on the whole scalp. The plot represents the
 996 time-course of the summed F-value over 64 electrodes for the main effect of confidence. The topography where a maximum
 997 F-value is reached (*) is shown next to each plot. **b. Left panel:** time course of the interaction between confidence and
 998 condition following a linear mixed model analysis within the same ROI as in (a). Although raw confidence ratings were
 999 used for the statistical analysis, the plot represents the difference in voltage amplitude between trials in the 4th vs. 1st
 1000 confidence quartile. **Right panel:** same analysis as shown in (b) on the whole scalp, with corresponding topography. In all
 1001 plots, grey bars correspond to significant main effects (a) or interactions (b), with $p < 0.05$ FDR-corrected. Significant
 1002 effects on topographies are highlighted with black stars ($p < 0.001$, uncorrected).

1003 **Figure 7. Alpha power time-locked to correct type 1 responses as a function of confidence. a. Left panel:** time course
 1004 of the main effect of confidence within a pre-defined ROI. Although raw confidence ratings were used for the statistical
 1005 analysis, they are depicted here as binned into four quartiles, from quartile 1 corresponding to trials with the 25% lowest
 1006 confidence ratings (light pink), to quartile 4 corresponding to trials with the 25% highest confidence ratings (dark red). The
 1007 size of each circle along the alpha power line is proportional to the corresponding F-value from mixed model analyses
 1008 within 50 ms windows. **Right panel:** same analysis as shown in (a) on the whole scalp. The plot represents the time-course
 1009 of the summed F-value over 64 electrodes for the main effect of confidence. The topography where a maximum F-value is
 1010 reached (*) is shown next to each plot. **b. Left panel:** time course of the interaction between confidence and condition
 1011 following a linear mixed model analysis within the same ROI as in (a). Although raw confidence ratings were used for the
 1012 statistical analysis, the plot represents the difference in voltage amplitude between trials in the 4th vs. 1st confidence
 1013 quartile. **Right panel:** same analysis as shown in (b) on the whole scalp, with corresponding topography. In all plots, grey
 1014 bars correspond to significant main effects (a) or interactions (b), with $p < 0.05$ FDR-corrected. Significant effects on
 1015 topographies are highlighted with black stars ($p < 0.001$, uncorrected).

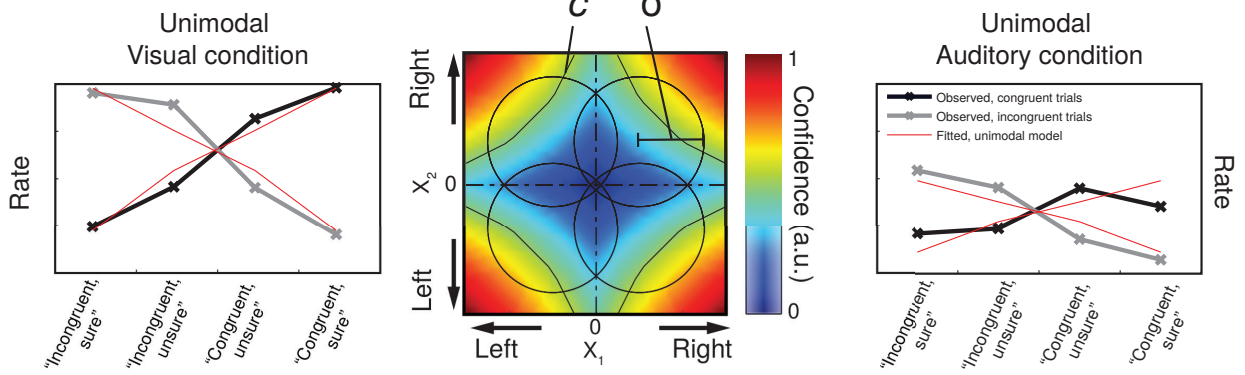
1016



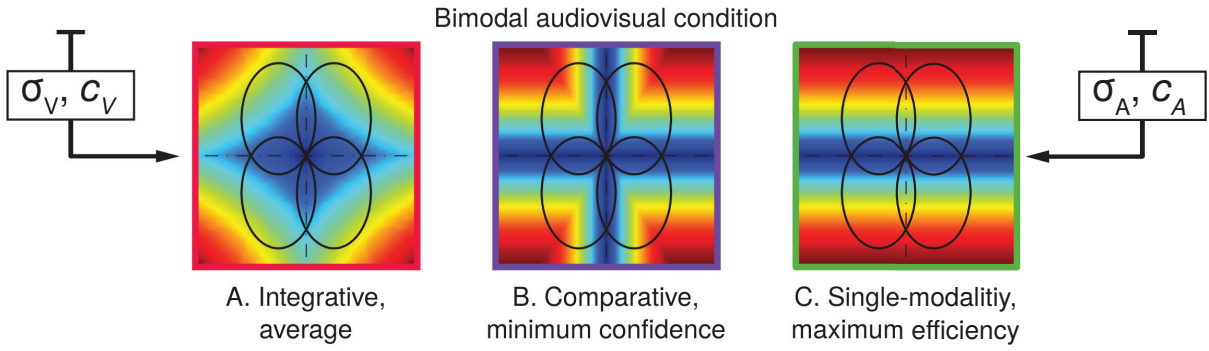




Parameters estimation



Model predictions



Model comparison

