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

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

JNeurosci
THE JOURNAL OF NEUROSCIENCE

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 Solovey[7], Simone Kühn[4,8] and Olaf Blanke[1,2,9]

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

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

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

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

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

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

[7]*Instituto de Cálculo, FCEyN, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina*

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

[9]*Department 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 Sk#odowska-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

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

Abbreviated title: Metacognition across senses and combination of senses

Authors: Nathan Faivre[1,2,3,*], Elisa Filevich[4,5,6], Guillermo Solovey[7], Simone Kühn[4,8], Olaf Blanke[1,2,9]

Affiliations:

1 Laboratory of Cognitive Neuroscience, Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Geneva, Switzerland
2 Center for Neuroprosthetics, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Geneva, Switzerland
3 Centre d'Economie de la Sorbonne, CNRS UMR 8174, Paris, France
4 Department of Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany
5 Department of Psychology, Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
6 Bernstein Center for Computational Neuroscience Berlin, Philippstr. 13 Haus 6, 10115 Berlin, Germany
7 Instituto de Cálculo, FCEyN, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina
8 Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistrasse 52 20246 Hamburg, Germany
9 Department of Neurology, University Hospital Geneva, Geneva, Switzerland

*** 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

Number of pages: 48
Number of figures: 7
Number of words: 150 (abstract), 680 (introduction), 1591 (discussion)

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

Acknowledgements
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.

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**.**

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

Abstract (150)

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

Significance statement (118 words)

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

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

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

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

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

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

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

General procedure

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

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

Behavioural analysis

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

Preprocessing of EEG data

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

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

Statistical analyses of EEG data

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

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

Signal-detection theory (SDT) models of behavior

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

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

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

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

Decision rule - type-1 task

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

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

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

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

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

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

If $d > 0$ the response given is "congruent", whereas if $d < 0$ the response is "incongruent". The values of (X_1, X_2) corresponding to $d = 0$, where the congruent and incongruent stimuli are equally likely, should satisfy the relation:

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

The solution to this relation should then satisfy:

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

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

$$\{X_1 = 0, X_2 \in \Re\}\ and\ \{X_2 = 0, X_1 \in \Re\}$$

Therefore, the internal response space (X_1, X_2) is divided in four quadrants such that an ideal observer 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 have different signs, the response will be incongruent.

Confidence judgment - type 2 task

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

$$conf \propto P(correct|\, decision, X_1, X_2)$$

If a response is "congruent", a participant's confidence in that response is then:

$$Conf(X_1, X_2) \propto P(RR|X_{1,}, X_2) + P(LL|X_{1,}, X_2)$$

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

Again applying Bayes' rule, and assuming that confidence in the unimodal condition is calculated on the basis of the joint distribution and hence $P(X1, X2|RR) = P(X1|R) \cdot P(X2|R)$, it follows that:

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

Assuming equal priors $P(L) = P(R)$ and given that $P(X_1) = P(X_1|R) + P(X_1|L)$ and $P(X_2) = P(X_2|R) + P(X_2|L)$ the expression above can be rewritten as:

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

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

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

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

Modeling strategy

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

Phase 1 - fits to the unimodal conditions

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

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

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

$$P(n_{obs}|N, P_{exp}) = \begin{cases} \frac{N!}{n^1_{obs}! \cdot \ldots \cdot n^8_{obs}!} \cdot p^1_{exp} \cdot \ldots \cdot p^8_{exp}, & when \sum_{i=1}^{8} n^i_{obs} = n \\ 0 & otherwise \end{cases}$$

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

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

Phase 2 - predictions of the bimodal condition

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

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

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

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

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

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

Model fits

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

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

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

Results

Experiment 1

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

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

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

We then turned to our main experimental question. We found positive correlations between metacognitive efficiency in the visual and tactile conditions [adjusted $R^2 = 0.21$, $p = 0.047$] (Figure 2e), and in the auditory and tactile conditions [adjusted $R^2 = 0.24$, $p = 0.038$] (Figure 2f). (The data were inconclusive regarding the correlation between the visual and auditory condition [adjusted $R^2 = 0.07$, $p = 0.17$, Bayes Factor = 0.86] (Figure 2d). These results reveal shared variance between auditory, tactile, and visual metacognition, in line with the supramodality hypothesis. Moreover, the absence of any correlation between first-order sensitivity and metacognitive efficiency in any of the conditions [all adjusted $R^2 < 0$; all p-values > 0.19], rules out the possibility that such supramodality during the 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, η p 2 = 0.02] or type 2 reaction times [F(1.77,24.84) =
428 1.77, p = 0.39, η p 2 = 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

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

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

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

[Figure 4 Here]

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

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

Behavioural data

The staircase procedure minimized variations in first-order sensitivity [$t(17) = 0.3$, $p = 0.76$, $d = 0.07$], such that sensitivity in the audiovisual [mean d' = 1.15 ± 0.07] and visual conditions [mean d' = 1.17 ± 0.05] were similar. Contrary to what was found in Experiments 1 and 2, response criterion varied across conditions [$t(17) = 4.33$, $p < 0.001$, $d = 0.63$], with a tendency to respond "congruent" more pronounced in the audiovisual [mean criterion = 0.27 ± 0.12] vs. visual condition [mean criterion = -0.02 ± 0.15]. This effect was unexpected but did not preclude from running subsequent analyses dealing with metacognitive sensitivity. We found no effect of condition on average confidence [$t(17) = 0.56$, $p = 0.14$, $d = 0.08$]. Average confidence ratings correlated between the visual and audiovisual conditions [adjusted $R2 = 0.65$, $p < 0.001$]. No difference in metacognitive sensitivity was found between conditions [$t(17) = 0.78$, $p = 0.44$, $d = 0.09$] or efficiency [$t(17) = 0.78$, $p = 0.44$, $d = 0.08$]. Crucially, we replicated our main results from Experiment 2, as we found a positive significant correlation between relative metacognitive accuracy in the audiovisual and visual conditions [adjusted $R^2 = 0.47$, $p < 0.001$], and no correlation between first-order sensitivity and metacognitive efficiency in either condition [both $R^2 < 0.01$; both p-values > 0.3] (Figure 5). Regarding the decisional locus of 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

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

[Figure 6 Here]

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

[Figure 7 Here]

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

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

The finding of lower alpha power with confidence in correct responses is compatible with the observation that participants with more distinct reaction times between correct and incorrect responses 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., & Wierzchoń 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.

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

Legends

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

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

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

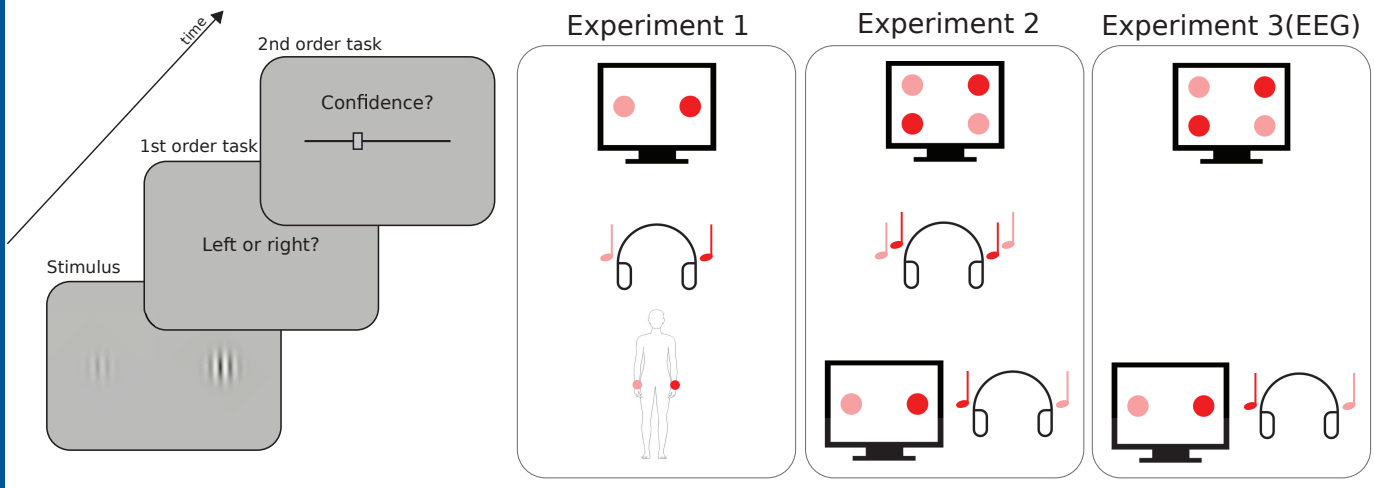
Figure 4: **Top row**: Parameters estimation in the unimodal visual and unimodal auditory conditions. In the middle panel, circles represent the partially overlapping bivariate internal signal distributions for each of the stimulus combinations, represented at a fixed density contour. The top right quadrant corresponds to congruent stimuli, where the stimuli in each pair were stronger on the right side. The colours represent the predicted confidence, normalized to the interval [0,1] for every combination of internal signal strength for each stimulus pair (X1, X2). Parameters for internal noise (σ) and criterion (c) were defined for each participant based on the fitting of response rates ("congruent"/"incongruent" and "sure"/"unsure" based on a median split of confidence ratings) in the unimodal visual (left panel) and auditory (right panel) conditions. The thick black and gray lines correspond respectively to observed responses in congruent and incongruent trials for a representative participant. The red lines represent the response rates predicted by the model with fitted parameters. **Middle row**: Model predictions. Modeling of bimodal data based on the combination of c_A, c_V and σ_A, σ_V according to integrative

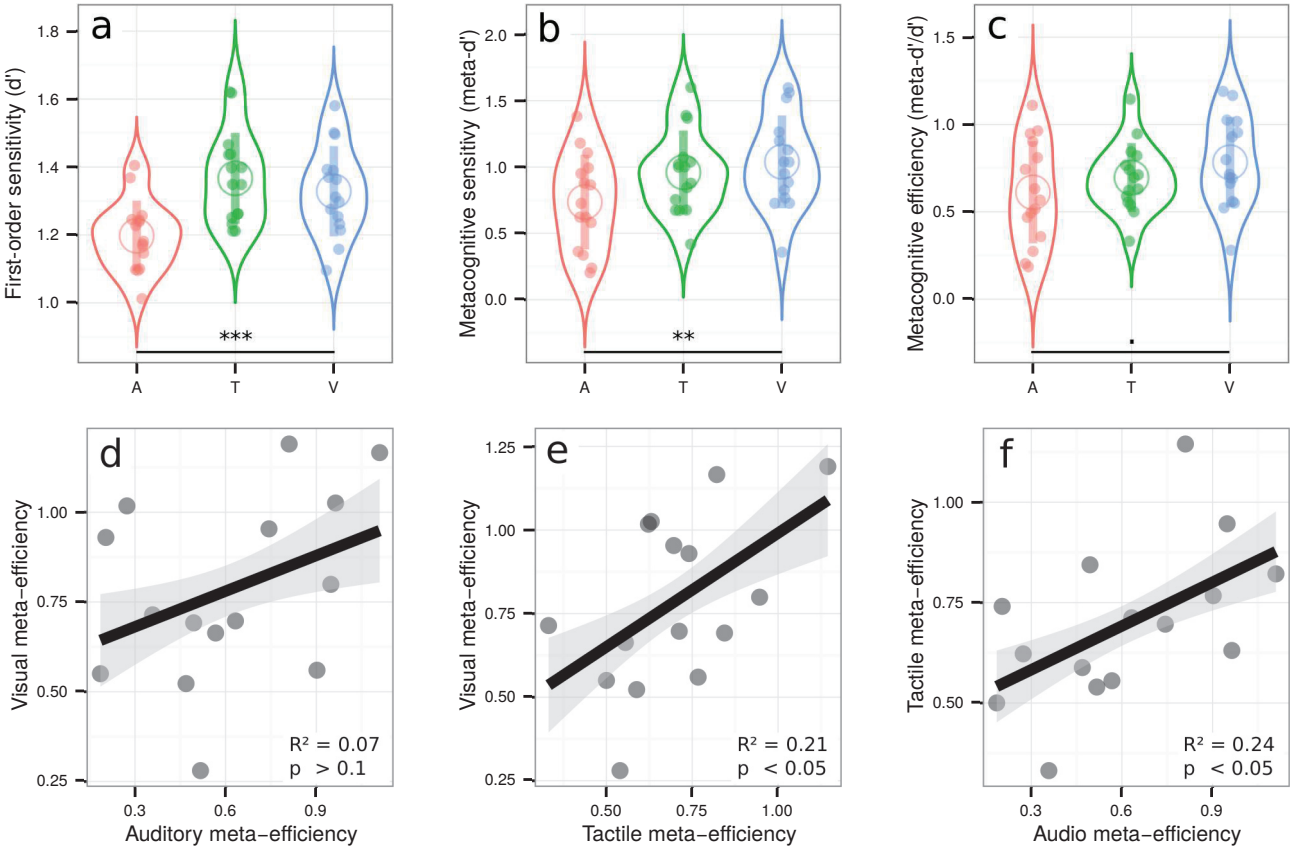
(A, in blue), comparative (B, in red), and single-modality rules (C, in green). Note that for models A and B, confidence increases with increasing internal signal level in both modalities, whereas in the single-modality model C, confidence depends on the signal strength of only one modality. **Lower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

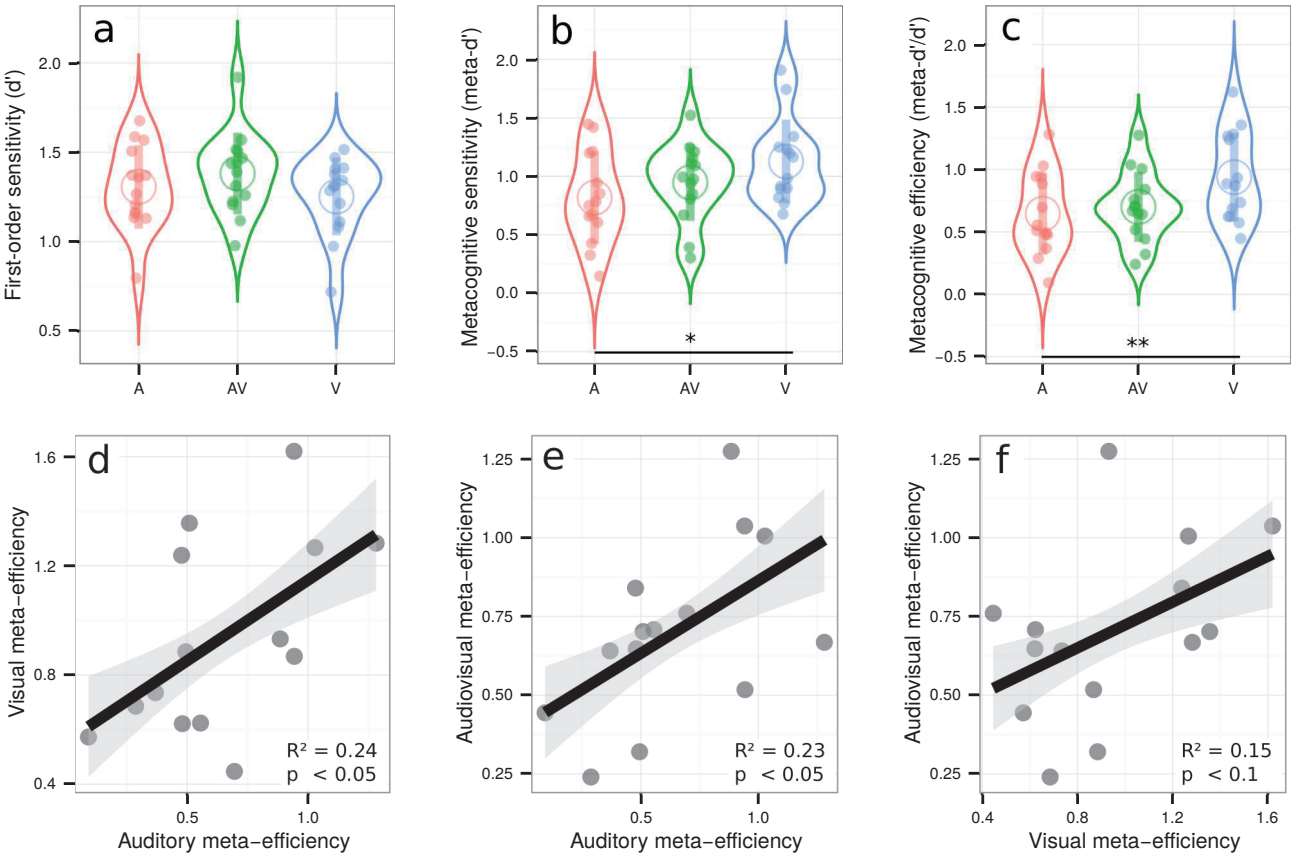
Figure 5: Violin plots representing first-order sensitivity (**5a**: d'), metacognitive sensitivity (**5b**: meta-d'), and metacognitive efficiency (**5c**: meta-d'/d') in the audiovisual (AV, in green), and visual conditions (V, in blue). Full dots represent individual data points. Empty circles represent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

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

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







Parameters estimation

Model predictions

Model comparison