

Cognition and Behavior

Response-Related Signals Increase Confidence But Not Metacognitive Performance

Elisa Filevich,^{1,2,3} Christina Koß,^{1,3} and Nathan Faivre^{4,5,6}<https://doi.org/10.1523/ENEURO.0326-19.2020>

¹Bernstein Center for Computational Neuroscience Berlin, 10115 Berlin, Germany, ²Research Training Group 2386 “Extrospection”, Humboldt-Universität zu Berlin, 10099 Berlin, Germany, ³Institute of Psychology, Humboldt-Universität zu Berlin, 10099 Berlin, Germany, ⁴Laboratory of Cognitive Neuroscience, Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology, 8092 Geneva, Switzerland, ⁵Center for Neuroprosthetics, Faculty of Life Sciences, Swiss Federal Institute of Technology, 8092 Geneva, Switzerland, and ⁶Laboratoire de Psychologie et Neurocognition, CNRS UMR 5105, Université Grenoble Alpes, 38400 Saint-Martin-d’Hères, France

Abstract

Confidence judgments are a central tool in metacognition research. In a typical task, participants first perform perceptual (first-order) decisions and then rate their confidence in these decisions. The relationship between confidence and first-order accuracy is taken as a measure of metacognitive performance. Confidence is often assumed to stem from decision-monitoring processes alone, but processes that co-occur with the first-order decision may also play a role in confidence formation. In fact, some recent studies have revealed that directly manipulating motor regions in the brain, or the time of first-order decisions relative to second-order decisions, affects confidence judgments. This finding suggests that confidence could be informed by a readout of reaction times in addition to decision-monitoring processes. To test this possibility, we assessed the contribution of response-related signals to confidence and, in particular, to metacognitive performance (i.e., a measure of the adequacy of these confidence judgments). In human volunteers, we measured the effect of making an overt (vs covert) decision, as well as the effect of pairing an action to the stimulus about which the first-order decision is made. Against our expectations, we found no differences in overall confidence or metacognitive performance when first-order responses were covert as opposed to overt. Further, actions paired to visual stimuli presented led to higher confidence ratings, but did not affect metacognitive performance. These results suggest that confidence ratings do not always incorporate motor information.

Key words: confidence; metacognition

Significance Statement

To measure metacognition, or the ability to monitor one’s own thoughts, experimental tasks often require human volunteers to, first, make a perceptual decision (“first-order task”) and, then, rate their confidence in their own decision (“second-order task”). In this paradigm, both first-order and second-order information could, in principle, influence confidence judgments. But only the latter is truly metacognitive. To determine whether confidence is a valid metacognitive measure, we compared confidence ratings between the following two conditions: with overt responses, where participants provided both first-order and second-order responses; and with covert responses where participants reported their confidence in a decision that they had not executed. Removing first-order decisions did not affect confidence, which validates confidence as an introspective measure.

Received August 15, 2019; accepted February 28, 2020; First published April 23, 2020.

The authors declare no competing financial interests.

Author contributions: E.F. and N.F. designed research; C.K. performed research; E.F., C.K., and N.F. analyzed data; E.F., C.K., and N.F. wrote the paper.

Introduction

Confidence judgments about one's own perception have been exploited in recent years as a useful way to probe introspection (Fleming and Dolan, 2012). In a now standard paradigm, participants first make a binary decision (typically, a perceptual or memory judgment, first-order task) and afterward rate the confidence in their response (second-order task). Metacognitive performance is measured as the relationship between accuracy in the first-order task and confidence in the second-order task (Fleming and Lau, 2014). Crucially, it is still unclear what confidence reports actually represent, as the variables participants compute to generate them remain latent.

Under a normative view, confidence is a finer-grained description of the same perceptual evidence that leads to the binary first-order decision and, specifically, correspond to the probability of giving a correct answer given the available perceptual discriminability (Pouget et al., 2016; Sanders et al., 2016). In other words, whereas participants choose between two options in the first-order task, they have the chance to more precisely describe the difficulty of their perceptual experience through confidence reports in the second-order task. In this view, introspection is required to produce accurate confidence reports. But recent results have challenged this standard view of confidence as a description of perceptual evidence by showing that, beyond perceptual evidence, sensorimotor signals associated with the response to the first-order task may also contribute to confidence. At its simplest, this effect is manifest as a negative correlation between first-order reaction times (RTs) and confidence reports (Henmon, 1911; Baranski and Petrusic, 1995), which can be explained by bounded evidence accumulation models (Pleskac and Busemeyer, 2010; Ratcliff and Starns, 2013; Moran et al., 2015). The dependency is strong when accuracy is stressed but is greatly reduced (Vickers and Packer, 1982) or disappears altogether when speed is emphasized, suggesting that the influence of predecisional and postdecisional cues on confidence depends on the task demands (Baranski and Petrusic, 1998). Nevertheless, overall, data from a wide range of recent tasks measuring confidence following discrimination decisions shows that an overwhelming majority of participants present a negative relationship between confidence and decision reaction times (Rahnev et al., 2019).

E.F. and C.K. are supported by a Freigeist Fellowship to E.F. from the Volkswagen Foundation (Grant 91620) and by the Deutsche Forschungsgemeinschaft Grant 337619223/RTG2386. N.F. is supported by a Starting Grant from the European Research Council (803122).

Acknowledgments: We thank Lukas Röd, Carina Forster, and Marco Wirthlin for help with data collection. We also thank Guillermo Bernabó for help with machine learning analyses, and Michael Pereira and Matthias Guggenmos for helpful comments on an earlier version of this manuscript.

Correspondence should be addressed to Elisa Filevich at elisa.filevich@bccn-berlin.de.

<https://doi.org/10.1523/ENEURO.0326-19.2020>

Copyright © 2020 Filevich et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Evidence from comparisons between participants further supports this idea: metacognitive performance was better in participants with large differences in response times between correct and incorrect responses (Faivre et al., 2018).

Beyond behavior alone, Gajdos et al. (2019) showed that confidence increases in the presence of subthreshold motor activity before first-order responses. Plus, we recently showed that α -desynchronization before first-order response (an electrophysiological signature of motor preparation) correlates with confidence over different perceptual tasks (Faivre et al., 2018), that metacognitive performance for decisions that are committed with a keypress is better than that for equivalent decisions that are observed (Pereira et al., 2020), and that sensorimotor conflicts alter confidence (Faivre et al., 2020). Finally, transcranial magnetic stimulation (TMS) directed at the premotor cortex involved in the first-order response was found to affect confidence ratings, suggesting a causal role of action-related signals for confidence (Fleming et al., 2015).

Experimental manipulations that artificially change the process of evidence accumulation have provided strong mechanistic explanations for this relationship (Fetsch et al., 2014; Kiani et al., 2014; Zylberberg et al., 2016). But these manipulations ultimately affected the evidence available to the observer or the process of accumulation itself. Here, we sought to compare confidence judgments and metacognitive performance between conditions that differed only on the sensorimotor information available for the decision, but that were indistinguishable from the point of view of perceptual evidence. We hypothesized that response-related sensorimotor activity carries information useful for confidence judgments, above and beyond the strength of the (perceptual) internal signal. We designed a paradigm in which participants saw visual stimulus that moved alternatively rightward or leftward for 5 s, and rated their confidence in their capacity to discriminate the motion direction that was presented for the longest duration. Following a preregistered plan (<https://osf.io/hnvsb/>), we compared conditions with and without overt motor discrimination responses and predicted that conditions with overt first-order two-alternative forced choice (2AFC) responses would reveal better metacognitive performance than those without them.

Importantly, we note that the logic of this design relies on the strong assumption that participants committed to a binary decision even in cases with no overt first-order responses, and that the confidence judgments reflected this. We discuss the implications that follow if these assumptions are not met.

Materials and Methods

Participants. Twenty-seven participants took part in this study, of whom 4 had to be excluded (see below). The results we report here correspond to a sample of 23 participants (13 males, 10 females) with a mean \pm SD age of 26.7 ± 5 years. All participants had normal or corrected-to-normal vision and no color blindness, and were right handed. Ten participants were tested in Berlin, the rest in

Table 1: Statistical table

	Data structure	Type of test	Power/effect size	Statistic
a: Performance in first-order task (CR+R+ vs CR– R+)	Mean per subject (continuous)	Bayesian <i>t</i> test	$d = -0.02$	$BF_{10} = 0.22$
b: Mean perceptual evidence (CR+R+ vs CR–R+)	Mean per subject (continuous)	Bayesian <i>t</i> test	$d = 0.36$	$BF_{10} = 0.54$
c: interaction between overt first-order (R+/R–) and continuous report (CR+/CR–) on mean confidence (all trials)	Single-trial confidence ratings (continuous)	Bayesian mixed-effects linear regression	Mean = -0.02 ± 0.01	Evidence ratio = 0.10
d: Main effect of CR+ on confidence (all trials)	Single-trial confidence ratings (continuous)	Bayesian mixed-effects linear regression	Mean = 0.04 ± 0.02	Evidence ratio = 75.92
e: Main effect of first-order RT on confidence (R+ trials)	Single-trial response times (continuous)	Bayesian mixed-effects linear regression	Mean = -0.15 ± 0.02	Evidence ratio > 4000),
f: Interaction between condition (CR+R+/CR+R–) and confidence on proxy accuracy	Single-trial accuracy (binomial)	Bayesian mixed-effects logistic regression	Mean = -0.11 ± 0.39	Evidence ratio = 1.57
g: Main effect of confidence on proxy accuracy (CR+ trials)	Single-trial confidence ratings (continuous)	Bayesian mixed-effects linear regression	Mean = 0.82 ± 0.32	Evidence ratio = 116.65
h: M-ratio between conditions (CR+R+ vs CR+R–)	Posterior probability distributions (continuous)	Differences between Highest density intervals	HDI = $[-1.42, 0.89]$	NA

NA, Not applicable.

Geneva. All received monetary compensation for their time. The procedures were approved by the corresponding local ethics committees, and institutional review board in conformity with the Declaration of Helsinki. Written and signed informed consents were obtained from all participants.

Procedure. The experimental task was written in MATLAB (MathWorks) using Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Stimuli were red (RGBA color, 0.75, 0, 0, 1) or green (RGBA color, 0, 0.75, 0, 1) vertical gratings that drifted sideways. The gratings were formed by a sine-wave function (0.27 cycles/°), drifting sideways at 15°/s, and drawn inside a square (8° height and width), presented at fixation. The green and red stimuli always drifted leftward and rightward, respectively.

Each 5-s-long trial was divided into four intervals of different durations, during which four red and green stimuli were presented in alternation. The total, summed duration of each pair of same-colored stimulus presentations corresponded to half the trial length (2.5 s) plus or minus a temporal difference determined by a staircase (see below). Further, each single stimulus presentation interval corresponded to half of the sum of the stimulus pair length. The first-order 2AFC task consisted of a duration comparison, followed by a confidence rating (Fig. 1).

To evaluate the effects of overt movement on metacognitive judgments, we asked for two kinds of reports. In continuous-report (CR+) trials, participants pressed two arrow keys using two fingers of their right hand to indicate which of the two colored stimuli was presented on the screen. In this condition, the task was simply to press a key that “tracked” the motion direction of the stimulus. In

conditions without continuous report (CR–), participants did not press any keys during stimulus presentation. In trials with first-order 2AFC response (R+) trials, participants did a temporal-summation task. Upon stimulus offset, they indicated with a single key press which of the two motion directions had been presented for a longer period of time (i.e., which of the summed stimulus durations was the longest over the course of the entire 5 s trial). The response keys and hands used for the first-order 2AFC response were the same as for the continuous report. In conditions without the first-order 2AFC response (R–), participants were also required to make a temporal summation decision (the decision was overt in R+ trials but covert in R– trials). Each trial corresponded to one of four possible conditions, combining CR (+, present/–, absent) and R (+, present/–, absent). At the end of each trial, participants rated their confidence in their decision by moving a slider with two keys on a vertical visual analog scale with the ends marked as “very sure” and “very unsure.”

The duration difference was determined separately for CR+ and CR– trials using two independent 1-up, 2-down staircases (updated only following R+ trials). We also ran two pre-experiment staircases of 25 trials each, without confidence ratings, to adjust the difference in duration of the two stimuli for each participant. After the staircases, each participant completed 240 trials in total (60 trials per condition). Trial types were interleaved, and the order of the trials was randomized for each participant. On any given trial, participants were not informed beforehand whether a first-order response would be required. That is, after stimulus offset, participants were either prompted to give a 2AFC on the color corresponding to the longest duration, or were directly prompted to give a confidence

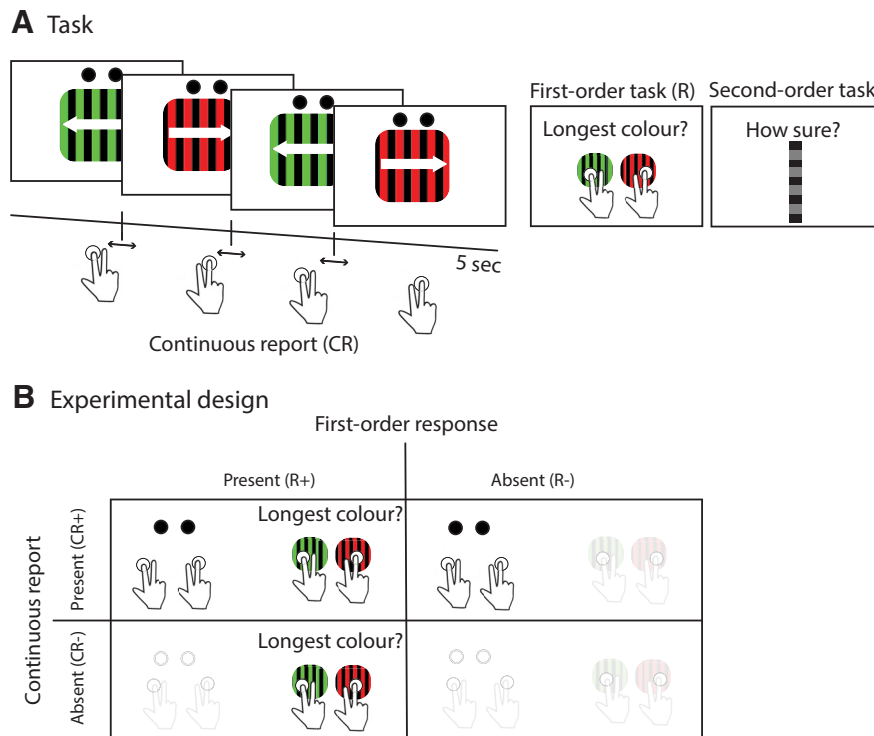


Figure 1. A, Task. **B**, Example trial with both continuous report (CR+) and first-order 2AFC response. On each 5 s trial, two stimuli pairs appeared serially in four consecutive intervals. Participants pressed one of two keys for the entire duration of the trial, tracking the visual presentation (continuous report). Following stimulus offset, participants reported which of the two stimuli had the longest duration overall. **B**, Experimental design. Each trial was one of the four possible conditions resulting from a combination of first-order 2AFC response and continuous report (CR+R+; CR+R-; CR-R+; or CR-R-). Participants rated their confidence in all conditions. Thus, the task demanded that participants make a first-order 2AFC judgment in every trial, but the corresponding overt action was only present in R+ conditions.

rating. Trials were self-paced and the experiment took on average 50 min.

Design rationale. As per the preregistration, we hypothesized that response-related sensorimotor activity carries information useful for confidence judgment, above and beyond the strength of the (perceptual) internal signal. We therefore expected that conditions with overt first-order 2AFC responses would be associated with better metacognitive performance than those without motor responses. In the same way, we expected that conditions in which a motor action was paired with the stimulus would also be associated with better metacognitive performance than those without motor responses.

As we anticipated in the Introduction and further elaborate in the Discussion, our analyses and conclusions are only valid if two assumptions are met. First, we assumed that participants committed to a binary decision even in cases with no overt first-order 2AFC responses. Related to that, we also assume that confidence reports in the two conditions are about the same quantity, and that participants reported their confidence in a binary decision that they (covertly) committed to and not, for example, to the uncertainty in the temporal accuracy of their continuous report.

Termination rule. Our plan at preregistration was to collect data until we reach a Bayes factor (BF_{10}) of either one-third or 3. We started by collecting a sample of 27

participants (four excluded) and examined the data once. With this sample size, we found evidence for the null hypothesis in our main test of interest (the interaction term between confidence and first-order response in the effect on accuracy as modeled by a logistic regression; see Confirmatory analyses below), so we halted data collection.

Analyses. We adhered to the exclusion criteria that were preregistered. Four participants were excluded because they did not follow the task instructions (in all cases, they did not press any keys during any of the trials in the continuous report conditions). No further participants were excluded, as none of them had first-order accuracy $<60\%$ or $>80\%$ in any task; and visual inspection of the staircases revealed no obvious problems. A total of 64 trials (from 17 participants) were excluded because first-order RTs were <200 ms or >5 s.

Metacognitive performance. As per the preregistration, we computed metacognitive efficiency (meta- d'/d') to quantify the capacity to adjust confidence regardless of the first-order task difficulty (Maniscalco and Lau, 2012) using the HMeta- d' (Fleming, 2017). For that, we scaled confidence judgments for each participant by subtracting from each rating the individual minimum rating and dividing the values by the total range. This procedure effectively “stretched” confidence distributions to fit the interval between 0 and 1 for all participants, thereby

eliminating biases between individuals while preserving mean differences between conditions. We then discretized scaled confidence values into four confidence bins. For the MCMC (Markov chain Monte Carlo method) procedure, we used four chains of 10,000 iterations including 1000 for adaptation, no thinning, and default initial values as generated by JAGS (Just Another Gibbs Sampler). Separate hierarchical estimates were computed for each condition. Potential scale reduction factors regarding average M-ratio estimates were equal to 1.02 (CR+R+), 1.02 (CR-R+), 1.06 (CR+R-), and 1.16 (CR-R-). Only the last value for CR+R- indicates a possible lack of convergence, so we refitted the model with 30,000 iterations including 10,000 for warmup, which resulted in scale reduction factors of 1.03 and 1.11, respectively, with no difference in M-ratios between conditions. These values still point to possible converge problems, presumably due to the relatively low number of trials in our sample.

In separate analyses, we estimated the slope parameter in a mixed-effects logistic regression with accuracy as the dependent variable and confidence as the independent variable. Because mixed-effects logistic regression analyses are not affected by subject-wise scaling of confidence (i.e., they include subject-wise random intercepts), we used raw confidence values as independent variables. For all models, we included a by-subject random slope for each of the main effects considered in the model, but not for their interactions. We ran Bayesian sampling of mixed regressions using the *brms* package (Bürkner, 2017, 2018) for all models, we report the estimate and its associated error mean \pm error and the 95% credibility interval (CI).

As no first-order 2AFC responses were provided in R-trials, we defined a proxy based on the percept associated with longer key presses during continuous report (i.e., covert first-order response). This allowed us to relate a proxy for first-order 2AFC responses and confidence ratings to compute metacognitive efficiency in CR+ trials.

Simulations for power estimations. We aimed at computing the power of our experimental design and analysis strategy. To do that, we estimated the proportion of simulated “experiments” in which we would have found a significant difference between two given conditions with different M-ratios. We used signal detection theory to simulate first-order and second-order responses from 80 trials for each of the 23 participants (see Fig. 4A). We set the distributions of the internal signals elicited by the stimuli to be a normal distribution with $\mu = \pm d'/2$, $\sigma = 1$ (the sign of μ depended on the longer stimulus presented). First-order responses were defined according to an optimal first-order criterion at 0.

Equation 1.1 describes the internal evidence e , as follows:

$$e = \begin{cases} \mathcal{N}(\frac{-1}{2}, \frac{1}{2}) & \text{if stimulus} = -1 \\ \mathcal{N}(\frac{1}{2}, \frac{1}{2}) & \text{if stimulus} = 1 \end{cases} \quad (1.1)$$

Equation 1.2 describes the perceptual decision d given the sampled evidence e , as follows:

$$d = \begin{cases} -1 & \text{if } e < 0 \\ 1 & \text{if } e \geq 0 \end{cases} \quad (1.2)$$

Next, to simulate the first-order proxy, we injected randomly distributed noise into the internal signal, sampled from a normal distribution centered at $\mu = 0$ and $\sigma = 0.8$.

Equation 2 describes our proxy for internal evidence e_{proxy} given the sampled internal evidence (e), as follows:

$$e_{\text{proxy}} = e + \mathcal{N}(0, 0.8). \quad (2)$$

This led to a correspondence of $\sim 70\%$ between real and proxy simulated responses, similar to our data. The rationale for adding noise to the internal signal rather than to the binary response variable itself was to preserve the structure of the data: trials with an internal signal closer to the decision boundary are associated with lower confidence and therefore are more likely to cross over the decision boundary as a consequence of adding noise, compared with trials with an internal signal strength that is far from the decision boundary. We obtained the simulated proxy by binarizing the noisy internal signal data based on the position relative to the same optimal first-order criterion placed at 0.

Finally, to simulate confidence ratings, we first added metacognitive noise by adding to the simulated internal signal an amount sampled from a normal distribution centered on 0 and with $m\sigma$, to achieve an M-ratio ranging from 0.1 to 4.

Equation 3 describes how we obtained degraded internal evidence e_{degraded}

$$e_{\text{degraded}} = e + \mathcal{N}(0, m\sigma). \quad (3)$$

Equation 4.1 describes how we assigned the absolute value of e_{degraded} to confidence if M-ratio < 1 , as follows:

$$\text{confidence} = |e_{\text{degraded}}|. \quad (4.1)$$

To simulate M-ratio values above 1, we then swapped the identity of the two distributions to make the second-order distributions sharper than the first-order ones. In a separate simulation, we established that these values of added σ corresponded to M-ratio values ranging between 0 and 1.1, which corresponds to the range of M-ratios in our experimental data (see Fig. 3). We set the simulated confidence as the absolute value of the internal signal; that is, the distance to the first-order decision criterion.

Equation 4.2 describes how we assigned the absolute value of e_{degraded} to the internal evidence if M-ratio > 1 . In this case, confidence was calculated as the absolute value of the internal evidence e , as follows:

$$e = e_{\text{degraded}}. \quad (4.2)$$

Thus, we added two kinds of noise to the original internal signals, with different meaning. The first type of noise simulated the imperfect relationship between covert/overt responses and their corresponding proxy. The second type of noise simulated the imperfect mapping between the strength of the internal signal at the point of the first-

order and second-order decisions, a relationship captured by M-ratio (Maniscalco and Lau, 2012).

We then submitted these simulated data (for 80 trials from 23 participants) to the same mixed-effects logistic regression we used to analyze empirical data. We repeated this procedure 250 times with each combination of M-ratio to estimate the number of times that a significant effect would occur in 250 experiments.

Data availability. The code described in the article is freely available online at https://gitlab.com/nfaivre/filevich_metareport/. The code is available as **Extended Data**. The preregistered analysis plan can be found at <https://osf.io/hnvsb/>. The raw data for analysis files used to reproduce all figures are available under https://gitlab.com/nfaivre/filevich_metareport/.

Results

Descriptive analyses: effects on confidence

The adaptive staircase procedures successfully fixed performance at ~71% correct, as follows: mean \pm SD accuracy was $72.0 \pm 4.2\%$ for continuous report and $72.1 \pm 4.6\%$ for no continuous report conditions, with no difference between conditions ($t_{(22)} = -0.12$, $p = 0.91$, $d = -0.02$, $BF_{10} = 0.22$, Table 1, a). Mean perceptual evidence did not differ across CR+R+ and CR-R+ trials ($t_{(22)} = 1.75$, $p = 0.09$, $d = 0.36$, $BF_{10} = 0.54$, Table 1, b), indicating that pairing motor information to the perceptual input was not informative for the first-order decision. Next, we tested for mean differences in confidence between all conditions using a linear mixed-effects regression model on confidence. The model included the two experimental manipulations (R and CR) and their interaction as fixed effects, intercepts for subjects as random effects, and a by-subject random slope for each of the factors. We found no interaction between overt first-order 2AFC responses and continuous report (mean = -0.02 ± 0.01 , evidence ratio = 0.10, Table 1, c), no strong effect of overt first-order 2AFC responses on mean confidence ratings (mean = 0.01 ± 0.02 , evidence ratio = 3.43), but a significant increase of mean confidence in conditions with continuous report (mean = 0.04 ± 0.02 , evidence ratio = 75.92, Table 1, d; Fig. 2A).

Importantly, to test the hypothesis that the monitoring of first-order 2AFC responses or their underlying processes contributed to confidence, we first established the existence of a relationship between reported confidence and first-order RT. We did so by fitting a mixed-effects linear regression to confidence in trials with overt first-order 2AFC responses (R+), including first-order accuracy, first-order RT, condition (CR+/CR-), and perceptual evidence as fixed effects, random intercepts for subjects, and by-subject random slopes for each fixed effect. As expected, we found a strong main effect of first-order 2AFC RT on confidence (mean = -0.15 ± 0.02 , evidence ratio > 4000, Table 1, e), confirming the relationship that has been reported in previous studies (Henmon, 1911; Vickers and Packer, 1982; Baranski and Petrusic, 1995; Patel et al., 2012; Fig. 2B). This effect was stronger for correct trials than for incorrect trials (interaction effect

estimate: mean = 0.04 ± 0.02 , evidence ratio = 46.06). We also found a main effect of accuracy (mean = -0.15 ± -0.03 , evidence ratio > 4000) and of perceptual evidence (mean = 0.22 ± 0.06 , evidence ratio > 4000), indicating that confidence was higher for correct responses, and in the presence of higher perceptual evidence. However, the model revealed no main effect of condition (mean = 0.01 ± 0.02 , evidence ratio = 0.41). No other model parameters were associated with confidence.

Together, these two analyses on mean confidence and the relationship between first-order 2AFC RT and confidence indicate that fast first-order responses were associated with higher confidence, but that response times are unlikely to play a causal role as removing first-order responses altogether had no effect on mean confidence.

Confirmatory analyses: effects on metacognitive sensitivity

Our first hypothesis was that sensorimotor activity related to first-order 2AFC responses carries information useful for confidence, above and beyond the strength of perceptual evidence. We therefore expected trials without first-order 2AFC responses to be associated with better metacognitive sensitivity (measured as the relationship between confidence and first-order accuracy) than those with responses to the 2AFC task. As we could not calculate response accuracy in trials with no first-order 2AFC responses (R-), we assumed that the percept associated with longer key presses during continuous report corresponded to the covert first-order 2AFC response. Hence, we limited this analysis to CR+ trials alone. In CR+R+ trials, this proxy based on continuous report predicted the actual first-order 2AFC response in $65.5 \pm 8\%$ of trials (ranging between 50% and 79.6%). For CR+R+ trials, we confirmed that response predictability based on the stimulus (i.e., the longest stimulus presented) and proxy (i.e., key pressed the longest) was significantly higher than that based on the stimulus alone (difference in Bayesian information criterion = 2.9, $\chi^2 = 10.04$, $p = 0.002$). In other words, the proxy, derived from the motor-tracking behavior, consistently added predictive power to the stimulus presented, which was already above chance (~71%) for the covert response provided in the first-order 2AFC task. This is why, despite low predictability scores, we proceeded with this analysis as per our preregistered plan and pursued alternative ways to analyze the data in the Exploratory analyses section described below.

We then compared metacognitive sensitivity between conditions with and without first-order 2AFC responses. It is only possible to estimate metacognitive sensitivity in R- trials if they are also CR+. In other words, we required the continuous report from CR+ conditions to estimate metacognitive sensitivity in cases of no first-order 2AFC response (R-). Therefore, we built a mixed-effects logistic regression for proxy accuracy that included condition (CR+R+/CR+R-) and confidence and their interaction as fixed effects, as well as subject-wise random intercepts, and random slopes for both confidence and condition. If metacognitive monitoring is affected by the presence of first-order 2AFC responses, this should

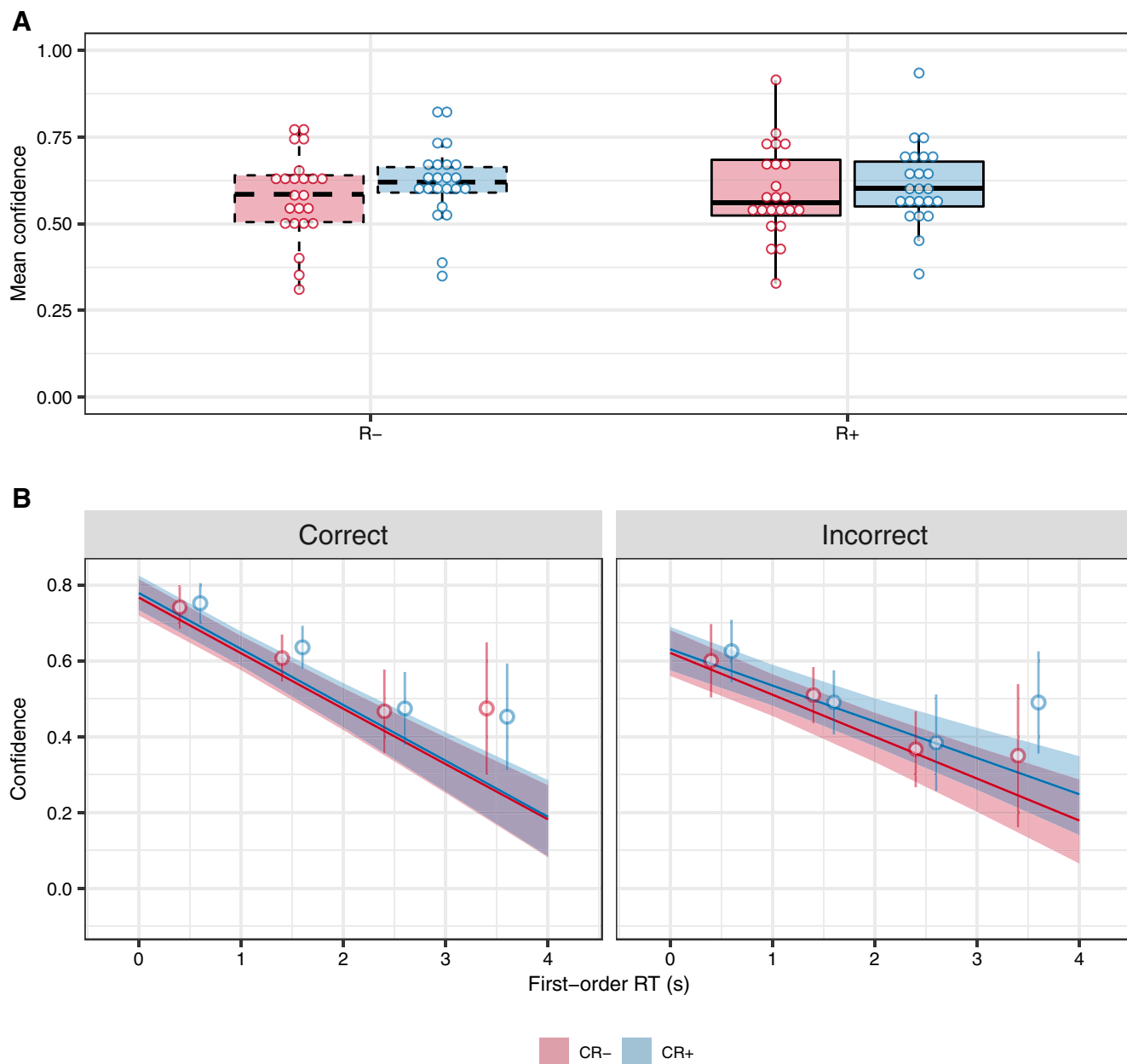


Figure 2. **A**, Differences in confidence judgments between conditions. A 2×2 ANOVA on mean confidence judgments revealed that trials with continuous report (CR+) were associated with higher confidence. **B**, Relationship between first-order reaction times and confidence judgments. Linear mixed-effects regressions revealed that, as expected, confidence judgments had a strong negative relationship with first-order reaction times on a trial-wise level. This relationship was present in all R+ trials (R- trials were not included in this analysis) but was stronger in the subset of correct trials. Regression lines and confidence intervals around them represent the model fit. The model took continuous reaction times as input. For illustrative purposes, we plot open circles and error bars that represent mean \pm 95% CI over participants after rounding reaction times and subtracting 0.5 s.

manifest as a significant interaction effect between confidence and the presence of a first-order response: the relationship (slope) between confidence and proxy accuracy should be stronger for trials with first-order responses than for those without them. Against our expectations, but in line with the results on mean confidence reported above, we found no interaction effect (mean = -0.11 ± 0.39 , evidence ratio = 1.57, Table 1, f). On the other hand, a main positive effect of confidence

(mean = 0.82 ± 0.32 , evidence ratio = 116.65, Table 1, g) indicated that the likelihood that the proxy was the correct answer increased with confidence and thus, simply put, that participants had some metacognitive access to their response accuracy.

The estimation of M-ratio (meta- d'/d') using the HMeta- d' toolbox (Fleming, 2017) revealed consistent results, as we found no differences between conditions in the M-ratio estimates (R+: M-ratio = 0.22, HDI = [0.12, 0.42], R-:

M-ratio = 0.25, HDI = [0.10, 0.45]; difference between conditions: highest-density interval (HDI) = [−1.42, 0.89], Table 1, h).

Our second preregistered hypothesis was that metacognitive performance between conditions with and without continuous report would differ because the key presses in the continuous report constitute an additional source of information for confidence responses. To test this hypothesis, we followed two approaches. First, using the same approach as above, we measured metacognitive sensitivity as the relationship between confidence and first-order 2AFC accuracy. Here again, a main effect of confidence on accuracy (mean = 2.51 ± 0.37 , evidence ratio > 4000) suggested that participants could monitor their performance. However, we found no interaction between confidence and condition (mean = 0.13 ± 0.35 , evidence ratio = 1.77), indicating that this effect was comparable with and without continuous report. This analysis included only trials with overt first-order 2AFC responses (R+), so it was possible to measure metacognitive accuracy with standard methods. Thus, we also estimated M-ratio (meta- d'/d') in trials with and without continuous report. Again, and consistent with our regression analyses, we found no differences between conditions in the M-ratio estimates (CR+: M-ratio = 1.06, HDI = [0.83, 1.32], CR−: M-ratio = 0.98, HDI = [1.24, 0.77]; difference between conditions: HDI = [−0.27, 0.40]).

Thus, our data revealed no differences in the relationship between confidence and first-order 2AFC accuracy between conditions.

To measure the effect of first-order responses (CR+R+ vs CR+R−), we relied on a proxy as the best informed guess for the covert first-order response; but the proxy was noisy and corresponded to the overt first-order response only for ~65% trials over all participants. In other words, with this analysis we injected noise into our first-order response, which might have in turn affected both the value of the confidence \times condition interaction estimates and our ability to find robust effects. To examine whether this was the case, and to what extent this affected our results, we simulated data from 250 “experiments” to compare the power of the logistic regression analysis based on the simulated first-order response and on the degraded first-order 2AFC proxy.

The results of these simulations (Fig. 4B–D) validated our analysis strategy. First, we found that power between the two analyses did not differ for values far from the diagonal (i.e., pairs of M-ratios with large differences between them). Second, and crucially, we found that even in regions where the proxy analysis fared worse (i.e., had lower power), power reductions in the range of 0.1–0.3, may partially, but not completely, explain our null results. This reduction in power reveals that we cannot strictly rule out that null effects in the proxy-based analyses are not due to low sensitivity. The loss of power is, however, intrinsic to a paradigm like ours, where the identity of covert responses is indirectly inferred and could only be avoided in principle if we had a lossless proxy with perfect accuracy. Interestingly, on the other hand, power estimations for the proxy-based analyses showed a somewhat

smoother pattern than those from actual responses. This result, presumably an effect of having an additional source of Gaussian noise, may be an unexpected advantage of the proxy-based analysis in preventing false inferences.

Effects of experimental manipulation

It has been recently shown that the variability in the stimuli presented may lead to inflated estimates of metacognitive performance (Rahnev and Fleming, 2019). To assess whether this was a problem in our data, we ran two separate analyses. We compared the range and SD of the stimuli presented to each participant in CR+ and CR− conditions (we did not compare R+ and R− conditions because these were yoked to their corresponding condition based on CR). In both cases, we found strong evidence for the null hypothesis (range of stimulus strengths presented: $\chi^2 = 0.1075$, $p = 0.743$, difference in Bayesian information criterion = −8.56, $BF_{10} = 0.014$; SD of stimulus strengths presented: $\chi^2 = 0.01$, $p = 0.752$, difference in Bayesian information criterion = −10.86, $BF_{10} = 0.004$), indicating that our estimates of metacognitive sensitivity are not inflated due to stimulus presentation.

Exploratory analysis: machine learning tools to predict first-order responses

We considered that the relatively low predictability of the continuous report-based proxy could be poor due to its simplicity: the proxy was based on nothing more than the longest reported percept in each CR+ trial. To extract as much information as possible from CR+ trials, we leveraged standard machine learning (ML) algorithms to predict first-order responses from CR+ information. First, for each CR+R+ trial we extracted features including the number of transitions in the key press response, the identity of the first and last stimuli shown and keys pressed, the total time with correct and incorrect key presses, and the delay between each stimulus presentation and the response. Using the scikit-learn module in Python (Pedregosa et al., 2011), we then trained three different classifiers on the data pooled over all participants using the following leave-one-out cross-validations: logistic regression, naive Bayes, and k-nearest neighbors. Their accuracy, based on the confusion matrix on CR+R+ trials revealed low overall predictability: 0.63, 0.61, and 0.64, respectively. These relatively low values are comparable to those of our simple proxy, and we therefore did not carry out any further analyses with the ML-based predictions.

Discussion

The past years have seen a growing interest in elucidating the sources of information that contribute to confidence judgments, as a window into potential computational processes that allow us to monitor our own thoughts. Converging evidence from very different experimental paradigms suggested that confidence is modulated by motor information concurrent with the first-order response (Vickers and Packer, 1982; Vickers et al., 1985; Baranski and Petrusic, 1995, 1998; Ratcliff and Starns, 2013; Fetsch et al., 2014; Kiani et al.,

2014; Moran et al., 2015; Zylberberg et al., 2016; for review, see Anzulewicz et al., 2019). Here, we set out to directly investigate this possibility. Concretely, we used a temporal-summation metacognitive task and asked whether committing a motor response associated with the response affected corresponding confidence judgments. Participants were instructed to rate their confidence in the accuracy of a binary decision both in trials with covert and overt first-order responses.

Effect of first-order responses on confidence ratings

As a precondition for our analyses, we first replicated what several studies had shown before (Fleming et al., 2010; Patel et al., 2012): in trials with overt first-order 2AFC responses (R+), reaction times to the first-order task showed a clear negative correlation with reported confidence. Based on these results alone, our data are in principle compatible with the hypothesis that first-order responses influence reported confidence. Crucially, we tested this hypothesis directly by comparing two conditions of the task that differed in whether participants had overtly responded with a key press to the first-order 2AFC task (CR+R+), or if their response remained covert (CR+R-). We first compared conditions in terms of average confidence judgments. Against our expectations, and despite the strong correlation between first-order reaction times and confidence, we found that absolute confidence judgments did not vary with the presence or absence of overt responses.

To further investigate the effects of overt responses, we then examined an important aspect of confidence judgments, namely their precision. That is, we considered that while participants may not have felt in general less confident in trials with covert responses, the quality of confidence judgments might have been degraded, resulting in a decrease in metacognitive performance relative to trials with overt responses. Measures of metacognition (metacognitive sensitivity, based on logistic regression and efficiency, based on M-ratio) rely on relating trial-wise confidence to accuracy. As the identity of covert responses remained latent by design, we inferred them relying on a proxy based on continuous reports (CR+). Concretely, we considered the percept with the longest key press as a proxy for both overt and covert first-order 2AFC responses. We then compared metacognitive sensitivity and efficiency based on the relationship between confidence and the proxy for responses. Here, mirroring the results from the analysis of absolute confidence values, we found no effect of overt first-order 2AFC responses. A concern with this analysis is that the proxy only corresponded to actual (overt) responses in an average of ~65% of trials, which resulted in a systematic underestimation of metacognitive performance (Fig. 3, compare dashed blue lines between panels). However, as the proxy predictive power does not differ across conditions, comparisons of metacognitive performance across conditions are still legitimate.

We note that in these analyses we assumed that participants had followed our instructions to rate confidence in the binary decision for both committed and omitted

responses. We discuss in the Limitations section the implications for our conclusions if participants did not follow these instructions.

Effect of continuous report on confidence ratings

In a factorial design, we also tested for the effect of continuous report paired to stimulus presentation on confidence judgments. Over conditions with and without first-order 2AFC responses (both R+ and R-), we found a consistent increase in confidence following continuous report (CR+ vs CR-) despite no changes in first-order performance. Previous studies have shown that different factors can affect first-order and second-order performance independently. These factors include experimental manipulations like changes in stimulus variability (Spence et al., 2016) or sensory reliability (Bang and Fleming, 2018), pharmacological silencing of different brain regions (Stolyarova et al., 2019), as well as the existence of sub-threshold motor activity (Gajdos et al., 2019), differences in movement parameters (Faivre et al., 2020), or voluntary control (Charles et al., 2020). Our study adds a novel kind of manipulation, namely, the occurrence of motor responses, to the list of experimental manipulations that affect confidence but not first-order accuracy. Alternatively, higher confidence ratings may result from criterion shifts. In fact, our model comparison showed that motor behavior could explain first-order 2AFC responses over and above perceptual evidence, suggesting that key presses in continuous report conditions were an additional source of information available for both the perceptual task (first-order) and the confidence task (second-order). With additional sources of information, participants may place their second-order criteria more liberally, resulting in higher confidence ratings.

Differences with the existing literature

To the best of our knowledge, this study is unique in that the effect of motor components on confidence was investigated by completely removing the first-order 2AFC response in some conditions, and replacing it instead with actions paired to the stimuli presentation. As a consequence, we never required participants to provide explicit responses in covert response conditions. Instead, we inferred them through participants' continuous report. Other studies have addressed the same question by using different experimental manipulations, which can be broadly grouped as following one of three approaches. A first set of studies has asked participants to rate the confidence of observed, rather than committed actions, by letting participants observe only first-order RTs in some of the experimental conditions (Patel et al., 2012; Vuillaume et al., 2019) or both RTs and stimuli (Pereira et al., 2020) before making the confidence judgment. A second group of studies has manipulated the perceptual evidence accumulation process (Fetsch et al., 2014; Kiani et al., 2014; Zylberberg et al., 2016). A third group of studies have instead manipulated the timing of the confidence judgment relative to that of the first-order response (Siedlecka et al., 2016; Wokke et al., 2020). Finally, a fourth approach

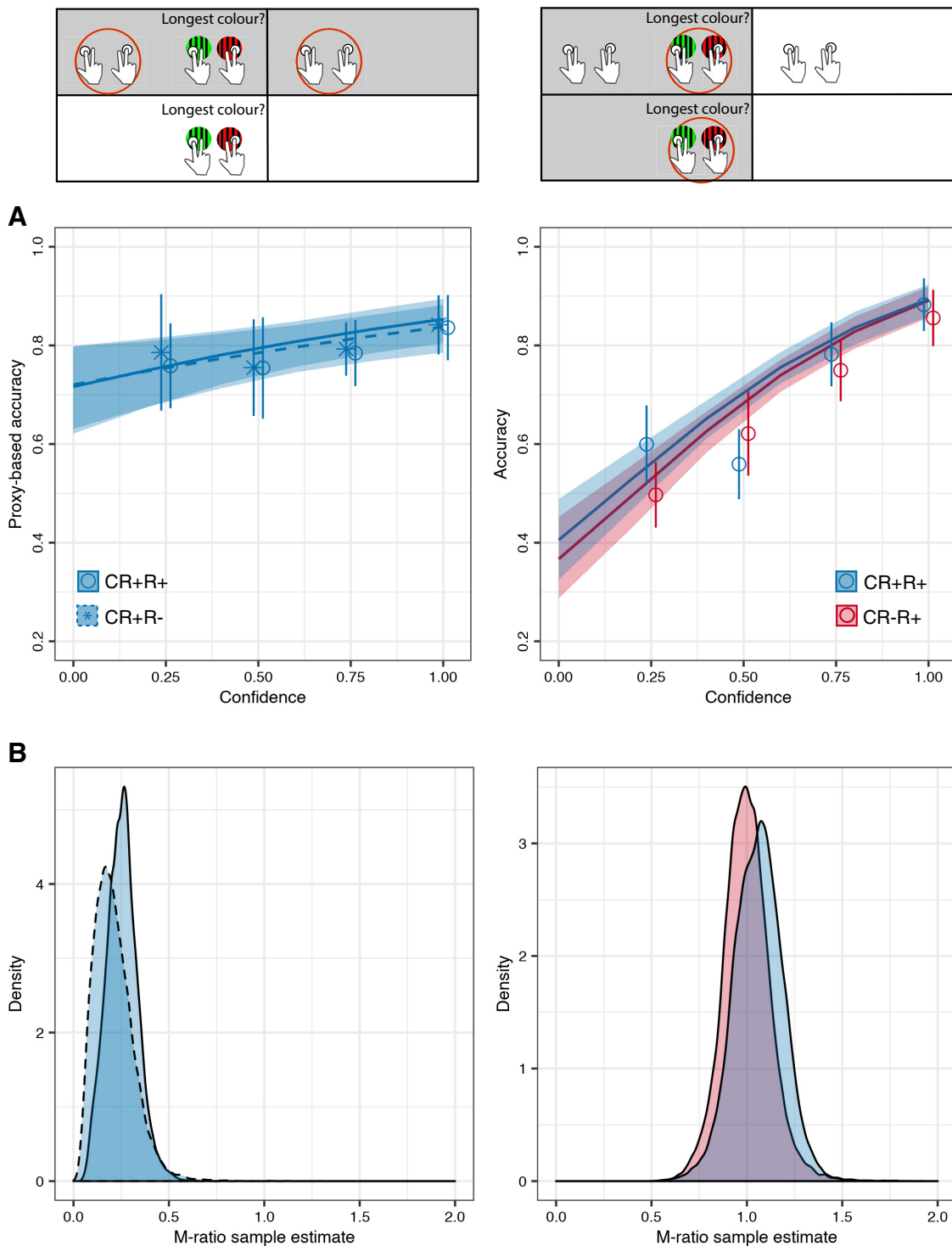


Figure 3. Differences in metacognitive performance between conditions. **A**, Metacognitive sensitivity quantified with a regression model on accuracy versus confidence. Estimated regression curves from the proxy for first-order 2AFC response (left) and overt first-order 2AFC response (right). The presence of a first-order 2AFC response did not affect the relationship between confidence and the first-order accuracy of the proxy. Open circles and error bars represent the mean \pm 95% CI over participants after rounding confidence ratings. **B**, Metacognitive efficiency quantified with M-ratio. As in **A**, we found no evidence that either giving an overt first-order response (left) or pairing an action to perceptual input (right) improved metacognitive efficiency. The insets above the panels highlight (in gray) which trials were used for each of the analyses.

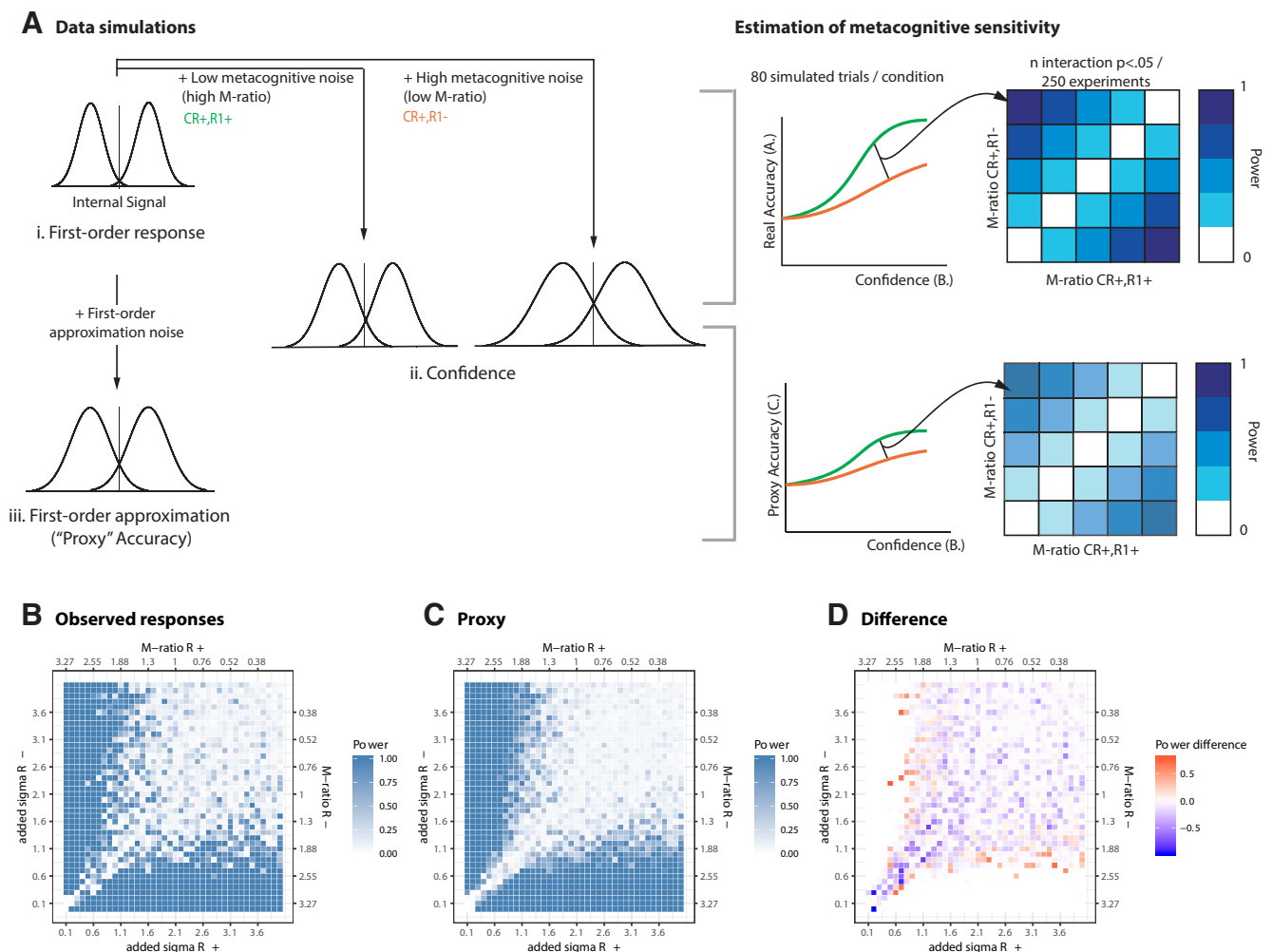


Figure 4. Power simulations. **A**, Data simulation strategy. We considered two conditions (in this case, CR+R+ and CR+R−) expected to differ in M-ratio. For each 1 of 250 experiments, we simulated 80 trials per condition, drawing three values for each of the “real” internal signal (top left row, i), a noisy confidence estimate (internal signal + metacognitive noise for each of two conditions; middle row, ii) and a value for the noisy proxy (bottom left row, iii). We fed the simulated trials into a logistic regression model, and determined the power of our analysis [i.e., the proportion of “experiments” in which the interaction term (representing a difference in metacognitive sensitivity between conditions) was significant (right)]. **B**, **C**, Results: power estimations for the analysis based on actual responses (**B**) and for proxy-based responses (**C**). **D**, The power difference between **B** and **C**. There are no differences in power when differences in M-ratios between two conditions are large (regions away from the diagonal), whereas there are small decreases in power for the proxy-based analysis for combinations of M-ratios that are closer to the diagonal.

consists in directly manipulating motor signaling either physiologically using TMS (Fleming et al., 2015) or behaviorally by instructions (Favre et al., 2018; Palser et al., 2018). Here, we followed the novel strategy of removing first-order responses and instead inferring them from stimulus-coupled responses. Against what has been reported in the literature and our expectations, we found that bypassing first-order 2AFC responses had no observable effect on metacognitive performance.

Our results also revealed that continuous motor responses contingent to perceptual evidence significantly increased confidence. A brief review of the literature reveals that motor activity impacts confidence biases and metacognitive performance distinctively, with large variations across experimental paradigms. On the one hand, our results are in line with what was reported by Gajdos et al. (2019), who

found that subthreshold motor activity before a decision increased confidence bias, with no impact of metacognitive performance. Other experimental manipulations produced the converse effect, namely, a modulation of metacognitive performance with no change in confidence bias. This includes the comparison of confidence in committed versus observed decisions (Pereira et al., 2020), and confidence under high or low sensorimotor conflicts (Favre et al., 2018). Using a similar design comparing prospective and retrospective confidence judgments, Siedlecka et al. (2016) found that both confidence bias and metacognitive performance increased in the presence of action-related signals. This set of mixed results questions the functional relevance of motor signals and suggests that the relationship might be more complex than previously thought. We speculate that the computation of confidence may be flexible and may

largely depend on the information that is globally available. In all previous studies, to the best of our knowledge, participants had access to some form of first-order reaction time information, at some point in time during the trial, as follows: either through observation from the third-person perspective, directly after the confidence report, or through simple access to reaction times produced under experimentally manipulated motor signals. In some conditions of our experiment, instead, responses were completely absent and may have shifted participants' global strategies for the computation of confidence. In other words, we contest that while first-order reaction time information is, under some experimental settings, used by participants to generate a confidence judgment, when motor information is not available at all, it may be replaced by other, equally precise sources of information, closer to the strength of evidence [e.g., the probability of being correct (Sanders et al., 2016), the internal signal noise (Navajas et al., 2017), and the evidence in favor of the chosen response alternative (Peters et al., 2017)]. This admittedly speculative account is in line with a previous study (Reyes and Sackur, 2014) showing that an introspective report in a visual search task (i.e., subjective reports about the number of items scanned, or the time required to scan them) may rely on different sources of information depending on the task context. This kind of introspective flexibility may explain our capacity to form confidence estimates about decisions that are not directly linked to a transient motor action, for instance, when controlling a brain machine interface (Schurger et al., 2017) or when making global confidence judgments in ecological contexts (Rouault et al., 2019).

Limitations and future directions

A limitation of our design lies in the capacity to identify covert first-order 2AFC responses from continuous reports. Voluntary key presses paired to the stimuli shown on the screen were a relatively poor predictor of first-order responses, and our simulations revealed that this led to lower statistical power in the proxy-based analyses, compared with analyses based on overt responses. However, we argue that the approach is promising given that future lines of research might take this first step further to develop "no-report" paradigms where covert decisions can be unequivocally inferred without a margin for error. Potential approaches include either eliciting an automatic response like the optokinetic nystagmus (Frässle et al., 2014), instead of a voluntary one like the key presses we used here; requiring voluntary key presses in highly trained participants, leading to low latencies between perception and response; or inferring responses through covert attention measured using steady-state visual evoked potentials (de Heering et al., 2019). Another limitation is the use of adaptive staircase procedures throughout the experiment. While maintaining task difficulty constant across trials, conditions, and participants is important to finely estimate metacognitive performance (Rahnev and Fleming, 2019), it also may hinder the relevance of sensorimotor signals as informative cues regarding the difficulty with which a decision was made (Kiani et al., 2014). Thus, a possibility is that sensorimotor

signals are more potent cues for confidence estimates under fluctuating task difficulty.

Finally, and importantly, we note that our interpretation of the results relies on the following two assumptions about trials without overt first-order 2AFC responses: first, we assume that participants committed to a binary decision on CR+R− trials, although they were not asked to overtly provide one; and, second, we assume that participants reported their confidence about the binary decision on both CR+R− and CR+R− trials. In other words, we assume that the only differences between CR+ and CR− conditions, and between R+ and R− conditions were the manipulations that we induced experimentally (continuous responses and first-order 2AFC responses, respectively), but that these differences had no impact on the cognitive processes that took place to produce the confidence judgments. If these assumptions were not met, our interpretation would not be valid.

If participants did not commit to a covert decision, despite our instructions and experimental design, this could imply that confidence ratings reflect different quantities in CR+R+ and CR+R− trials, making the comparison between conditions problematic. Specifically, while confidence in CR+R+ trials presumably reflects the probability that a binary decision was correct, given the external evidence (Pouget et al., 2016; Sanders et al., 2016; Fleming and Daw, 2017), it might reflect other quantities like the precision of the internal representation (Meyniel et al., 2015; Meyniel and Dehaene, 2017). Further, previous studies have shown that committing to a binary decision can affect the internal representation of the evidence, both at the first-order level (Stocker and Simoncelli, 2007; Luu and Stocker, 2018) and second-order level (Navajas et al., 2016, Peters et al., 2017). Because the internal representation is modified by a decision, confidence in two conditions that differ in whether a decision has been made may plausibly also differ in the biases and additional evidence accumulation taking place, making a direct comparison of confidence between R+ and R− trials potentially problematic.

Nevertheless, we have reasons to believe that our assumptions are, in fact, justified. First, by instruction, we asked participants to make a decision (and rate their confidence in its accuracy) even in cases where they were not prompted to explicitly provide the answer. Additionally, participants did not know beforehand whether, on each trial, they would have to provide a first-order response. Then, from the participants' point of view, R+ and R− trial types were indistinguishable until the point of stimulus offset.

Notwithstanding these limitations, we note that we found a clear null effect on absolute confidence differences between conditions with covert and overt first-order 2AFC responses. This result, which is not contaminated by imprecision in our identification of covert first-order 2AFC responses, more strongly argues for our interpretation that motor signals need not be used in metacognitive monitoring.

Conclusion

Identifying the sources of information that feed into confidence judgments is a core issue in metacognition

research. This study suggests that, while confidence judgments correlate with first-order reaction times, this relationship may be merely correlational, as removing the execution of first-order decisions altogether had no visible impact on confidence or metacognitive performance. By contrast, motor actions paired to stimulus presentation boosted confidence, but not metacognitive performance. These results, then, do not support the emerging idea that metacognition relies on the monitoring of sensorimotor signals and call for further research to find the underpinnings of metacognitive judgments.

References

- Anzulewicz A, Hobot J, Siedlecka M, Wierchoń M (2019) Bringing action into the picture. How action influences visual awareness. *Atten Percept Psychophys* 81:2171–2176.
- Bang D, Fleming SM (2018) Distinct encoding of decision confidence in human medial prefrontal cortex. *P Natl Acad Sci*, 115:6082–6087.
- Baranski JV, Petrusic WM (1995) On the calibration of knowledge and perception. *Can J Exp Psychol* 49:397–407.
- Baranski JV, Petrusic WM (1998) Probing the locus of confidence judgments: experiments on the time to determine confidence. *J Exp Psychol Hum Percept Perform* 24:929–945.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Bürkner P-C (2017) brms: an R package for Bayesian multilevel models using Stan. *J Stat Softw* 80:1–28.
- Bürkner P-C (2018) Advanced Bayesian multilevel modeling with the R package brms. *R J* 10:395–411.
- Charles L, Chardin C, Haggard P (2020) Evidence for metacognitive bias in perception of voluntary action. *Cognition* 194:104041.
- Faivre N, Filevich E, Solovey G, Kühn S, Blanke O (2018) Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. *J Neurosci* 38:263–277.
- Faivre N, Vuillaume L, Bernasconi F, Salomon R, Blanke O, Cleeremans A (2020) Sensorimotor conflicts alter metacognitive and action monitoring. *Cortex* 124:224–234.
- Fetsch CR, Kiani R, Newsome WT, Shadlen MN (2014) Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* 83:797–804.
- Fleming SM (2017) HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neurosci Conscious* 3:nix007.
- Fleming SM, Daw ND (2017) Self-evaluation of decision-making: a general Bayesian framework for metacognitive computation. *Psychol Rev* 124:91–114.
- Fleming SM, Dolan RJ (2012) The neural basis of metacognitive ability. *Philos Trans R Soc Lond B Biol Sci* 367:1338–1349.
- Fleming SM, Lau HC (2014) How to measure metacognition. *Front Hum Neurosci* 8:443.
- Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G (2010) Relating introspective accuracy to individual differences in brain structure. *Science* 329:1541–1543.
- Fleming SM, Maniscalco B, Ko Y, Amendi N, Ro T, Lau H (2015) Action-specific disruption of perceptual confidence. *Psychol Sci* 26:89–98.
- Frässle S, Sommer J, Jansen A, Naber M, Einhäuser W (2014) Binocular rivalry: frontal activity relates to introspection and action but not to perception. *J Neurosci* 34:1738–1747.
- Gajdos T, Fleming SM, Saez Garcia M, Weindel G, Davranche K (2019) Revealing subthreshold motor contributions to perceptual confidence. *Neurosci Conscious* 2019:niz001.
- de Heering A, Beauny A, Vuillaume L, Salvésen L, Cleeremans A (2019) SSVEP as a no-report paradigm to capture phenomenal experience of complex visual images. *bioRxiv*. doi: <https://doi.org/10.1101/588236>.
- Henmon VAC (1911) The relation of the time of a judgment to its accuracy. *Psychol Rev* 18:186–201.
- Kiani R, Corthell L, Shadlen MN (2014) Choice certainty is informed by both evidence and decision time. *Neuron* 84:1329–1342.
- Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in Psychtoolbox-3. *Perception* 36:1–1.
- Luu L, Stocker AA (2018) Post-decision biases reveal a self-consistency principle in perceptual inference. *Elife* 7:e33334.
- Maniscalco B, Lau H (2012) A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious Cogn* 21:422–430.
- Meyniel F, Dehaene S (2017) Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proc Natl Acad Sci U S A* 114:E3859–E3868.
- Meyniel F, Sigman M, Mainen ZF (2015) Confidence as Bayesian probability: from neural origins to behavior. *Neuron* 88:78–92.
- Moran R, Teodorescu AR, Usher M (2015) Post choice information integration as a causal determinant of confidence: novel data and a computational account. *Cogn Psychol* 78:99–147.
- Navajas J, Bahrami B, Latham PE (2016) Post-decisional accounts of biases in confidence. *Curr Opin Behav Sci* 11:55–60.
- Navajas J, Hindocha C, Foda H, Keramati M, Latham PE, Bahrami B (2017) The idiosyncratic nature of confidence. *Nat Hum Behav* 1:810–818.
- Palser ER, Fotopoulou A, Kilner JM (2018) Altering movement parameters disrupts metacognitive accuracy. *Conscious Cogn* 57:33–40.
- Patel D, Fleming SM, Kilner JM (2012) Inferring subjective states through the observation of actions. *Proc Biol Sci* 279:4853–4860.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Pereira M, Faivre N, Iturrate I, Wirthlin M, Serafini L, Martin S, Desvachez A, Blanke O, Ville DVD, Millan JDR (2020) Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proc Natl Acad Sci U S A* 117:8382–8390.
- Peters MAK, Thesen T, Ko YD, Maniscalco B, Carlson C, Davidson M, Doyle W, Kuzniecky R, Devinsky O, Halgren E, Lau H (2017) Perceptual confidence neglects decision-incongruent evidence in the brain. *Nat Hum Behav* 1:0139.
- Pleskac TJ, Busemeyer JR (2010) Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychol Rev* 117:864–901.
- Pouget A, Drugowitsch J, Kepecs A (2016) Confidence and certainty: distinct probabilistic quantities for different goals. *Nat Neurosci* 19:366–374.
- Rahnev D, Fleming SM (2019) How experimental procedures influence estimates of metacognitive ability. *Neurosci Conscious* 2019:niz009.
- Rahnev D, Desender K, Lee ALF, Adler WT, Aguilar-Lleyda D, Akdoğan B, Arbuzova P, Atlas L, Balci F, Bang J, Bègue I, Birney DP, Brady T, Calder-Travis J, Chetverikov A, Clark TK, Davranche K, Denison R, Dildine T, Double K, et al. (2019) The confidence database. *PsyArXiv*. doi: <https://doi.org/10.31234/osf.io/h8tju>.
- Ratcliff R, Starns JJ (2013) Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychol Rev* 120:697–719.
- Reyes G, Sackur J (2014) Introspection during visual search. *Conscious Cogn* 29:212–229.
- Rouault M, Dayan P, Fleming SM (2019) Forming global estimates of self-performance from local confidence. *Nat Commun* 10:1141.

- Sanders JI, Hangya B, Kepecs A (2016) Signatures of a statistical computation in the human sense of confidence. *Neuron* 90:499–506.
- Schurger A, Gale S, Gozel O, Blanke O (2017) Performance monitoring for brain-computer-interface actions. *Brain Cogn* 111:44–50.
- Siedlecka M, Paulewicz B, Wierzchoń M (2016) But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Front Psychol* 7:218.
- Spence ML, Dux PE, Arnold DH (2016) Computations underlying confidence in visual perception. *J Exp Psychol Hum Percept Perform* 42:671–682.
- Stocker AA, Simoncelli EP (2007) A Bayesian model of conditioned perception. *Adv Neural Inf Process Syst* 2007:1409–1416.
- Stolyarova A, Rakhshan M, Hart EE, O'Dell TJ, Peters MaK, Lau H, Soltani A, Izquierdo A (2019) Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. *Nat Commun* 10:1–14.
- Vickers D, Packer J (1982) Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychol (Amst)* 50:179–197.
- Vickers D, Burt J, Smith P, Brown M (1985) Experimental paradigms emphasising state or process limitations: I effects on speed-accuracy tradeoffs. *Acta Psychol (Amst)* 59:129–161.
- Vuillaume L, Martin J-R, Sackur J, Cleeremans A (2019) Comparing self-and hetero-metacognition in the absence of verbal communication. *bioRxiv*. doi: <https://doi.org/10.1101/585240>.
- Wokke ME, Achoui D, Cleeremans A (2020) Action information contributes to metacognitive decision-making. *Sci Rep* 10:3632.
- Zylberberg A, Fetsch CR, Shadlen MN (2016) The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *Elife* 5:e17688.