Metacognitive Awareness of Difficulty in Action Selection: The Role of the Cingulo-opercular Network

Kobe Desender[1,2*], Martyn Teuchies[1*], Carlos Gonzalez Garcia[1], Wouter De Baene[3], Jelle Demanet[1], and Marcel Brass[1,4]

Abstract

■ The question whether and how we are able to monitor our own cognitive states (metacognition) has been a matter of debate for decades. Do we have direct access to our cognitive processes, or can we only infer them indirectly based on their consequences? In the current study, we wanted to investigate the brain circuits that underlie the metacognitive experience of fluency in action selection. To manipulate action-selection fluency, we used a subliminal response priming paradigm. On each trial, both male and female human participants additionally engaged in the metacognitive process of rating how hard they felt it was to respond to the target stimulus. Despite having no conscious awareness of the prime, results showed that participants rated incompatible trials (during which subliminal primes interfered with the required response) to be more difficult than compatible trials (where primes facilitated the required response), reflecting metacognitive awareness of difficulty. This increased sense of subjective difficulty was mirrored by increased activity in the rostral cingulate zone and the anterior insula, two regions that are functionally closely connected. Importantly, this reflected activations that were unique to subjective difficulty ratings and were not explained by RTs or prime–response compatibility. We interpret these findings in light of a possible grounding of the metacognitive judgment of fluency in action selection in interoceptive signals resulting from increased effort. ■

INTRODUCTION

To what extent are humans able to monitor their own cognitive processes? Looking at the literature, there is some controversy surrounding this question. Some research suggests that humans are poor judges of their own cognitive processes (Johansson, Hall, Sikström, & Olsson, 2005; Wilson & Dunn, 2004; Nisbett & Wilson, 1977), whereas others have shown that humans are remarkably good at monitoring their own cognition, as participants are often aware when they made an error (Murphy, Robertson, Harty, & O'Connell, 2015), and can provide very precise estimates of the probability of being correct (Boldt & Yeung, 2015). This process of self-monitoring is also referred to as metacognition, as it describes insights into our own cognitive processes (Metcalfe & Shimamura, 1994; Brown, 1978). The neurocognitive mechanisms that underlie this self-monitoring ability are still poorly understood. Self-monitoring of behavior has been the focus of a number of recent brain imaging studies in the domain of error awareness and decision-making (Fleming, Weil, Nagy, Dolan, & Rees, 2010; Ullsperger, Harsay, Wessel, & Ridderinkhof, 2010; Klein et al., 2007). The anterior insula (AI) has been found to be involved in error awareness (Ullsperger et al., 2010; Klein et al., 2007), whereas anterior prefrontal regions were found to be involved in decision confidence (Fleming et al., 2010).

More recently, the process by which humans monitor the difficulty in action selection has attracted increasing attention (see, e.g., Questienne, van Dijck, & Gevers, 2018; Desender, Van Opstal, Hughes, & Van den Bussche, 2016; Desender, Van Opstal, & Van den Bussche, 2014). An interesting aspect of action selection is that one can manipulate its difficulty without participants becoming aware of the manipulation. It is known that subliminal response conflict hampers performance: It slows down response speed and increases error rates (see, e.g., Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003). By using a subliminal response priming paradigm, one can thus manipulate conflict between two response options outside participants' awareness. Consequently, a metacognitive representation of this manipulation cannot be based on a conscious interpretation of the events (i.e., the visually conflicting information) but has to be based on the interpretation of internal signals caused by these events. A recent brain imaging study revealed that subliminal response conflict is registered in the brain by increased activity in both the rostral cingulate zone (RCZ) and the left AI (Teuchies et al., 2016). Although participants are typically unaware of the subliminal conflict-inducing stimulus, previous research has shown that they nevertheless report increased

[1]Ghent University, [2]KU Leuven, [3]Tilburg University, [4]Humboldt Universtät zu Berlin
*Shared first coauthorship.

levels of subjective difficulty when responding to trials with subliminal conflict (Desender et al., 2014, 2016). This raises the question about the source of such metacognitive judgments. A previous study indicates that the experience of subjective difficulty does not simply reflect a read-out of response speed (Desender et al., 2016) but rather seems to be based on motor conflict induced by the subliminal primes (Questienne et al., 2018). This leads to the prediction that the metacognitive judgment is related to brain processes that are involved in subliminal conflict processing itself, namely, the RCZ and the AI (Teuchies et al., 2016), independently from RTs and prime–response compatibility.

METHODS

Participants

Participants in this study were 30 Dutch-speaking students from Ghent University (19 female; mean age = 23.77 years, *SD* = 3.20); each one reported to be healthy and with no history of neurological, pain, or circulatory disorders and normal or corrected-to-normal vision. One participant was removed because of excessive head motion. All participants gave written informed consent, and the study was approved by the medical ethical review board of the Ghent University Hospital, in accordance with the Declaration of Helsinki. All participants were right-handed, as assessed by the Edinburgh Inventory (Oldfield, 1971), and were compensated 30 Euros for their participation.

Stimuli

Stimulus presentation and response registration were done using Tscope software (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006). In the scanner, the task was presented using a Brainlogics 200MR digital projector that uses digital light processing running at a refresh rate of 60 Hz with a viewing distance of 120 cm. Using digital light processing, it took 1 msec to deconstruct the image on the screen allowing our subliminal primes to be presented with great precision. The mean presentation time was 18.00 msec (*SD* = 0.24, range = 15.91–18.91 msec). Three types of gray-colored primes were used (Figure 1): left- or right-pointing arrows or a neutral prime (which consisted of overlapping left- and right-pointing arrows). The primes were followed by superimposed metacontrast masks of the same luminance. The metacontrast masks were embedded within target arrows that pointed left or right. Primes subtended visual angles of 0.8° × 1.86°; and the targets, 1.09° × 3.47°. Prime and target stimuli could appear randomly above or below a fixation cross at a visual angle of 1.38°. The unpredictable location was included to enhance the masking effect (Vorberg et al., 2003). A circular rating scale was adapted from Kahnt, Heinzle, Park, and Haynes (2011). The x and y coordinates of the mouse response were converted into polar coordinates ranging from 0° (easiest) to 360° (most difficult). The thickness of the scale increased with difficulty. The easiest point on the scale was the tail of the circle; the most difficult point was the thickest point of the circle. The orientation of the scale was randomly chosen on each trial so that the starting point of the scale was unpredictable. This prevented participants from preparing a motor response before seeing the actual scale.

Procedure

Except for the ratings, the experimental design was identical to Teuchies et al. (2016). Primes were presented for 16.7 msec (1 refresh rate at 60 Hz), followed by a blank

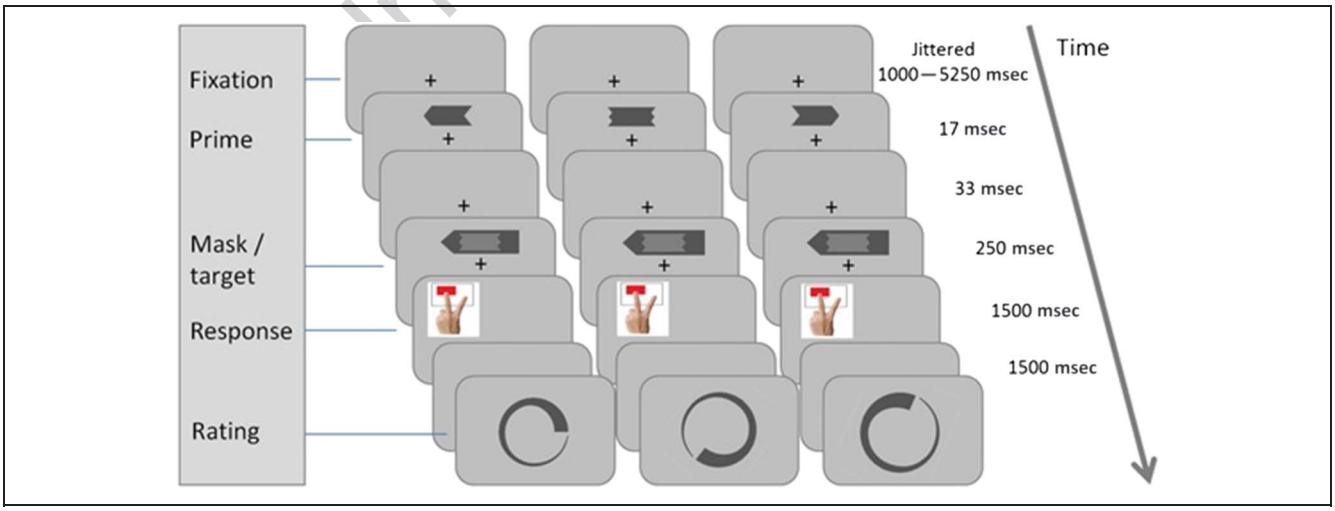

Figure 1. Schematic of an experimental trial. Three possible combinations of the factor prime–response compatibility (compatible: left; neutral: center; incompatible: right). Participants were instructed to respond to the target stimuli (with the left hand) and were unaware of the presence of the arrow primes. Primes and targets could appear randomly above or below fixation on each trial. After their response, participants indicated their subjective feeling of difficulty using a circular rating scale. The thin tail is the easiest point, and the scale continuously increases in thickness and difficulty up to the thick end representing the most difficult point. Participants were instructed to use the whole scale.

screen for 33.3 msec and a target that also functioned as a mask. Target duration was 250 msec. The response window was set to 1500 msec. Participants were instructed to respond as fast and accurate as possible to the direction of the target arrows with their left middle finger (left-pointing targets) and left index finger (right-pointing targets) using an MR-compatible response box. If participants failed to respond within this time window, they saw "te laat" (too late) for 1000 msec after the trial. After each response, a blank screen was shown for 1500 msec followed by the rating part of the trial during which the rating scale was shown until participants had given their response with their right hand using an MR-compatible optical trackball mouse to select a point on the rating scale that matched their subjective sense of difficulty. The response was registered only when the mouse was actually on the rating scale. Mouse clicks outside the rating scale were not registered. Participants were instructed to use the entire scale and were informed that the extremities of the scale represented their personal most difficult and easiest points. Once they clicked on the scale, a blank screen was shown for the intertrial interval. The intertrial interval was jittered with values ranging between 1000 and 5250 msec. The jitter values followed a distribution with pseudologarithmic density (range = 1000–5250 msec, in steps of 250 msec; mean jitter = 2625 msec).

Before doing the experiment in the scanner, participants carried out two training blocks of 48 trials each. In the first training block, they were only presented with the response priming task, without the rating to let them experience the response priming task. When asked, all participants indicated that they made mistakes and that some trials felt more difficult than others. In the second training block, the rating was added after every individual trial, and participants were instructed to rate on each trial how difficult they found it to respond as fast and accurate as possible to the target stimulus. Participants were never alerted to the possibility of primes being presented. The main task inside the MRI scanner consisted of three blocks of 72 trials each. Within each block, each prime–response compatibility condition (compatible, incompatible, and neutral) occurred equally often. At the end of the task, participants were asked whether they noticed anything unusual about the stimuli during the task. None of the participants noticed the primes, but three of them reported seeing a "flash" before the target was presented. These participants were included in the final sample. After the test phase, participants were explicitly told about the presence of the primes and performed a prime-visibility test. This test allowed us to check if the prime stimuli were indeed presented subliminally or not. In this test, participants were asked to identify the direction of the primes (left or right) on each individual trial by using the same left and right response buttons as used during the test phase. During this test, participants remained in the scanner, so environment and apparatus were identical to the main experiment. To minimize indirect priming effects on the recognition of the primes, participants were required to respond at least 600 msec after the mask was presented. A visual cue ("*") signaled when they were allowed to respond. The test consisted of two blocks of 50 trials each.

fMRI Data Acquisition and Preprocessing

Data were acquired with a 3-T Siemens Magnetom Trio MRI system (Siemens Medical Systems) using a 32-channel radiofrequency head coil. Participants were positioned head first and supine in the magnet bore. First, 176 high-resolution anatomical images were acquired using a T1-weighted 3-D magnetization prepared rapid gradient echo sequence (repetition time = 2250 msec, echo time = 4.18 msec, inversion time = 900 msec, image matrix = 256 × 256, field of view = 256 mm, flip angle = 9°, and voxel size = 1 × 1 × 1 mm). Whole-brain functional images were then collected using a T2-weighted EPI sequence, sensitive to BOLD contrast (repetition time = 2000 msec, echo time = 35 msec, image matrix = 64 × 64, field of view = 224 mm, flip angle = 80°, slice thickness = 3.0 mm, distance factor = 17%, voxel size = 3.5 × 3.5 × 3.0 mm, and 30 axial slices). A varying number of images were acquired per run because of individual differences in choice behavior and RTs. All data were preprocessed and analyzed using MATLAB and the SPM8 software (Wellcome Department of Cognitive Neurology). To account for possible T1 relaxation effects, the first four scans of each EPI series were excluded from the analysis. The ArtRepair toolbox for SPM was used to detect outlier volumes concerning global intensity or large scan-to-scan movement (Mazaika, Whitfield-Gabrieli, & Reiss, 2007). First, a mean image for all scan volumes was created, to which individual volumes were spatially realigned using rigid body transformation. Thereafter, they were slice time corrected using the first slice as a reference. The structural image of each participant was coregistered with their mean functional image, after which all functional images were normalized to the Montreal Neurological Institute (MNI) T1 template. Motion parameters were estimated for each session separately. The images were resampled into 3 × 3 × 3 mm voxels and spatially smoothed with a Gaussian kernel of 8 mm (FWHM). A high-pass filter of 128 Hz was applied during fMRI data analysis to remove low-frequency drifts.

Behavioral Data Analysis

Mean RTs, error rates, and subjective ratings were submitted to a repeated-measures ANOVA, with prime–response compatibility (compatible vs. incompatible vs. neutral) as a factor. The responses to the primes in the visibility check were categorized using signal detection theory (Green & Swets, 1966). Measures of prime discriminability (d') for each participant were computed. We then used a

one-sample t test to see whether the mean d' of the sample deviated from zero.

ROI Analyses

In the ROI analyses, we focused on the RCZ and the AI as these were our principal ROIs based on our previous study (Teuchies et al., 2016). Accordingly, the peak coordinates were taken from this previous study. To create ROIs, we created spheres with a 5-mm radius around the peak coordinates of the RCZ (MNI: 6, 20, 43) and the AI (MNI: −36, 20, –2). We then extracted single-trial beta estimates using a general linear model (GLM) approach, in which each trial was modeled as one regressor. We used linear mixed models, as implemented in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015), to analyze the relationship between difficulty ratings and brain activity. Using this type of analysis, both variables can be fit at the single-trial level. Random slopes were added for each variable when this increased the model fit, as assessed by model comparison. For these models, F statistics are reported and the degrees of freedom were estimated by Satterthwaite's approximation, as implemented in the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2014). Finally, for each model, we checked and confirmed that the variance inflation factors were below 3.

GLM Analyses

The participant-level statistical analyses were performed using the GLM. In a first analysis, compatibility conditions (compatible, incompatible, and neutral) were modeled in a single regressor of interest, and raw subjective rating values for each trial were added as an extra parameter allowing us to look at brain activity related to the raw subjective difficulty ratings. In a second analysis, we wanted to capture variance in brain activity that was unique to the subjective difficulty ratings, independent of the variables prime–response compatibility and RT (which are both known to affect subjective difficulty ratings; Questienne et al., 2018; Desender, Buc Calderon, Van Opstal, & Van den Bussche, 2017). To capture variance related to compatibility, three different regressors of interest (compatible/incompatible/neutral) were modeled for this variable. To look at brain activation uniquely attributed to subjective ratings independent of RTs (i.e., both variables showed modest negative relation: mean $r = -.30$, $SD = .16$, range $= -.545$ to .011), we introduced them both as parametric modulators. Because the order of the parametric regressors matters (i.e., the second regressor will only capture variance that has not been captured yet), we first entered RT as a parametric regressor and subjective rating as the second parametric regressor.

In both analyses, erroneous trials and the first trials of each block were always modeled as separate regressors of no interest (4.9% of the trials). The events of interest were the periods after the onsets of the different targets in the response priming task. Vectors containing the event onsets were convolved with the canonical hemodynamic response function to form the main regressors in the design matrix (the regression model). Motion parameters for each individual participant were added. No derivatives were added to the model for this analysis. The statistical parameter estimates were computed separately for each voxel for all columns in the design matrix. Contrast images were constructed for each individual to compare the relevant parameter estimates for the regressors containing the canonical hemodynamic response function. The group-level random effects analysis was then performed. Using one-sample t tests, we looked at the effects of the subjective difficulty ratings and RTs across prime–response compatibility conditions. The subjective difficulty ratings and the RTs had been added as parametric regressors during the first-level analysis. To correct for multiple comparisons, first we identified individual voxels that passed a "height" threshold of $p < .001$, and then, the minimum cluster size was set to the number of voxels corresponding to $p < .05$, FWE corrected. This combination of thresholds has been shown to control appropriately for false positives (Eklund, Nichols, & Knutsson, 2016). The resulting maps were overlaid onto a structural image of a standard MNI brain, and the coordinates reported correspond to the MNI coordinate system.

Mediation Analyses

As described below, activity in both the RCZ and the AI was related to difficulty ratings. To shed light on the direction of these effects, we performed post hoc mediation analyses. For this analysis, we created a new GLM in which all the trials were entered as separate regressors, so we obtained brain activation for the RCZ and the AI on a trial-by-trial basis. Our main question was whether the influence of the RCZ on difficulty ratings was mediated by the AI, conditional on congruency. A mediator and an outcome model were fitted on the data using mixed regression modeling, using the same model building strategy as reported above. A mediator mixed model was fit in which activity in the AI was predicted by activity in the RCZ, RTs, and compatibility. An outcome mixed model was fit in which difficulty ratings were predicted by activity in the AI, the RCZ, RTs, and compatibility. A mediation analysis was then performed on these two models (using the mediation package; Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). This method partitions the total effect on ratings into an indirect effect (i.e., the effect of RCZ on ratings that is mediated by the AI) and a direct effect (i.e., correlation between RCZ and ratings that is not explained by the AI), conditional on RTs and compatibility. If this indirect effect is significant, this is evidence for a significant mediation effect. Note that this latter observation is equivalent to showing that the

influence of a direct path decreases when a mediation path is added to the model. Second, we tested the reversed hypothesis that the influence of AI on difficulty ratings was mediated by the RCZ. For this, the same mediation analysis was run but after exchanging RCZ and AI in all models.

RESULTS

Behavioral Results

Main Task

Trials where participants did not respond within the 1500-msec response window were removed from the data (0.6% of the trials). For the remaining data, mean RTs on correct trials, mean error rates, and mean difficulty judgments on correct trials were submitted to separate repeated-measures ANOVAs with prime–response compatibility (prime–response compatible vs. incompatible vs. neutral) as a factor. For RTs (Table 1), this analysis yielded a significant effect of prime–response compatibility, $F(2, 28) = 39.24$, $p < .001$, $\eta_p^2 = .737$. Prime-compatible responses ($M = 426.8$ msec) were significantly faster than prime-incompatible responses ($M = 453.9$ msec; incompatible − compatible = 27 msec), $t(29) = 7.94, p < .001, d = 0.68$. Prime-compatible responses were not faster than prime-neutral responses ($M = 430.6$ msec; neutral − compatible = 7 msec), $t(29) = -1.27, p = .22, d = 0.11$, meaning that directional primes did not lead to a significant facilitation effect. There was, however, a significant interference effect, meaning that prime-incompatible responses were slower than responses to neutral primes (incompatible − neutral = 23 msec), $t(29) = 7.97$, $p < .001$, $d = 0.62$.

The error rates showed a similar effect of prime–response compatibility, $F(2, 28) = 12.53$, $p < .001$, $\eta_p^2 = .472$. Participants made significantly more errors on prime-incompatible trials ($M = 7.93\%$) than on prime-compatible trials ($M = 2.92\%$), $t(29) = 5.1$, $p < .001$, $d = 0.93$, and on neutral trials ($M = 3.84\%$), $t(29) = 4.3$, $p < .001$, $d = 0.73$. Error rates were also slightly higher on neutral-prime trials than on prime-compatible trials, but this difference was not significant, $t(29) = -1.7$, $p = .1$, $d = 0.28$.

For the subjective difficulty ratings, we also observed a main effect of prime–response compatibility, $F(2, 28) = 9.60$, $p < .001$, $\eta_p^2 = .407$. Because of the circular nature of the scale, ratings lie between 0° (easy) and 360° (difficult). Participants rated prime-incompatible trials ($M = 156.6$) as significantly more difficult than prime-compatible trials ($M = 146.3$), $t(29) = 4.1, p < .001, d = 0.20$, and more difficult than neutral trials ($M = 145.8$), $t(29) = -4.3, p < .001, d = 0.21$. Ratings for neutral-prime trials did not differ from ratings for prime-compatible trials, $t(29) = -0.29, p = .77, d = 0.01$.

Table 1. RTs, Percentage of Errors, and Difficulty Ratings as a Function of Prime–Action Compatibility

	RT (msec)	*Errors (%)*	*Subjective Difficulty Rating (°)*
Compatible	426.8 (8.1)	2.92 (0.6)	146.3 (9.5)
Incompatible	453.9 (6.8)	7.93 (1.3)	156.6 (9.5)
Neutral	430.6 (6.7)	3.84 (0.6)	145.8 (9.4)

Numbers in parentheses show standard errors of the means across participants. The subjective difficulty ratings are reported in degrees ranging from 0 to 359.

Prime Visibility

On the basis of the data of the prime visibility task, a d' value was computed for each participant as an index of prime visibility. The d' values were not significantly different from chance-level performance (i.e., zero; mean $d' = 0.077$, $SD = 0.37$; one-sample t test, $t(29) = 1.13, p = .27$). Thus, it can be concluded that participants show no reliable sign of awareness of the direction of the prime stimuli. Furthermore, when correlating the compatibility effect in the subjective ratings with the individual d' values, we found no significant correlation, $r(28) = .12, p = .54$, indicating that the subjective ratings were not influenced by prime visibility.

fMRI

ROI Analysis Results

In our previous study, which was identical to the current work except that we did not query subjective difficulty, we observed that the RCZ and the AI both were sensitive to conflicts in response selection (Teuchies et al., 2016). Therefore, in a first set of analyses, we focused specifically on these two brain regions. To do so, we extracted single-trial beta estimates from the RCZ (MNI: 6, 20, 43) and the AI (MNI: −36, 20, –1), both defined a priori based on our previous study. We then used linear mixed models to examine whether these regions are sensitive to differences in subjective difficulty.

An analysis predicting activity in RCZ by difficulty ratings showed a significant effect of difficulty ratings, $F(1, 28.23) = 17.71$, $p < .001$. As can be seen in Figure 2A, the easier a trial was judged to be, the lower the activity in RCZ. A similar analysis predicting activity in the AI by difficulty ratings also showed a significant effect, $F(1, 28.29) = 35.87$, $p < .001$. As can be seen in Figure 2B, increased subjective difficulty (i.e., higher ratings) was associated with enhanced activity in AI.

Although this analysis is an important first step, it cannot unravel whether activity in the RCZ and AI reflects actual variation in difficulty ratings or whether both are driven

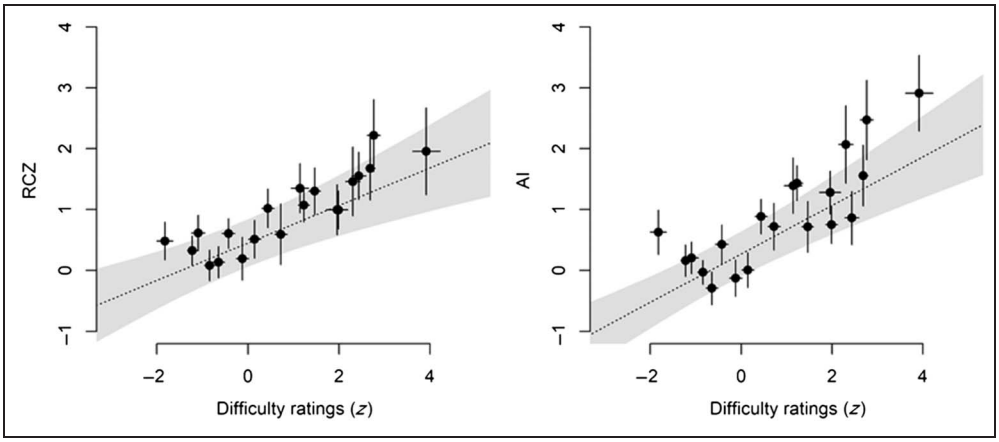

Figure 2. Relation between subjective difficulty ratings and activity in RCZ (A) and AI (B). Dots show the average neural activity in the specified ROI as a function of average difficulty rating, divided in 20 equal-sized bins; dotted lines show the fixed effects slope from the mixed model fit; and errors bars and shades reflect standard errors.

by another variable that could in principle covariate with difficulty (such as prime–response compatibility or RT; Desender et al., 2017). Therefore, in a second set of analyses, we looked at whether both these brain regions showed activity uniquely correlated with subjective ratings, that is, after controlling for the variables compatibility and RTs. To do so, we extended the mixed regression models reported before and included compatibility and RTs.

First, we report the results of a model predicting single-trial RCZ activity by subjective difficulty ratings, compatibility (three levels: congruent, neutral, or incongruent), RTs, and the interaction between compatibility and RTs. Replicating our previous work, there was a main effect of compatibility, $F(2, 5814.2) = 4.508$, $p = .011$. This main effect reflected that RCZ activity was higher on incongruent trials than on congruent trials, $z = 2.45$, $p = .014$, and on neutral trials, $z = 2.75$, $p = .006$, whereas congruent and neutral trials did not differ, $p > .757$. We also observed a significant main effect of RTs, $F(1, 33.4) = 27.02$, $p < .001$, reflecting increased RCZ activity for trials with slower RTs. Critically, even after controlling for both these factors, we still observed a significant main effect of difficulty ratings, $F(1, 27.9) = 9.80$, $p = .004$. The interaction between compatibility and difficulty ratings did not reach significance, $p = .072$.

Second, highly similar results were found in a model predicting single-trial AI activity by difficulty ratings, compatibility (three levels: congruent, neutral, or incongruent), RTs and the interaction between compatibility and RTs. Also here, we observed a main effect of compatibility, $F(2, 5825.7) = 5.11$, $p = .006$, reflecting that activity in AI was higher on incongruent trials than on congruent trials, $z = 2.13$, $p = .033$, and neutral trials, $z = 3.14$, $p = .001$, whereas congruent and neutral trials did not differ, $p > .302$. We also observed a significant main effect of RTs, $F(1, 5654.4) = 43.97$, $p < .001$, reflecting increased AI activity for trials with slower RTs. Critically, even after controlling for both these factors, we still observed a significant main effect of difficulty ratings, $F(1, 31.3) = 18.91$, $p < .001$. The interaction between difficulty ratings and congruency did not reach significance, $p = .085$.

Whole-Brain Analysis Results

In the ROI analyses, we found that difficulty ratings significantly predicted activity in the RCZ and AI even when controlling for RT and compatibility. To corroborate these findings, we next performed two whole-brain univariate analyses. We first looked for brain regions where activation magnitude was correlated with subjective ratings. This analysis revealed a large set of regions with significant activation, including the RCZ and insula (see Figure 3). When using a more conservative threshold, clusters surviving correction were located in the RCZ (MNI: 3, 23, 55), right insula (MNI: 51, 17, 4), and left insula (MNI: −36, 23, −2). This first whole-brain analysis corroborates the

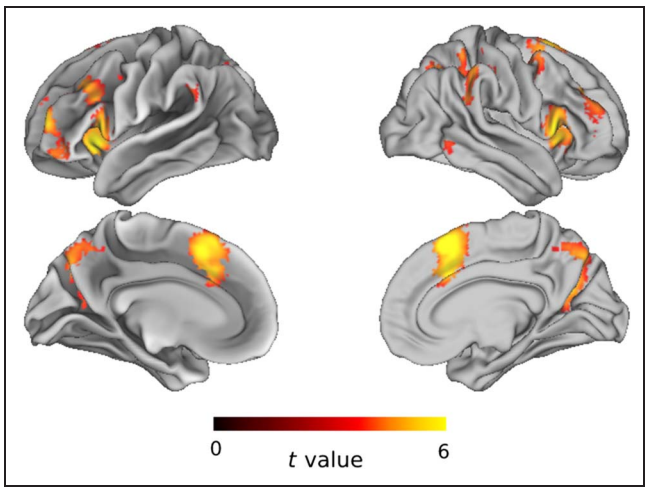

Figure 3. GLM contrast for the effect of subjective difficulty. Warm colors show regions where activation magnitude is correlated with difficulty ratings (primary voxel threshold [$p < .001$ uncorrected] and cluster-defining threshold [FWE $p < .05$]).

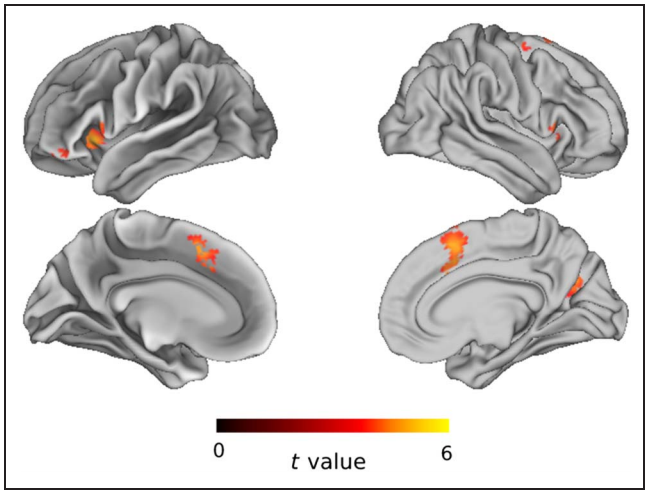

Figure 4. Main areas of interest showing higher activation when the sense of subjective difficulty increased, independent from prime–response compatibility and after regressing out the effect of RTs. Warm colors show regions where activation magnitude is correlated with difficulty ratings (primary voxel threshold [$p < .001$ uncorrected] and cluster-defining threshold [FWE $p < .05$]).

previous analyses showing that increased activity in the RCZ and insula is related to increased subjective difficulty.

Similar to the ROI analyses, we then looked for brain regions that showed activity uniquely correlated with subjective ratings. To do so, we performed a GLM analysis with one regressor for each compatibility level (compatible, incompatible, neutral) and two parametric modulators: (1) RT and (2) subjective ratings. Importantly, the order of the parametric modulators matters, as the second one will only operate over the variance not explained by the first modulator. Therefore, by looking at activity correlated with subjective ratings, we were able to ascertain what brain regions code for subjective ratings, while controlling for RTs and compatibility. Results from this model showed that the parameter of subjective rating residuals revealed again a set of frontoparietal regions (see Figure 4), which included significant clusters of FWE-corrected activation in the RCZ (MNI: −3, 14, 52) and the left (MNI: −33, 26, 5) and right (MNI: 57, 23, 7) AI. The left AI cluster is closely located to the AI (MNI: −36, 20, –2) that we observed in our previous study (Teuchies et al., 2016). These results indicate that the RCZ and the AI showed increased activity with increased subjective difficulty, independent from prime–response compatibility or RTs. Note that we did not observe activation near these two regions when performing the same analysis locked to the rating scale, suggesting that the observed activity indeed reflects the processing of conflict in action selection and not an evaluation or decision process.

Mediation Analysis

To shed light on the directionality between both identified brain regions, we next performed causal mediation analysis. First, we tested the hypothesis that the influence of the RCZ on ratings is mediated by the AI. A prerequisite for mediation analysis is that all three paths are significant. In the mediator model, the RCZ predicted activity in the AI, $F(1, 28.4) = 511.08$, $p < .001$, and in the outcome model, subjective ratings were predicted by the AI, $F(1, 5813) = 11.05$, $p < .001$, and RCZ, $F(1, 5814.1) = 3.011$, $p = .083$, although the latter failed to reach significance. Results of the mediation analyses showed that, conditional on compatibility and RTs, there was a significant part of the influence of the RCZ on subjective ratings that was mediated by activity in the AI, $\beta = .451$, $p < .001$, whereas there was no direct effect of the RCZ on subjective rating, $\beta = .404$, $p = .078$. These results are shown in Figure 5. Note that these results should be interpreted with caution, given that we did not observe a significant effect of RCZ in the outcome model. Because mediation analysis is a correlational technique, we also tested the reversed causal flow, namely, that the influence of the AI on ratings is mediated by the RCZ. This analysis showed a highly significant direct effect of AI on ratings, $\beta = .701$, $p < .001$, and no mediation effect by the RCZ, $\beta = .209$, $p = .099$. In summary, results from the mediation analyses suggest that the influence of the RCZ on ratings is mediated by activity in the AI.

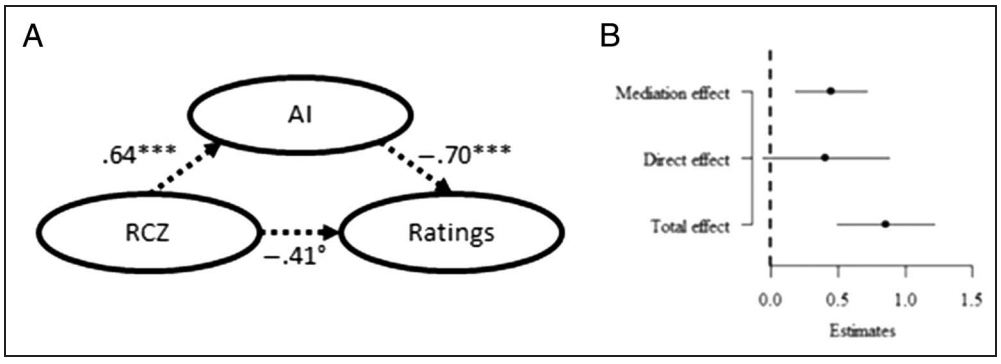

Figure 5. Results of the mediation analysis testing whether the influence of RCZ on ratings is mediated by the AI. Unstandardized regression coefficients are shown from the mediation and the outcome mixed regression models. Degrees of freedom for calculating significance were based on Satterthwaite's approximation. Effect sizes from the causal mediation analysis are shown in B. Error bars reflect quasi-Bayesian 95% confidence intervals. ***$p < .001$, °$p < .1$.

DISCUSSION

In the current study, we set out to test which brain regions are involved in self-monitoring difficulty in action selection. To accomplish this, a subliminal response priming task was used that influences the difficulty of action selection. The benefit of this task is that participants remained unaware of the conflict-inducing stimulus itself, thereby eliminating the possibility that participants rated their sense of subjective difficulty based on perceiving the conflict-inducing stimuli or based on how they believe they should respond. In the current study, we observed that participants reported increased subjective difficulty on trials in which subliminal response conflict was induced. In a first analysis step, activation in the RCZ was found that was related to the raw subjective difficulty ratings. In a second step, we then wanted to unravel activity specific to subjective ratings, which was not explained by prime–response compatibility or RTs. This analysis showed that both the RCZ and the AI were related to unique variation in subjective ratings. Furthermore, replicating previous work, we found that activation in both these regions was increased in the presence of subliminal response conflict. In the remainder of this section, we discuss how these findings are compatible with a grounding of metacognitive experiences of difficulty within interoceptive signals.

Metacognitive Computations of Subjective Difficulty

The RCZ is part of ACC, which is a central hub in the cerebral cortex. This brain region plays a key role in cognitive control processes, as it is argued to be involved in conflict processing (Botvinick, Braver, Barch, Carter, & Cohen, 2001), in computing the expected value of control (Shenhav, Botvinick, & Cohen, 2013), in detecting violations from predictions (Alexander & Brown, 2011), and in effort processing (Walton, Bannerman, & Rushworth, 2002). These different functions have been integrated by positing that ACC controls the degree of effort invested in a certain task (Holroyd & McClure, 2015). Indeed, focal damage to rat medial frontal cortex decreased the frequency of high-effort responses to obtain a reward. Within this framework, the sensitivity of the anterior cingulate to response conflict results from an increased need for effort in difficult trials (i.e., conflict trials). Confirming this prediction and replicating previous work, the current study observed higher activation in the RCZ for incompatible trials compared to compatible and neutral trials. Indeed, we observed highly similar peaks in the current work compared to those observed in our previous study using the same design (Teuchies et al., 2016). This was the case both for the RCZ ([3, 23, 55] vs. [6, 20, 43], respectively) and the AI ([−36, 23, –2] vs. [−36, 20, –1], respectively), and the peaks of that previous study both fell within the clusters observed in the current study. Different from our previous study, however, in the current work, we also asked our participants to rate their metacognitive experience of difficulty after each response. This allowed us to go beyond our previous work and examine not only whether the RCZ and AI are sensitive to objective manipulations of response but also whether these brain regions are sensitive to the accompanying metacognitive experience. Given that subjective difficulty judgments track response conflict, a relation between subjective difficulty and RCZ was expected. Critically, however, we were able to demonstrate that the relation between difficulty ratings and RCZ was present, even after controlling for the influence of prime–response compatibility and RTs. This shows that metacognitive computations of subjective difficulty do not merely track experimental manipulations; rather, they are based on brain regions, such as the RCZ, that code for the required degree of effort, over and above that induced by the experimental manipulation.

Whether or not metacognitive computations of subjective difficulty are directly related to RCZ activity or only indirectly remains an open question. One interesting possibility is that the RCZ only codes for the required degree of effort, and this is subsequently implemented by other brain regions. In this regard, it is interesting that, apart from the RCZ, we also observed that the AI was sensitive to subjective difficulty ratings, over and above the effect of response conflict and RTs. The AI is a key brain region involved in interoceptive awareness (Grupe & Nitschke, 2013; Gu, Hof, Friston, & Fan, 2013; Craig, 2009). Interoception can be described as the sense of the physiological condition of the body, or the perception of sensory events occurring within one's body (Grupe & Nitschke, 2013; Craig, 2002, 2003). The AI is thought to monitor and control internal, embodied states, such as the degree of arousal. Thus, when the RCZ detects the need for increased effort allocation, this might subsequently be implemented by the AI that increases arousal, via interactions with the sympathetic nervous system. This interpretation is further supported by the results of the mediation analysis carried out in the current study, which suggest that the influence of the RCZ on subjective difficulty ratings is mediated by activity in the AI, although the mediation analysis should be interpreted with caution given that we did not observe a main effect of the RCZ in the outcome model. Interestingly, using intracranial recordings, Bastin et al. (2017) were able to show the same flow of information from AI to RCZ on correct trials, but the reverse information flow occurred on error trials. Given that we only studied (subjective difficulty on) correct trials in the current work, our findings would be in line with work of Bastin et al. It would be interesting for future studies to query subjective difficulty ratings in a more difficult task, so that it would be possible to examine the role of subjective difficulty in the information flow between AI and RCZ on errors trials too. Given that humans are typically unaware of their own brain activity (Prinz, 1992), this raises the interesting possibility that

metacognitive evaluations of difficulty are based on bodily signals in response to required effort. Thus, when judging whether a trial was easy or difficult, participants might integrate, among other things, their autonomous bodily reactions toward subliminal response conflict (i.e., cardiac acceleration, increased skin conductance; Hauser et al., 2017; Allen et al., 2016) to come to a single judgment of difficulty. Given that the AI is not specifically activated by interoceptive processes, currently, this proposal remains speculative. Future work could test this proposal by asking participants to not only indicate metacognitive experiences of difficulty after each response but also engage on each trial in interoceptive monitoring (e.g., reporting subjective heartbeat, perceived stress) and measure this activity objectively (e.g., using cardiac activity and skin conductance, respectively). The proposed account predicts that the relation between response conflict and metacognitive experiences of difficulty is fully mediated by interoceptive processes. In line with this proposal, a recent study demonstrated that participants relied on motor activity in their response effectors (i.e., in the thumbs of both hands) when judging the difficulty of a trial (Questienne et al., 2018).

Domain-General versus Domain-Specific Metacognition

In recent years, the metacognitive evaluation of performance has been tackled from different angles. This has raised the question whether this metacognitive evaluation of performance is supported by a set of domain-general mechanisms or whether there is domain specificity. To tackle this question, McCurdy et al. (2013) compared metacognition about visual performance with metacognition about memory performance. Although metacognitive performance was correlated across both domains (see also Faivre, Filevich, Solovey, Kühn, & Blanke, 2018; Song et al., 2011), different neural structures were involved in each. Whereas metacognitive performance about visual decisions was related to volume in anterior pFC (see also Fleming et al., 2010), metacognitive performance about memory decisions was related to the precuneus. In a subsequent study, metacognitive performance about visual decisions was linked to white matter microstructure in ACC, whereas metacognitive performance about memory was linked to white matter microstructure in the inferior parietal lobule (Baird, Mrazek, Phillips, & Schooler, 2014). The current work adds to this debate, by demonstrating that in a different type of self-evaluation, subjective difficulty in decision-making, both the RCZ and AI are involved. The latter is particularly interesting, because although the AI is critically involved in self-referential processes such as self-awareness (Craig, 2009) and becoming aware of your own errors (Klein et al., 2007), previous studies did not, to our knowledge, implicate the AI in the metacognitive evaluation of performance on correct trials. As such, the current findings lend further support for the domain-specific view of metacognition.

Conclusion

In the current work, we observed that the subjective sense of difficulty is represented in the RCZ and the AI, two regions that are functionally closely connected. Importantly, this was observed when controlling for prime–response compatibility and RTs. Because the RCZ and the AI typically activate in unison, future research is needed to test our hypothesis that the RCZ codes for the required level of effort and the AI implements this by increasing arousal, which is what participants become aware of.

Acknowledgments

The research reported in this article was funded by the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office (IUAPVII/33). K. D. is an FWO [PEGASUS]^2 Marie Skłodowska-Curie fellow (Grant Number 12T9717N). C. G. G. was supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Grant Agreement No. 835767. M. B. is supported by an Einstein Strategic Professorship of the Einstein Foundation Berlin.

Reprint requests should be sent to Kobe Desender, Brain and Cognition, KU Leuven, Tiensetraat 102, 3000 Leuven, Belgium, or via e-mail: Kobe.Desender@kuleuven.be.

Author Contributions

Kobe Desender: Conceptualization; Formal analysis; Visualization; Writing—Original draft. Martyn Teuchies: Conceptualization; Data curation; Formal analysis; Investigation; Writing—Original draft. Carlos Gonzalez Garcia: Formal analysis; Visualization; Writing—Review & editing. Wouter De Baene: Supervision. Jelle Demanet: Investigation. Marcel Brass: Conceptualization; Supervision; Writing—Review & editing.

Funding Information

Carlos Gonzalez Garcia, H2020 Marie Skłodowska-Curie Actions (https://dx.doi.org/10.13039/100010665), grant number: 835767. Kobe Desender, Federaal Wetenschapsbeleid (https://dx.doi.org/10.13039/501100002749), grant number: IUAPVII/33. Kobe Desender, Fonds Wetenschappelijk Onderzoek (https://dx.doi.org/10.13039/501100003130), grant number: 12T9717N.

Diversity in Citation Practices

A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed

a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .408, W(oman)/M = .335, M/W = .108, and W/W = .149, the comparable proportions for the articles that these authorship teams cited were M/M = .579, W/M = .243, M/W = .102, and W/W = .076 (Fulvio et al., *JoCN*, 33:1, pp. 3–7). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*, 1338–1344. https://doi.org/10.1038/nn.2921, PubMed: 21926982

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., et al. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *eLife*, *5*, e18103. https://doi.org/10.7554/eLife.18103, PubMed: 27776633

Baird, B., Mrazek, M. D., Phillips, D. T., & Schooler, J. W. (2014). Domain-specific enhancement of metacognitive ability following meditation training. *Journal of Experimental Psychology: General*, *143*, 1972–1979. https://doi.org/10.1037/a0036882, PubMed: 24820248

Bastin, J., Deman, P., David, O., Gueguen, M., Benis, D., Minotti, L., et al. (2017). Direct recordings from human anterior insula reveal its leading role within the error-monitoring network. *Cerebral Cortex*, *27*, 1545–1557. https://doi.org/10.1093/cercor/bhv352, PubMed: 26796212

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, *35*, 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015, PubMed: 25716847

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652. https://doi.org/10.1037/0033-295X.108.3.624, PubMed: 11488380

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 77–165). Hillsdale, NJ: Erlbaum.

Craig, A. D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience*, *3*, 655–666. https://doi.org/10.1038/nrn894, PubMed: 12154366

Craig, A. D. (2003). Interoception: The sense of the physiological condition of the body. *Current Opinion in Neurobiology*, *13*, 500–505. https://doi.org/10.1016/S0959-4388(03)00090-4, PubMed: 12965300

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*, 59–70. https://doi.org/10.1038/nrn2555, PubMed: 19096369

Desender, K., Buc Calderon, C., Van Opstal, F., & Van den Bussche, E. (2017). Avoiding the conflict: Metacognitive awareness drives the selection of low-demand contexts. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1397–1410. https://doi.org/10.1037/xhp0000391, PubMed: 28368164

Desender, K., Van Opstal, F., Hughes, G., & Van den Bussche, E. (2016). The temporal dynamics of metacognition: Dissociating task-related activity from later metacognitive processes. *Neuropsychologia*, *82*, 54–64. https://doi.org/10.1016/j.neuropsychologia.2016.01.003, PubMed: 26777465

Desender, K., Van Opstal, F., & Van den Bussche, E. (2014). Feeling the conflict: The crucial role of conflict experience in adaptation. *Psychological Science*, *25*, 675–683. https://doi.org/10.1177/0956797613511468, PubMed: 24395737

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences, U.S.A.*, *113*, 7900–7905. https://doi.org/10.1073/pnas.1602413113, PubMed: 27357684

Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. *Journal of Neuroscience*, *38*, 263–277. https://doi.org/10.1523/JNEUROSCI.0322-17.2017, PubMed: 28916521

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*, 1541–1543. https://doi.org/10.1126/science.1191883, PubMed: 20847276

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, *14*, 488–501. https://doi.org/10.1038/nrn3524, PubMed: 23783199

Gu, X., Hof, P. R., Friston, K. J., & Fan, J. (2013). Anterior insular cortex and emotional awareness. *Journal of Comparative Neurology*, *521*, 3371–3388. https://doi.org/10.1002/cne.23368, PubMed: 23749500

Hauser, T. U., Allen, M., Purg, N., Moutoussis, M., Rees, G., & Dolan, R. J. (2017). Noradrenaline blockade specifically enhances metacognitive performance. *eLife*, *6*, e24901. https://doi.org/10.7554/eLife.24901, PubMed: 28489001

Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, *122*, 54–83. https://doi.org/10.1037/a0038339, PubMed: 25437491

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*, 116–119. https://doi.org/10.1126/science.1111709, PubMed: 16210542

Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J.-D. (2011). Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. *Neuroimage*, *56*, 709–715. https://doi.org/10.1016/j.neuroimage.2010.05.058, PubMed: 20510371

Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *Neuroimage*, *34*, 1774–1781. https://doi.org/10.1016/j.neuroimage.2006.11.014, PubMed: 17185003

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (R package, version 2.0-6). Available at: https://cran.r-project.org/web/packages/lmerTest/index.html.

Mazaika, P., Whitfield-Gabrieli, S., & Reiss, A. L. (2007). *Artifact repair for fMRI data from high motion clinical subjects*. Chicago: Organization for Human Brain Mapping.

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception.

Journal of Neuroscience, *33*, 1897–1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013, PubMed: 23365229
Metcalfe, J., & Shimamura, A. P. (Eds.) (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001
Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *eLife*, *4*, e11946. https://doi.org/10.7554/eLife.11946, PubMed: 26687008
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259. https://doi.org/10.1037/0033-295X.84.3.231
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4, PubMed: 5146491
Prinz, W. (1992). Why don't we perceive our brain states? *European Journal of Cognitive Psychology*, *4*, 1–20. https://doi.org/10.1080/09541449208406240
Questienne, L., van Dijck, J.-P., & Gevers, W. (2018). Introspection of subjective feelings is sensitive and specific. *Journal of Experimental Psychology: Human Perception and Performance*, *44*, 215–225. https://doi.org/10.1037/xhp0000437, PubMed: 28504525
Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*, 217–240. https://doi.org/10.1016/j.neuron.2013.07.007, PubMed: 23889930
Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Consciousness and Cognition*, *20*, 1787–1792. https://doi.org/10.1016/j.concog.2010.12.011, PubMed: 21256051
Stevens, M., Lammertyn, J., Verbruggen, F., & Vandierendonck, A. (2006). Tscope: A C library for programming cognitive experiments on the MS Windows platform. *Behavior Research Methods*, *38*, 280–286. https://doi.org/10.3758/BF03192779, PubMed: 16956104
Teuchies, M., Demanet, J., Sidarus, N., Haggard, P., Stevens, M. A., & Brass, M. (2016). Influences of unconscious priming on voluntary actions: Role of the rostral cingulate zone. *Neuroimage*, *135*, 243–252. https://doi.org/10.1016/j.neuroimage.2016.04.036, PubMed: 27138208
Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*, 1–38. https://doi.org/10.18637/jss.v059.i05
Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure & Function*, *214*, 629–643. https://doi.org/10.1007/s00429-010-0261-1, PubMed: 20512371
Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences, U.S.A.*, *100*, 6275–6280. https://doi.org/10.1073/pnas.0931489100, PubMed: 12719543
Walton, M. E., Bannerman, D. M., & Rushworth, M. F. S. (2002). The role of rat medial frontal cortex in effort-based decision making. *Journal of Neuroscience*, *22*, 10996–11003. https://doi.org/10.1523/JNEUROSCI.22-24-10996.2002, PubMed: 12486195
Wilson, T. D., & Dunn, E. W. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Annual Review of Psychology*, *55*, 493–518. https://doi.org/10.1146/annurev.psych.55.090902.141954, PubMed: 14744224