Temporal-sensitive prefrontal involvement in associating confidence with task performance illustrates metacognitive introspection in Old World monkeys (*Macaca mulatta*)

Yudian Cai[1,2], Zhiyong Jin[1,2], Chenxi Zhai[1], Huimin Wang[1,3,4], Jijun Wang[5,6,7], Yingying Tang[5 #], Sze Chai Kwok[1,2,4 #]

[1] *Shanghai Key Laboratory of Brain Functional Genomics, Key Laboratory of Brain Functional Genomics Ministry of Education, Shanghai Key Laboratory of Magnetic Resonance, Affiliated Mental Health Center (ECNU), School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China*
[2] *Division of Natural and Applied Sciences, Duke Kunshan University, Kunshan, Jiangsu 215316, China*
[3] *NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, Shanghai 200062, China*
[4] *Shanghai Changning Mental Health Center, Shanghai 200335, China*
[5] *Brain Science and Technology Research Center, Shanghai Jiao Tong University, Shanghai 200030, China*
[6] *CAS Center for Excellence in Brain Science and Intelligence Technology (CEBSIT), Chinese Academy of Science, China*
[7] *Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China*

Denote corresponding authors. Email: sze-chai.kwok@st-hughs.oxon.org; yytang0522@gmail.com

Acknowledgment
This research received support from Science and Technology Commission of Shanghai Municipality (201409002800), National Natural Science Foundation of China (32071060), and Jiangsu Qinglan Talent Program Award (S.C.K.). We thank Yong-di Zhou for his advice on NHP research; Makoto Kusunoki for headpost implementation; Lei Wang, Shuzhen Zuo, Angie Xie, Aihua Chen, Hakwan Lau, and Alicia Izquierdo for their input in the preparation of the manuscript.

Abstract

Metacognition refers to the ability to be aware of one's own cognition. Ample evidence indicate that metacognition in the human primate is highly dissociable from cognition, specialized across domains, and subserved by distinct neural substrates. However, these aspects remain relatively understudied in the macaque monkeys. Here we investigated the functionality of macaques' metacognition by combining a confidence proxy, hierarchical Bayesian meta-d' computational modelling, and single-pulse transcranial magnetic stimulation technique. We found a critical role of Brodmann Area 46d (BA 46d) in supporting metacognition without implicating task performance and its critical time-sensitive role in meta-calculation. We additionally revealed that macaque metacognition is highly domain-specific with respect to memory and perception decisions. These findings carry implications for our understanding of metacognitive introspection within the primate lineage.

Keywords: confidence; domain-specificity; metacognition; monitoring; dorsolateral prefrontal cortex; Transcranial magnetic stimulation

Introduction

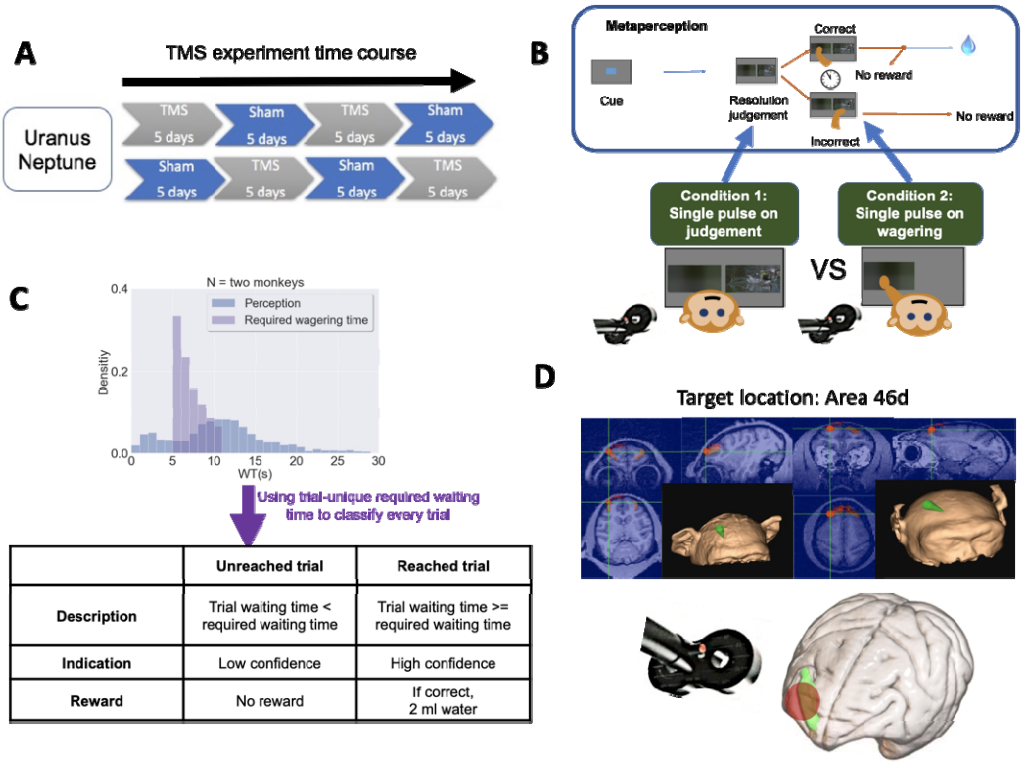
Metacognition, the ability to monitor and evaluate one's own cognitive processes, is believed to be unique in human. Ample evidence indicate that neural underpinnings supporting metacognitive abilities are different from cognitive processes (Fleming et al., 2010; Fleming et al., 2012; McCurdy et al., 2013; Fleming et al., 2014b; Rahnev et al., 2015; Morales et al., 2018; Rouault et al., 2018; Brown et al., 2019; Gilbert et al., 2020). A number of human transcranial magnetic stimulation (TMS) studies has implicated the dorsolateral prefrontal cortex (dlPFC) in meta-perceptual judgments more than in perceptual judgements (Rounis et al., 2010; Rahnev et al., 2016; Shekhar and Rahnev et al., 2018). These evidence indicate that the prefrontal cortex, especially the lateral prefrontal cortex (lPFC), is a key region of the metacognitive mechanism (Odegaard et al., 2017; Brown et al., 2019; Lapate et al., 2020).

However, is the importance of dlPFC conserved in other species like non-human primates? Only one extant study has investigated the role of macaques' dlPFC in meta-perceptual processes. They found that in a visual-oculomotor task, single neurons in the dlPFC would encode metacognitive components of decision (Middlebrooks and Sommer, 2012). To extend on their finding, our first aim was to test for any functional role of the monkey dlPFC in meta-perception independently of perception itself. To attain this, we applied single-pulse transcranial magnetic stimulation to the dlPFC (BA 46d) of monkeys while they underwent a perceptual resolution judgement task. We adopted a temporal-wagering paradigm to measure the animals' trial-wise confidence (Lak et al., 2014; Stolyarova et al., 2019; Masset et al., 2020). Following each perceptual decision, the animals had to wait for an unknown and variable period by holding their hand on the screen before qualifying for any possible reward. The amount of time wagered on their trial-wise decision was used as a proxy for decision confidence.

Taking advantage of the high temporal resolution by TMS, we intended to ascertain the precise window in which meta-computation is carried out. An electrophysiology study reported that information carried by lateral intraparietal cortex (LIP) neurons at the time of decision is sufficient for predicting subsequent confidence-related neural responses (Kiani and Shadlen, 2009). However, single pulse TMS on the dorsal premotor cortex (PMd) impairs confidence reports at both

pre-response and post-response windows (Fleming et al., 2015), suggesting late-stage evidence accumulation might also be required for metacognitive processes. In order to more precisely determine the critical phase of meta-calculations, here we included two time-sensitive TMS conditions: on-judgement and on-wagering stimulation. Specifically, we applied TMS either at 100ms after the stimulus onset (on-judgement phase) or at 100ms after the animal's decision (on-wagering phase). We hypothesized that if the critical phase of meta-calculation was within the decision stage, we would expect metacognition deficits when TMS was applied during the on-judgement phase. In contrast, if the meta-computation was at a later stage (e.g., together with processes associated with "wagering"), we would expect metacognition deficits when TMS was applied during the on-wagering phase.

There is evidence that efficient metacognition in one task can predict good metacognition in another task (Baird et al., 2013a; Ais et al., 2016; Faivre et al., 2017; Samaha and Postle, 2017; Alan et al., 2018; Carpenter et al., 2019). In the nonhuman primates, monkeys' ability to transfer their metacognitive judgement from a perceptual test to a memory test show that they could employ domain-general signals to monitor the status of cognitive processes and knowledge levels (Kornell et al., 2007; Brown et al., 2017), suggesting that metacognition is domain-general. However, mounting anatomical (McCurdy et al., 2013; Maniscalco et al., 2017), functional (Morales et al., 2018), and neuropsychological (Fleming et al., 2014b; Ye et al., 2018) evidence in the human literature increasingly point to domain-specificity of metacognition, indicating that humans posses specialized metacognitive abilities across domains (Kelemen et al., 2000; Baird et al., 2013b; Vo et al., 2014; Morales et al., 2018). Here, we posed the question of whether macaques show metacognitive domain-specific components (cf. Kornell et al., 2007). To this end, we additionally trained two more monkeys to perform a temporal-memory task in combination of the wagering part. Making use of the data collected in both experiments, we assessed both covariation and divergence between metacognitive abilities in the two domains.



	Unreached trial	Reached trial
Description	Trial waiting time < required waiting time	Trial waiting time >= required waiting time
Indication	Low confidence	High confidence
Reward	No reward	If correct, 2 ml water

Figure 1. TMS experiment time design. TMS Experiment schedule with TMS/Sham conditions counterbalanced between monkeys (Uranus and Neptune) (A). Perceptual judgement task with temporal wagering. Each trial is consisted of a starting (blue) cue, a 1-s ~ 6-s delay, and two simultaneously presented pictures. The monkeys needed to choose a picture with a lower (or higher, counterbalanced across monkeys) resolution by holding their hand on the touchscreen. The waiting process is initiated as soon as they lay their hand on the picture. Their decision-related confidence is measured by temporal wagering: that is, they could wait for a reward if they are confident or opt out to abort the current trial. There are two time-sensitive TMS conditions. On each trial, the monkeys either received a TMS single pulse right after the picture stimulus onset (on-judgement phase) or 100ms after they made their perceptual decision (on-wagering phase) (B). Required WT distribution and the actual WT distribution (only catch trials and incorrect trials) with WT bin size set as 1 s. The table depicts a classification of low confidence trial (unreached trials) and high confidence trial (reached trials) (C). An illustration for TMS modulation location, as indicated by the green arrows. Bottom: The green area indicates BA 46d on a rendered macaque brain; the red disc indicates the target area (D).

Results

Metacognition in monkeys in both memory and perception domains

To show macaques are metacognitively capable, we used a bias-free metacognitive efficiency (H-model *meta-d' /d'*) to quantify monkeys' metacognition. We compared animals' score with one-sample t-tests against zero and showed above zero meta-index values in all monkeys for both tasks (Fig. 2C & 2D; meta-perception: H-model *meta-d' /d'*: Mars, $t_{(19)} = 5.685$, $p < 0.001$; Saturn, $t_{(19)} = 5.639$, $p < 0.001$; Uranus: $t_{(19)} = 10.55$, $p < 0.001$; Neptune, $t_{(19)} = 9.458$, $p < 0.001$; meta-memory: H-model *meta-d' /d'*: Mars, $t_{(19)} = 9.012$, $p < 0.001$; Saturn, $t_{(19)} = 5.639$, $p < 0.001$; Uranus: $t_{(19)} = 4.159$, $p < 0.001$; Neptune, $t_{(19)} = 3.621$, $p < 0.001$).

We then replicated the results with Phi coefficient (meta-perception: Phi coefficient: Mars, $t_{(19)} = 3.643$, $p < 0.001$; Saturn, $t_{(19)} = 6.245$, $p < 0.001$; Uranus: $t_{(19)} = 6.722$, $p < 0.001$; Neptune, $t_{(19)} = 3.423$, $p < 0.001$; meta-memory: Phi coefficient: Mars, $t_{(19)} = 4.135$, $p < 0.001$; Saturn, $t_{(19)} = 2.962$, $p = 0.004$; Uranus: $t_{(19)} = 2.252$, $p = 0.018$; Neptune, $t_{(19)} = 1.838$, $p = 0.041$).

To further validate these results, we combined all trials per monkey across all days and then performed subject-based distribution simulations to each monkey. Through randomly shuffling all the pairings between "responses" (correct/incorrect) and their corresponding "confidence level" (high/low) within each subject, we generated 2,000 random pairings for each animal and simulated 4,000 meta-scores per animal (both H-model *meta-d' /d'* and Phi coefficient). These scores represent cases in which the animals had no meta-ability. We then tested these simulated scores with animals' actual scores using a minimum statistic method (Nichols et al., 2005) and showed that the animals indeed performed significantly above chance meta-ability in both tasks (all *p* values < 0.001; Table 1).

As a control, we ruled out training effects as a contributing factor by comparing the animals' metacognition scores between the first ten days and the second ten days of testing. We found no difference between first ten days and second ten days on metacognitive performance in neither perception (H-model *meta-d' /d'*: ($t_{(39)}$ = -0.314, p = 0.755) nor in memory (H-model *meta-d' /d'*: ($t_{(39)}$ = 0.89, p = 0.378). These results show their metacognitive ability was stable across the whole testing period. For completeness, we checked the monkeys' cognitive performance and found that they improved moderately in the second half in the memory task (accuracy: $t_{(39)}$ = -2.266, p = 0.029), but not in the perception task ($t_{(39)}$ = -1.083, p = 0.285).

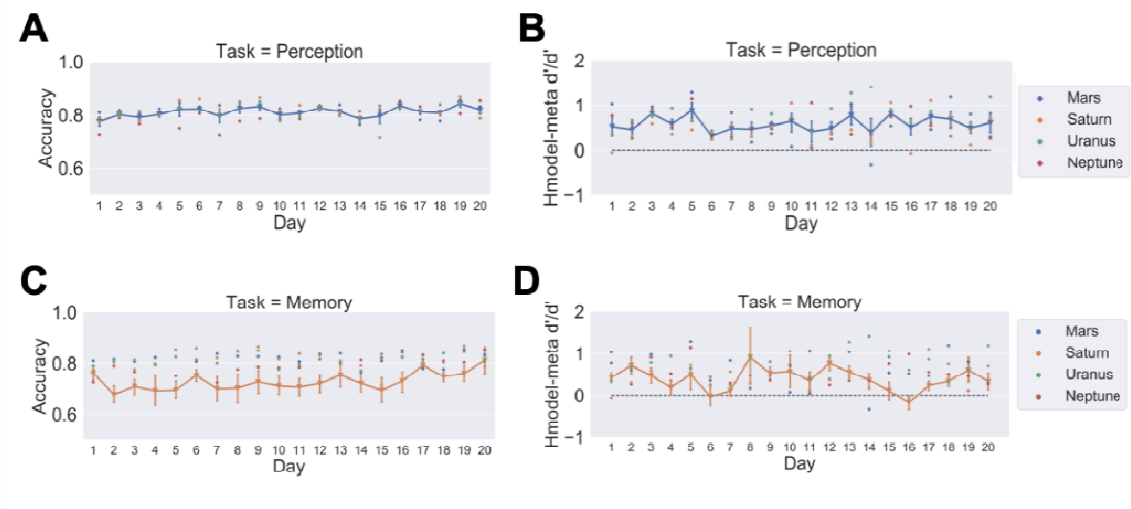


Figure 2. Task performance and metacognitive capability remained steady across days. Plots depict the daily accuracy (A & C) and metacognitive efficiency (B & D) across 20 days for four monkeys for two tasks. Colored dots represent each monkey's value. Error bars indicate ± one standard error.

TMS on BA 46d impairs metacognitive performance but not cognitive performance

We then turned to our main question. We tested whether TMS on the BA 46d will affect metacognition on perceptual decision. We performed a 2 (TMS-phase: on-judgement / on-wagering) × 2 (TMS: TMS-46d / TMS-Sham) mixed-design repeated-measures ANOVA for metacognitive efficiency with TMS-phase as a within-subjects

factor and with TMS as a between-subjects factor. We found a significant interaction between TMS-phase and TMS modulation in both monkeys (Neptune, $F_{(1,18)} = 6.431$, $p = 0.021$; Uranus, $F_{(1,18)} = 10.718$, $p = 0.004$). The interaction was driven by lower metacognitive efficiency following TMS to BA 46d relative to sham in the on-judgement phase condition (paired t-tests: Neptune, $t_{(9)} = 3.675$, $p = 0.002$; Uranus, $t_{(9)} = 2.741$, $p = 0.013$), whereas no difference in metacognitive efficiency was found in the on-wagering phase (paired t-tests: Neptune, $t_{(9)} = -0.3$, $p = 0.768$; Uranus, $t_{(9)} = -0.841$, $p = 0.411$), see Fig. 3A & 3B. We replicated the metacognition-deficit in the on-judgement phase with Phi coefficient (paired t-tests: Neptune, $t_{(9)} = 3.51$, $p = 0.002$; Uranus, $t_{(9)} = 5.637$, $p < 0.001$).

These meta-indices are based on how the subjects rate their confidence, which refer to how meaningful a subject's confidence (wagering time here) is in distinguishing between correct and incorrect responses. We accordingly performed a three-way ANOVA (TMS-phase: on-judgement / on-wagering × TMS: TMS-46d / TMS-Sham × Confidence: unreached / reached) on task performance (accuracy) and obtained a significant three-way interaction in both monkeys (Neptune, $F_{(1,2313)} = 5.530$, $p = 0.019$; Uranus $F_{(1,2295)} = 6.910$, $p = 0.009$). The TMS effect was disproportionally stronger in the on-judgement TMS-phase (TMS × Confidence interaction: Neptune, $F_{(1,1167)} = 10.672$, $p = 0.001$; Uranus $F_{(1,1160)} = 10.404$, $p < 0.001$, Fig. 3C), than in the on-wagering TMS-phase (TMS × Confidence interaction: Neptune, ($F_{(1,1146)} = 0.003$, $p = 0.954$; Uranus $F_{(1,1135)} = 0.309$, $p = 0.579$; Fig. 3D). The effects in the on-judgement TMS-phase were driven by higher accuracy following TMS-46d than TMS-sham in the unreached trials (Mann-Whitney tests: Neptune, $p = 0.001$; Uranus, $p < 0.001$), but not in the reached trials (Mann-Whitney tests: Neptune, $p = 0.235$; Uranus, $p = 0.192$). These findings confirmed that TMS targeted at BA 46d impairs metacognitive ability at the trial-by trial levels.

We further verified that type 1 task performance and mean wagering time were not affected by TMS. As expected, task performance (daily accuracy), reaction time (RT), and wagering time (WT) were not different between the two TMS conditions in neither on-judgement phase (paired T-test, all p values > 0.1 for both monkeys in accuracy, RT, and WT) nor on-wagering phase (paired T-test, all p values > 0.1 for both monkeys in accuracy, RT, and WT). These findings confirmed our first hypothesis that the monkey dlPFC is critical for meta-perception and such effects are independent of perception processes.

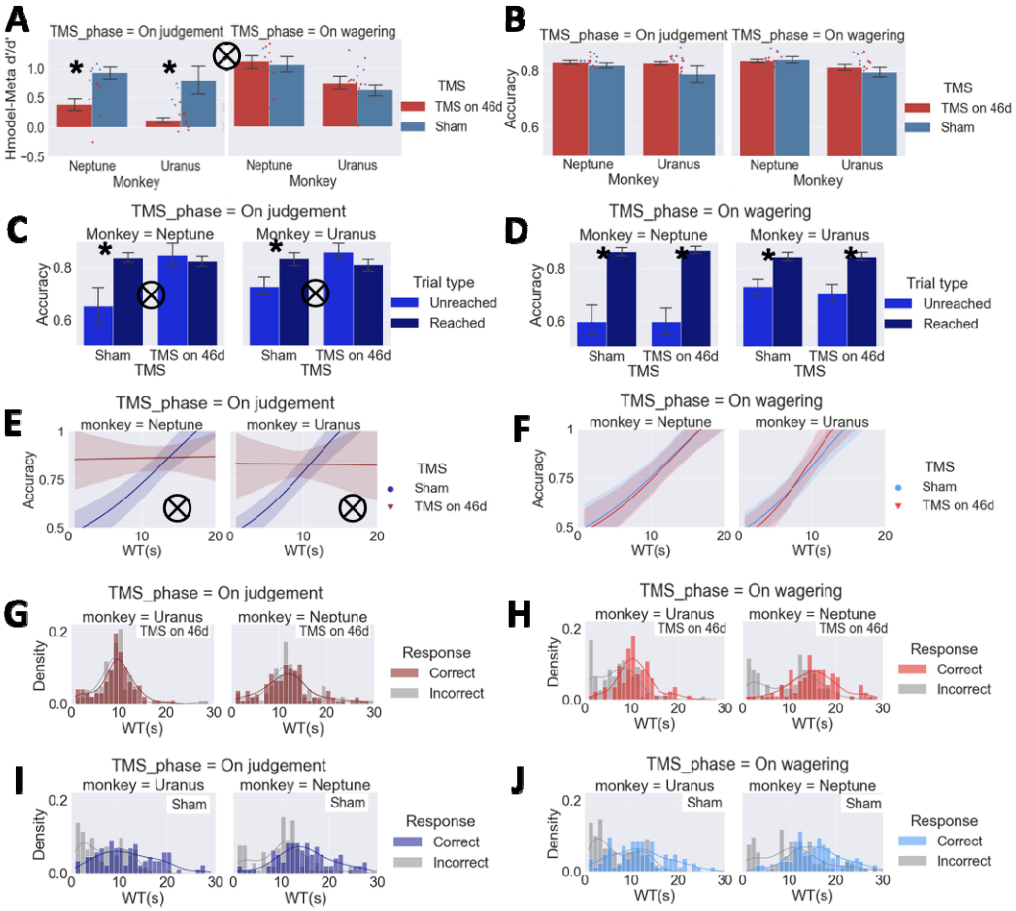


Figure 3. TMS during on-judgement phase disrupts metacognition and response outcome tracking ability of wagering time (WT). The monkeys demonstrate an impairment in metacognitive efficiency in TMS-46d condition during on-judgement phase but not during on-wagering phase (A). TMS on area 46d does not affect task accuracy (B). Difference of accuracy between unreached trials (low confidence) and reached trials (high confidence) in on-judgment phase and on-wagering phase (C & D). The lines are logistic regression fits on accuracy with WT as a factor separately for TMS-Sham and TMS-46d condition. WT reliably tracks response outcome in TMS-Sham condition but not in TMS-46d condition during on-judgement phase. WT tracks response outcome in both TMS-Sham and TMS-46d condition during TMS on-wagering phase (E & F). Distributional differences between correct and incorrect WT. Largest effects are seen in TMS-Sham conditions in which the monkey BA 46d was not perturbed (G, H, I, J). WT bin size set as 1 s, color lines indicate kernel density estimation. Error bars indicate ± one standard error, *

indicates p < 0.05. □ indicates a significant WT and TMS (TMS on 46d / Sham) interaction (p < 0.05). Shaded areas indicate bootstrap-estimated 95% confidence interval for the regression estimate.

Instantiation of the TMS impairment: Reduced accuracy-tracking ability of wagering time; altered reaction time-wagering time association, and altered trial-difficulty psychometric curve

We examined whether TMS would affect the ability of tracking task performance with WT in the two TMS-phase (on-judgement / on-wagering). We focused our analysis on catch trials and incorrect trials since we could not have measured the precise WT for some trials (i.e., correct reached trials, see methods). We performed logistic regression on correctness with WT, TMS (TMS-46d / TMS-Sham), and cross-product items as factors to test whether TMS on BA 46d affected the WT response-tracking precision. We found a significant interaction between TMS and WT in on-judgement TMS-phase (both monkeys: β_3 = -0.149, Standard Error = 0.029, Odds Ratio = 0.862, z= -5.115, p < 0.001, Fig. 3E), but not during on-wagering phase (both monkeys: β_3 = 0.010, Standard Error = 0.030, Odds Ratio = 1.010, z= 0.321, p = 0.748, Fig. 3F). Such effect in the on-judgement phase was driven by higher WT in correct trials than incorrect trials in the TMS-sham condition (Mann-Whitney tests: Neptune, p < 0.001; Uranus, p < 0.001, Fig. 3I&J), but not in the TMS-46d condition (Mann-Whitney tests: Neptune, p = 0.98; Uranus, p = 0.45, Fig. 3G). We also confirmed that WT can predict the trial outcome in a graded manner in on-wagering phase (β_1 = 0.152, Standard Error = 0.020, Odds Ratio = 1.164, z= 7.631, p < 0.001). These results revealed that TMS to Area 46d affects the metacognitive performance when it was administered during on-judgement phase. We obtained the same results when performing these logistic regressions on the two monkeys separately (Table 2).

Secondly, metacognitive abilities in animals are often confounded by behavioral association (Hampton, 2009). For example, animals are believed to make use of cues (environmental cues like stimulus condition, self-generated cues like response time) to index confidence instead of performing the task metacognitively. To rule out this possibility, we calculated the correlation between RT and WT in both experiments to check if the monkeys rely on RT as an associative cue to index confidence. The results showed no RT × WT correlation in the domain-comparison experiment (Fig. 4A), indicating the macaques did not rely on RT as an associative

cue to index their WT. We then utilized this phenomenon to verify the effect of TMS. The WT is significantly negatively correlated with RT only in TMS-46d condition during the on-judgement TMS-phase (r = -0.195, p < 0.001), but not in TMS-sham condition (Fig. 4B). We found a significant difference in correlation coefficient between TMS-46d and TMS-sham in on-judgement phase (z = -2.24, p = 0.0251). It is possible that monkeys started to rely on RT as an associative cue after having received TMS on area 46d, which hampered their metacognitive ability. As a control comparison, no difference was found between TMS conditions in the on-wagering phase (Fig. 4C).

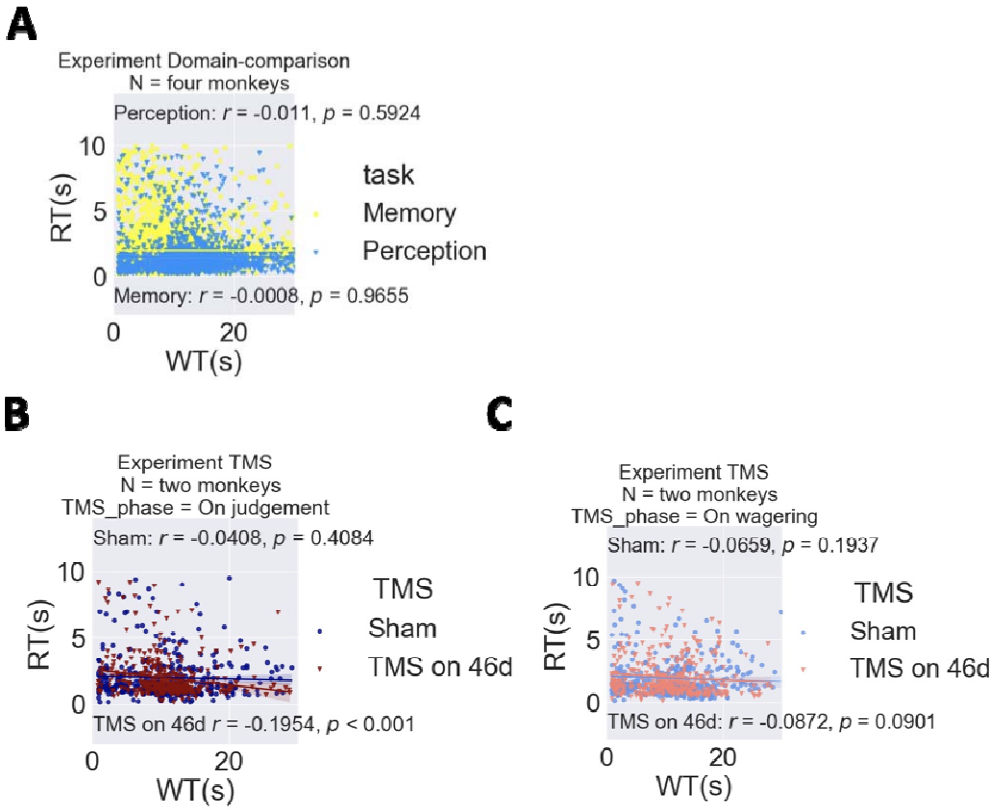


Figure 4. On-judgement TMS alters correlation between reaction time (RT) and wagering time (WT). No correlation between RT and WT in domain-comparison experiment (A). The Pearson correlation between RT and WT was statistically significant for the TMS 46d condition ($p < 0.001$) but not significant for TMS Sham during on-judgment phase (B). The correlations were not significant for both TMS conditions during on-wagering phase (C).

Moreover, as seen in the rodents literature, WT could be expressed as a function of strength of evidence (e.g., odor mixture ratio in their task) and response outcome (correct / incorrect) (Masset et al., 2020); the level of confidence should increase with evidence strength (resolution difference in our experiments) for correct trials and decrease with evidence strength for incorrect trials. We performed GLM to predict WT with four variables: TMS (TMS-46d / TMS-Sham), TMS-phase (on-judgement / on-wagering), resolution-difference, and correctness and their cross-product items. We found a four-way interaction in the monkeys (Neptune, $\beta_{\text{TMS} \times \text{TMS-phase} \times \text{correctness} \times \text{resolution-difference}} = -60.66$, $p = 0.010$; Uranus, $\beta_{\text{TMS} \times \text{TMS-phase} \times \text{correctness} \times \text{resolution-difference}} = -44.76$, $p = 0.019$). Plotting these results out with trial-difficulty psychometric curves, the effects were driven by stronger correctness × resolution-difference interaction in the TMS-Sham condition (including trials in both on-judgement TMS-phase and on-wagering TMS-phase) (Neptune, $\beta_{\text{correctness} \times \text{resolution-difference}} = 48.99$, $p < 0.001$; Uranus, $\beta_{\text{correctness} \times \text{resolution-difference}} = 42.20$, $p < 0.001$) and no effect in the TMS-46d on-judgement condition (Neptune, $\beta_{\text{correctness} \times \text{resolution-difference}} = 13.55$, $p = 0.119$; Uranus, $\beta_{\text{correctness} \times \text{resolution-difference}} = -2.50$, $p = 0.753$, Fig. 5C).

Critically, the correctness × resolution-difference interaction was driven by the increased WT for correct trials in TMS-Sham condition (including trials in both on-judgement TMS-phase and on-wagering TMS-phase) (Neptune, $\beta_{resolution\text{-}difference} = 27.47$, $p < 0.001$; Uranus, $\beta_{resolution\text{-}difference} = 27.76$, $p < 0.001$), and decreased WT for incorrect trials (Neptune, $\beta_{resolution\text{-}difference} = -21.51$, $p < 0.001$; Uranus, $\beta_{resolution\text{-}difference} = -14.43$, $p < 0.001$, Fig. 5D-F). These results suggest that under TMS-sham condition, WT increase with resolution-difference for correct trials and decrease with resolution-difference for incorrect trials irrespective of TMS-phase, whereas such pattern was disrupted during on-judgement stage in the TMS-46d condition. Additionally, we confirmed that perceptual performance is intact by performing logistic regression on response outcome with resolution-difference, TMS (TMS-46d / TMS-Sham), and cross-product item as factors. We found no interactions for both on-judgement TMS-phase and on-wagering TMS-phase in the monkeys (all *Ps* > 0.05).

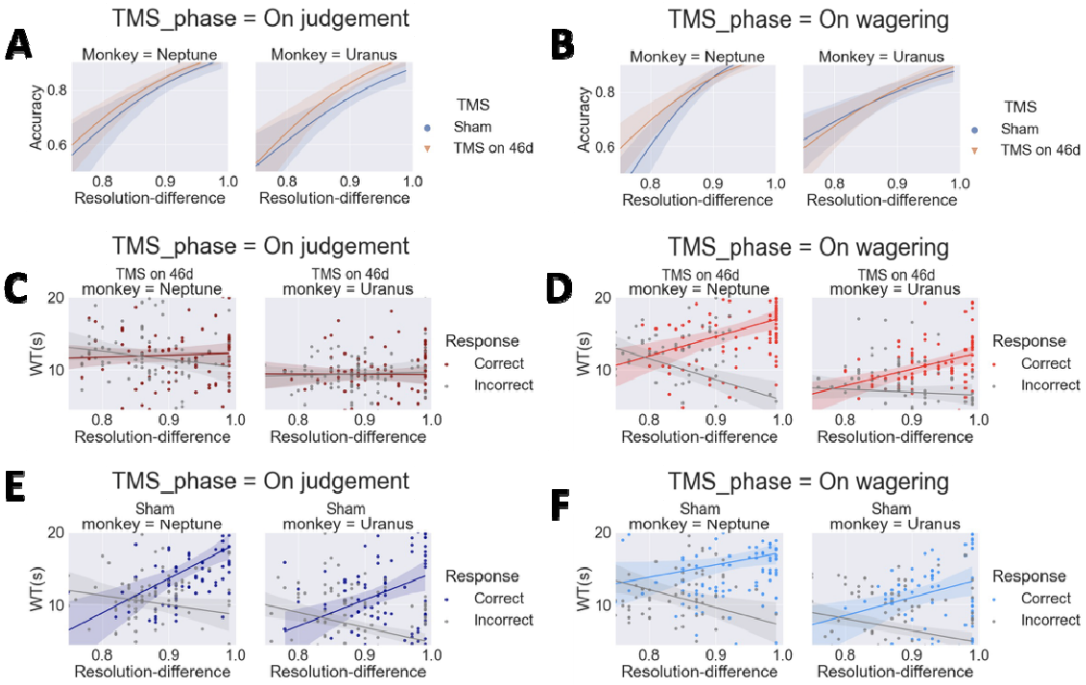


Figure 5. On-judgement TMS distorts trial-difficulty psychometric curve. Accuracy decreases with task difficulty (resolution-difference, higher values indicate lower task difficulty). The lines are logistic regression fits on accuracy with resolution-difference as a factor separately for TMS-sham and TMS-46d condition in on-judgement phase (A) and on-wagering phase (B). WT decrease with task difficulty in correct trials and increase with task difficulty in incorrect trials in all control conditions (D-F), but this pattern is eradicated in the on-judgement phase TMS-46d condition (C). Shaded areas indicate bootstrap-estimated 95% confidence interval for the regression estimate.

Qualifier on monkey's metacognition: Wagering time (WT) is diagnostic of the animals' performance

To further instantiate these results, we expected that monkeys can index their confidence using their trial-by-trial wagering time. We showed that wagering time is diagnostic of the animals' performances with a number of analyses. First, by comparing the accuracy in reached (high confidence) and unreached (low confidence) trials, Chi-Square tests revealed that monkeys have higher accuracy in high confidence trials in meta-perception (all four monkeys: $\chi^2_{(1)} = 31.88$, $p < 0.001$; for individual monkeys: all p values < 0.05, Fig. 6A) and in metamemory (all four monkeys: $\chi^2_{(1)} = 13.41$, $p < 0.001$; for individual monkeys: all p values < 0.05, Fig. 6B). To test if the monkeys could track their response outcome with their WT, we performed logistic regression on response outcome with WT, task (memory/perception), and cross-product item as factors. We confirmed that the WT can accurately predict the trial outcome ($\beta_1 = 0.033$, Standard Error = 0.007, Odds Ratio = 1.033, z= 4.586, $p < 0.001$; Fig. 6E). We found no interaction between Task and WT ($\beta_3 = 0.0014$, Standard Error = 0.011, Odds Ratio = 1.014, z = 1.335, p = 0.182), indicating that WT in both memory and perception tasks produced the same response outcome tracking. These results showed that trial-wise wagering time are diagnostic the animals' decision outcome, reflecting that the monkeys were aware of their judgement outcome. All results held when we performed the analyses for each monkey individually (Table 3).

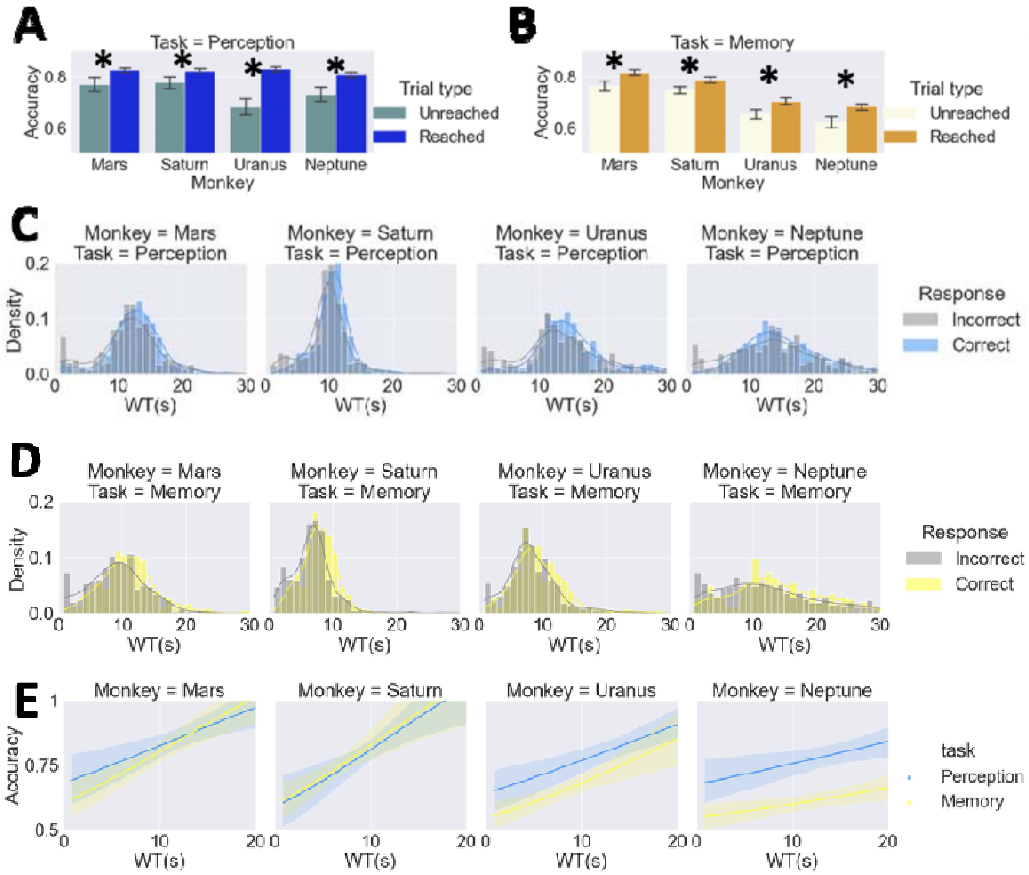
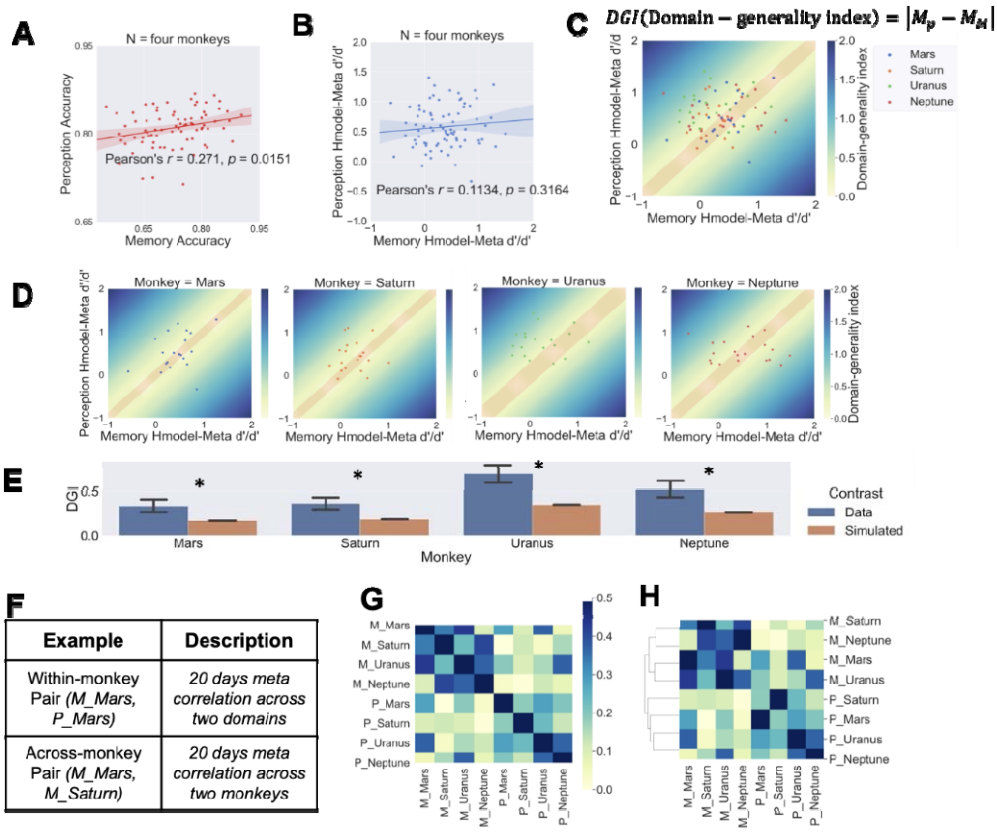


Figure 6. Wagering time reflect monkeys' task performance (correctness) in both memory and perception tasks. Difference of accuracy between unreached trials and reached trials accuracy in perception (A) and memory task (B). Distributional differences between correct and incorrect WT for each monkey in perception (C) and memory task (D). WT tracks response outcome (correct / incorrect) in both memory and perception tasks. The lines are logistic regression fits on accuracy with WT as a factor. WT bin size set as 1 s; color lines indicate kernel density estimation (E). Error bars indicate ± one standard error, * indicates $p < 0.05$. Shaded areas indicate bootstrap-estimated 95% confidence interval for the regression estimate.

Qualifier on monkey's metacognition: Evidence on domain-specificity

While we found a positive correlation between perception and memory domains in each days' individual accuracy ($r_{(80)}$ = 0.271; p = 0.0151; Fig. 7A), their respective metacognitive efficiency scores do not correlate ($r_{(80)}$ = 0.1134; p = 0.3164; right panel in Fig. 7B). This thus prompted us to look into the domain-specificity with the bias-free metacognitive efficiency (H-model *meta-d' /d'*). To assess a potential covariation between metacognitive abilities, we calculated for each subject a domain-generality index (DGI). We quantified each monkey's domain-generality, as well as their mean across the two tasks (Fig. 7C & 7D). Specifically, we shuffled the task labels (memory/perception) of all 40 days (20 days' memory and 20 days' perception) within each subject. This procedure was shuffled 1,000 times and we obtained 40,000 simulated DGI values for each monkey. We found all monkeys' DGI are above simulated data by Mann-Whitney tests against the mean of simulated data (Mars: 0.167; Saturn: 0.182; Uranus: 0.350; Neptune: 0.260; Mann-Whitney test results: all p values < 0.001, Fig. 7E). Additionally, we employed pairwise correlation to assess the similarity of two tasks across and within subjects (Fig. 7G). The matrix of pairwise correlation was hierarchically clustered (Fig. 7H), revealing two distinct clusters in which monkeys from the same domain grouped together (rather than within-monkeys data from grouped together). This indicates that within-task meta efficiency's similarity are stronger than those for within-subjects similarity. Altogether, these results suggest domain-specific constraints on metacognitive ability transcending across individual animals.



Example	Description
Within-monkey Pair *(M_Mars, P_Mars)*	*20 days meta correlation across two domains*
Across-monkey Pair *(M_Mars, M_Saturn)*	*20 days meta correlation across two monkeys*

Figure 7. Domain-specific metacognition in monkeys. Task performance in terms of percentage correct was correlated across perceptual and memory domains (A). In contrast, their metacognitive efficiency was not correlated across perceptual and memory domains (B). DGI quantifies the similarity between their metacognitive efficiency scores in each domain. Greater DGI scores indicate less metacognitive consistency across domains. The darker color indicates lower metacognitive generality across domains, the red area indicates the simulated DGI values. Daily domain generality index (DGI) is shown for each monkey (C) and for all four monkeys (D). The monkeys demonstrate a greater DGI than shuffled data (chance) (E). Two example pairs of pairwise correlation analysis are described (F). Pairwise correlation matrix indicates a pairwise correlation between every pair of monkey and domain (G). Cluster results from pairwise correlation matrix, revealing two distinct clusters where monkeys from the same domain grouped together (H). Error bars indicate ± one standard error, * indicates $p < 0.05$.

Discussion

Our findings on deficits following TMS to BA 46d demonstrate functional and biological dissociation of cognition and metacognition in animals (Lak et al., 2014; Miyamoto et al., 2017). Together with evidence on metacognitive domain-specificity, our results characterize specialization of metacognition in the primates.

The metacognition deficit by TMS revealed here is specific to the correspondence between accuracy and confidence (cf. criteria for producing subjective ratings, Rounis et al., 2010) rather than to the animals' task performance (neither RT nor accuracy). Mechanistically, TMS affects neural functioning by inducing a short-lasting electric field at suprathreshold intensities via electromagnetic induction (Valero-Cabré et al., 2017). Considering that we combined T1-weight images with the stereotaxic system, we have reliably confined our stimulation focally to BA 46d (possibly implicating subregions in dlPFC e.g., 9m, 9d, 46v, 46f) without affecting the OFC or aPFC. Our results corroborate with the human literature. The human lateral PFC has been associated with a unique type of metacognitive process—the feeling of knowing (Lapate et al., 2020). Studies inactivating the dlPFC to diminish the metacognitive ability without altering perceptual discrimination performance and confidence criteria (Rounis et al., 2010) as well as decoded multivariate patterns in the lPFC pertaining to metacognitive judgements indicate the lPFC's involvement in conscious experiences (Morales et al., 2018). Our results confirmed that the dorsal part of lPFC in the monkeys plays a critical role in mediating perceptual experiences. We should note that the metacognitive functions of lPFC are distinct from those neuronal activities in the LIP (Kiani and Shadlen, 2009), supplementary eye field (SEF) (Middlebrooks and Sommer, 2012), and middle temporal visual area (MT) (Fetsch et al., 2014), which have been shown to carry information that correlate with both perceptual decision and metacognition indistinguishingly. Our results are in line with the views that a general role of dlPFC lies in information monitoring and maintaining (Fleck et al., 2006; Fleming and Dolan, 2012). It is possible that the neural signal is an inter-status form first-order representations to high-order representations (Brown et al., 2019), which enables the perceptual content to enter consciousness.

In terms of the temporal window of meta-computation, by applying TMS with high temporal resolution to the monkey dlPFC in on-judgement and on-wagering

phase, we revealed that meta-calculation processes were carried out in the relatively early-stage. This is in line with findings that LIP in monkeys computes perceptual evidence at the time of judgment (Kiani and Shadlen, 2009). However, interestingly, the human's aPFC (Pereira et al., 2019) and dorsal premotor cortex (Fleming et al., 2015) along with the rodents' OFC (Lak et al., 2014; Masset et al., 2020) support late-stage meta-calculation. For example, single neurons in rodents' OFC showed neural activities that predict trial-difficulty psychometric curve during wagering (Masset et al., 2020), indicating the role of OFC in late-stage meta-calculation. Some computational models also proposed that the post-decisional (late stage) processes are essential for meta-calculation (Van Den Berg et al., 2016; Pleskac and Busemeyer, 2010). To tap further into these issues, a recent study employing online TMS pulses on trials (three consecutive pulses: 250, 350, 450 ms after stimulus onset) on human's dlPFC showed that TMS alters subjective confidence, but not metacognitive ability (Shekar and Rahnev, 2018). By comparing their TMS timing with ours, it indicates processes necessitated for meta-calculation might have happened earlier than those required for confidence-calculation (TMS at 250ms led to deficits in confidence-calculation, whereas TMS at 100ms led to deficits in meta-calculation in our study). In this case, the dlPFC operates meta-calculation around 100ms - 250ms and permits the confidence expression at a later-stage. The very short duration (100ms - 250ms) during which meta-calculation could be affected seems to suggest meta-calculation is heuristic (Ferrigno et al., 2017).In contrast to humans, where metacognitive ability can be assessed by quantifying trial-by-trial correspondence between objective performance and subjective confidence (Tunney and Shanks, 2003; Tunney, 2005; Maniscalco and Lau, 2012; Fleming and Lau, 2014), studies on aniamls have used binary report for confidence expression like betting (Kornell, 2007; Middlebrooks and Sommer, 2011, 2012; Brown et al., 2017; Ferrigno et al., 2017; Miyamoto et al., 2017; Miyamoto et al., 2018), opt-out (Janssen and Shadlen, 2005; Kornell, 2007; Kiani and Shadlen, 2009; Kiani et al., 2014; Odegaard et al., 2018), or some secondary metrics such as reaction time (Weidemann and Kahana, 2016; Kwok et al., 2019) and saccadic end point (Kiani et al., 2014). However, binary reports have several shortcomings. For example, we cannot preclude the possibility that information being integrated before report, merging various putative processes underlying metacognitive control (Redford, 2010) and monitoring (Shields et al., 2005; Son and Kornell, 2005). Since the relationship between response and confidence is affected by the distribution assessments (Juslin

et al., 1996), binary or even scaled confidence report will make it impossible to obtaina confidence distribution (Ais et al., 2016). As a result, information carried in the intermediate confidence range in the calibration of confidence and accuracy will also be missed (Baranski and Petrusic, 1994; Tenney et al., 2008; Fischer et al., 2019). For these considerations, we therefore adopted Lak et al. (2014)'s paradigm and provided a quantitative and continuous proxy for confidence akin to self-reporting in humans.

The results obtained with this paradigm allowed us to address a long-standing controversy in the animal cognition literature. Previous studies established that several other species are capable of monitoring their own behavior (Smith et al., 1998; Sole et al., 2003; Hampton et al., 2004; Janssen and Shadlen, 2005; Kornell et al., 2007; Kiani and Shadlen, 2009; Rosati and Santos, 2016; Odegaard et al., 2018; Sumie Iwasaki, 2019). However, due to extensive training that is often required, animals' metacognitive ability can be confounded by various kinds of cue associations (Hampton, 2009). Importantly, with the time wagering paradigm, the monkeys' introspective capability of their memory/perception state in our studies are unlikely confounded by these associative factors. The observation that their RT is not associated with WT under normal circumstances show that monkeys did not use RT as a behavioral cue for wagering decisions (Hampton et al., 2004; Miyamoto et al., 2017). Only when area 46d was perturbed, the monkeys turned to rely on trial-wise RT as an associative cue to index confidence, potentially as a means to compensate their metacognitive deficits to some extent (note that their metacognitive scores remained above zero in all conditions). This pattern shift suggests that the monkeys might have turned to rely on public information (like behavioral cues such as RT) when their introspective ability was suppressed (Hampton, 2009), satisfying the established criterion required for animal metacognition.

Our domain-generality index and intraday correlation analysis serve to reveal the existence of such domain-specific metacognition in monkeys. The pairwise correlation shows the domain-specificity is more robust than the within-individual correlation. Behavioral studies have found that efficient metacognition in one task predicts good metacognition in another task (Baird et al., 2013a; Ais et al., 2016; Faivre et al., 2017; Samaha and Postle, 2017; Alan et al., 2018; Carpenter et al., 2019). Coexistence of domain-general and domain-specific BOLD signals has been reported in humans (Morales et al., 2018). Here, we found that the monkeys have

successfully generalized their metacognitive ability form memory to perception (or vice versa). Such generalization suggests monkeys are capable of using domain-general cues to monitor the status of cognitive processes and knowledge states (Kornell, 2007; Brown et al., 2017), carrying theoretical implications for how metacognition and decision confidence is formed in the animals.

In summary, we provided evidence for a high-level cognitive facility in a nonhuman primate species. We pinpointed a critical functional role of BA 46d in supporting metacognition without implicating their task performance and found that metacognition in macaques is highly domain-specific for memory versus perception processes.

Methods

Experimental protocol

Animals. Four male adult macaque monkeys (*Macaca Mulatta*, mean age: 6 yr, mean weight: 8.2 ± 0.4kg) took part in this study. They were initially housed in a group of 4 in a specially designed spacious enclosure (max capacity = 12–16 adults) with enrichment elements (e.g., swings and climbing structures). During the experiment, the monkeys were kept in pairs according to their social hierarchy and temperament. They were given with portions of 180 g monkey chow and pieces of fruit twice a day (9:00 am / 3:00 pm). Except on experimental days, the monkeys had unlimited access to water and were routinely given treats such as peanuts and raisins. The monkeys were procured from a nationally accredited colony in the Beijing outskirts, where the monkeys were bred and reared. The room in which they are housed has a 12-hour light-dark cycle and is kept at a temperature of 18–23°C with a humidity of 60–80%. The experimental protocol was approved by the Institutional Animal Care and Use Committee (permission code: M020150902 & M020150902-2018) at East China Normal University.

Behavioral tasks

Perception task. We used resolution difference judgement as our perceptual task (Ye et al., 2018), see Fig. 1B. The monkeys began a perceptual trial by touching a blue rectangle in the center of the screen (which served as a self-paced start-cue), and after a variable delay duration (1 – 6 s), two pictures (which differed in resolution and shrunk in both length and width) were displayed bilaterally on the screen. The monkeys were trained to choose and hold onto the target picture (either higher or lower resolution; counterbalanced across monkeys). In order to maintain stable cognitive performance across days, we controlled cognitive performance using a 4 up – 1 down staircase procedure with resolution difference as a variable.

Memory task. We used temporal order judgement as our mnemonic task (Zuo et al., 2020). Monkeys initiated a mnemonic trial by touching a red rectangle in the center of the screen, and following a 4-s clip and a variable delay duration (1 – 6 s) two frames extracted from the clip were displayed bilaterally on the screen. Monkeys were trained to choose and hold onto the frame that was shown earlier in the clip. Memory

and perception tasks shared the same picture pool, which allowed us to avoid the interference of stimuli context, allowing us a matched comparison of memory and perception tasks.

*TMS experiment design (only perceptual test), time schedule, and preliminary training***.** Uranus and Neptune received 20 days of meta-perception testing with single pulse TMS intervention (Uranus: 2303 trials, Neptune: 2321 trials). There were two experimental factors. The first factor was TMS stimulation condition in which TMS was either administered on their BA 46d (right hemisphere) or TMS was a sham. The second factor was the timing of TMS: In the on-judgement condition, the monkeys received a single pulse 100ms after stimulus onset. In the on-wagering condition, the monkeys received a single pulse 100ms after they made their decision, see Fig. 1B. The timing conditions were completed in two with-session blocks (on-judgement, on-wagering) with an interval of 5 minutes between them. The order of TMS-46d/sham and on-judgement/on-wagering was counterbalanced within- and across-monkeys (Fig. 1A). The TMS experiment was conducted 10 months after the domain-comparison experiment.

*Domain-comparison experiment design, time schedule, and preliminary training***.**. The monkeys were tested for 20 days in the meta-memory task (Saturn: 2165 trials, Neptune: 2196 trials; Mars: 1694 trials, Uranus: 2200 trials), and 20 days for the meta-perception task (Saturn: 1923 trials, Neptune: 2061 trials; Mars: 1851 trials, Uranus: 2087 trials). The testing order for the two tasks was counterbalanced across monkeys: Saturn and Neptune (meta-memory task → meta-perception task), whereas Mars and Uranus (meta-perception → meta-memory). A daily session required the animals to complete 120 trials. All monkeys completed the testing, except Mars who did not complete enough trials during the meta-memory task in some days. We accordingly conducted extra 10 days of testing on Mars to complete the number of trials required.

TMS protocol. Single pulse TMS (monophasic pulses, 100 µs rise time, 1 ms duration) was applied using a Magstim 200 Stimulator (Magstim, UK) with an MC-B35 butterfly coil with 35-mm circular components. Based on feasibility analysis of cross-species TMS comparison (Rossi et al., 2009; Alekseichuk et al., 2019), we made use of smaller coils for more focal induced electric fields to compensate for the smaller head size in monkeys (Deng et al., 2013). The pulse intensity was at 120% of

resting motor threshold (rMT), which was defined as the lowest TMS intensity delivered over the right motor cortex necessary to elicit visible twitches in at least 5 of 10 consecutive pulses (Rossini et al., 2015). For the stability of the TMS set up, a headpost (Crist Instruments) was implanted on the monkey's skull with nonmagnetic material screws. The TMS coil was held in place by means of an adjustable metal arm. In the sham condition, we rotated the coil 90 degrees, and placed the coil in BA 46d as well, thereby ensuring that the sound and vibration (by-products) of the stimulation were identical between the TMS-46d and sham conditions.

*Stimulation sites and localization procedure***.** The structural T1-weighted images from post-training MRI scanning were used to enable subject-specific neuro-navigation. Brainsight 2.0, a computerized frameless stereotaxic system (Rogue Research), was used to localize the target brain regions. To determine the area of BA 46d in each monkey, we first performed non-linear registration of T1w image to the D99 atlas and resampled the D99 macaque atlas in native space (Reveley et al., 2017). Then using the same atlas to define each monkey's BA 46d. We uploaded each monkey's mask of BA 46d into the system along with the T1-weighted images for navigation. The stimulated site was located in the BA 46d (coordinate x = 13, y = 16, z = 12 in monkey atlas) for each monkey (Fig. 1D). To align each monkey's head with the MRI images, the location information of each monkey's head was obtained individually by touching three fiducial points, which are the nasion, and the intertragal notch of each ear using an infrared pointer. The real-time locations of reflective markers attached to the coil and the subject were monitored by an infrared camera with a Polaris Optical Tracking System (Northern Digital).

Requirement for reward delivery and post-decision confidence measured by wagering time (WT). Our study measured monkeys' confidence via a post-decision time-based wagering paradigm. Following the monkey's perceptual or mnemonic decision, they needed to hold onto the target (instead of touch) to initiate a waiting process. Monkeys would get a reward (2 ml water) by having chosen the correct picture *and* having waited till the required WT set for that trial. The required WT for each trial was drawn from an exponential distribution with a decay constant equal to 1.5 (Lak et al., 2014) and it differs from trial to trial ranging from 5250 ms to 11250 ms (setting up a value each 500 ms) (Fig. 1C). We did not impose additional punishment measures like blank screen, considering the WT itself served as an effective means for metacognitive feedback. The time duration that animals were

willing to invest in each trial for a potential reward provided us with a quantitative measure of his trial-wise decision confidence. We included catch trials (approximately 20% of correct trials) to reflect the maximum amount of wagering time, similar to a previous study (Lak et al., 2014). In catch trials, we delivered the liquid reward after the monkeys released their hand off the screen.

Training. For the preliminary training, there were three main stages. First, we trained naïve monkeys to perform the perception and memory task separately. Note the perceptual and mnemonic tasks only require touching as their responses, we thus avoided any preliminary training on confidence expression (no touching required). Second, we introduced the requirement of hand-holding on touchscreen for reward delivery: monkeys were trained to place their hand onto the screen and subsequently obtain a water reward with a single discrimination task (choosing between a white vs. a yellow rectangle). The monkeys learned to hold onto the target for 3 s in this stage. Third, we introduced a contingency of random WTs, which contain a gradually increasing max-WT from 5 s to 12 s. Catch trials were introduced in this stage. By the time of the experiments proper, we had the monkeys combine the perception and memory tasks with the hand-holding wagering requirement from its outset.

Data analysis

In total, we have registered 4,624 trials for the TMS experiment and 16,177 trials for the domain-comparison experiment. Trials with RT longer than 10 s (6.3%) or shorter than 0.2 s (4.1%) were discarded from analysis in the domain-comparison experiment. We limited our WT-related analysis for trials with WT < 30 s (99.7 % and 98.5% of trials included in the TMS and the domain-comparison experiment respectively).

Meta-index with hierarchical Bayesian estimation (Hierarchical model meta-d′/d′). Here, we calculated the meta-d′/d′, a metric for estimating the metacognitive efficiency (level of metacognition given a level of performance or signal processing capacity) with a hierarchical Bayesian estimation method, which can avoid edge-correction confounds and enhance statistical power (Fleming and Daw, 2017). Meta-d′ is a measure of metacognitive accuracy from the empirical Type II receiver operating characteristic curve, which reflects the subject's ability to link confidence with performance. To ensure that our results were not due to any idiosyncratic

violation of the parametric assumptions of SDT, we additionally calculated a contingency index of preference for optimal choice (Kornell, 2007; Middlebrooks and Sommer, 2011) using the number of trials classified in each case [*n*(case)]:

$$Phi\ coefficient\ (\Phi) = \frac{n(Correct\ High) \times n(Incorrect\ Low) - n(Correct\ Low) \times n(Incorrect\ High)}{\sqrt{n(Correct) \times n(Incorrect) \times n(High) \times n(Low)}}$$

*Classification of high and low confidence trials***.** For the computation of meta-d′ /d′ and Phi coefficient, four types of trials and their distribution are required: high confidence/correct, low confidence/incorrect, low confidence/correct, and low confidence/incorrect. We used trial-unique required waiting time to classify every trial into high confidence and low confidence, similar to how confidence scale is classified into high and low in human studies (Fleming et al., 2014b; Morales et al., 2018; Crystal, 2019). Specifically, we designated the unreached trials (actual wagering time shorter than the required wagering time, in which case the monkeys would not obtain a reward) as low confidence trials. We designated the reached trials (actual wagering time longer than (or equal to) the required wagering time, in which case the monkeys would obtain a reward if the response was correct) as high confidence trials. We obtained one meta-d′ /d′ and one Phi coefficient per monkey per daily session.

Logistic regression to probe wagering time (WT) response-tracking precision. By running logistic regression to capture how well WT might align with accuracy at the trial level, we tested for the differences between tasks in domain-comparison experiment (memory/perception) and between the two conditions in the TMS experiment (Sham / TMS-46d) in terms of their respective WT response-tracking precision. We used only catch and incorrect trials in the logistic regression analysis.

In the domain-comparison experiment, we fit the percentage of correct response as a function of WT, task (memory/perception), and the cross-product term of WT and task to a logistic function:

$$P(correct) = \frac{1}{1 + e^{-(\beta_1 \times WT + \beta_2 \times task + \beta_3 \times WT \times task)}}$$

, where β_1 reflects the response-tracking precision of WT, β_2 reflects the difference of accuracy between two tasks, and β_3 reflects the difference of WT response-tracking precision between tasks (memory/perception).

In the TMS experiment, we fit the percentage of correct response as a function of WT, TMS condition (TMS-46d / Sham), and the cross-product term of WT and TMS condition to a logistic function:

$$P(correct) = \frac{1}{1 + e^{-(\beta_1 \times WT + \beta_2 \times TMS + \beta_3 \times WT \times TMS)}}$$

, where β_1 reflects the response-tracking precision of WT, β_2 reflects the difference of accuracy between two tasks, and β_3 reflects the difference of WT response-tracking precision between TMS conditions (TMS-46d / Sham).

Generalized linear model (GLM). We used GLMs to exam how WTs might vary as a function of task difficulty levels (see trial-difficulty psychometric curves in Fig. 5C-F). We used "*Enter*" method to include several variables and their cross-product items into the GLMs:

$$E(Y) = g^{-1}(X\beta)$$

, where the dependent variable *Y* is WT, *β* is an unknown parameter to be estimated, and *g* is a Gaussian estimated function. The independent variables *X* are: resolution-difference, a binary regressor indicating correctness, a binary regressor indicating TMS modulation (TMS-46d / TMS-Sham), a binary regressor indicating TMS-phase (on-judgement / on-wagering), and their cross-product items. *Domain-generality Index (DGI) & pairwise correlation assessing metacognitive efficiency similarity of two tasks across and within subjects. DGI* quantifies the similarity between scores in each domain (Fleming et al., 2014b) as follows:

$$DGI(Domain - generality\ index) = \left|M_p - M_M\right|$$

, where M_P is the perceptual H-model *meta-d' /d'* and M_M is the memory H-model *meta-d'/d'*. Lower DGI scores indicate more similar metacognitive efficiencies between domains (DGI = 0 indicates identical scores).

In terms of pairwise correlation matrices, we built a matrix inwhich each entry E (task, monkey) represents meta efficiency correlation of a pair of monkey and task over a period of 20 days. For example, (M_Mars, P_Mars) represents the correlation between meta efficiency of 20 days' memory task and 20 days' perception task for Mars (Fig. 7F). A single-linkage clustering method (Hastie et al., 2009) was employed to compute the minimum of the pairwise distance and generate a

hierarchical cluster. These allowed us to test whether the within-task similarity exceeds the within-subject similarity of two domains.

Apparatus. The training and testing were conducted in an automated test apparatus. The subject sat in a Plexiglas monkey chair (29.4 cm × 30.8 cm × 55 cm) fixed in position in front of an 18.5-inch capacitive touch-sensitive screen (Guangzhou TouchWo Co., Ltd, China) on which the stimuli could be displayed, and allow the monkeys to move their hands and hold onto the target. An automated water delivery reward system (5-RLD-D1, Crist Instrument Co., Inc, U.S.) delivered water through a tube positioned just beneath the mouth of the monkeys, in response to the correct choices made by the subject. Apart from the backdrop lighting from the touch screen, the entire chair was placed in an experimental cubicle that was dark. The stimulus display and data collection were controlled by Python programs on a computer with millisecond precision. An infrared camera and video recording system (EZVIZ-C2C, Hangzhou Ezviz Network Co., Ltd, China) was used to monitor the subjects.

Material. Documentary films on wild animals was gathered from YouTube and bilibili, including Monkey Kingdom (Disney), Monkey Planet (Episode 1–3; BBC), Monkey Thieves (http://natgeotv.com/asia/monkey-thieves), Monkeys: An Amazing Animal Family (https://skyvision.sky.com/programme/15753/monkeys--an-amazing-animal-family), Nature's Misfits (BBC), Planet Earth (Episode 1–11; BBC), Big Cats (Episode 1–3; BBC), and Snow Monkey (PBS Nature). A total of 36 hours of video were collected. We used Video Studio X8 (Core Corporation) to split the film into smaller clips (2s each) and CV2 package in Python to eliminate any blank-screenframes . We chose 800 2-s clips which do not contain snakes, blank screens, or altered components such as typefaces as the video pool. We extracted 1600 still frames (two frames per video: 10th and 10th last frames) from these 800 clips.

References

Abzug, Zachary, M., Sommer, Marc, A. (2018) Neuronal correlates of serial decision-making in the supplementary eye field. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 38:7280-7292.

Ais J, Zylberberg A, Barttfeld P, Sigman M (2016) Individual consistency in the accuracy and distribution of confidence judgments. Cognition 146:377-386.

Alan LF, Lee, Eugene, Ruby, Nathan, Giles, Hakwan, Lau (2018) Cross-domain association in metacognitive efficiency depends on first-order task types. Frontiers in psychology.

Alekseichuk, I., Mantell, K., Shirinpour, S., & Opitz, A. (2019). Comparative modeling of transcranial magnetic and electric stimulation in mouse, monkey, and human. Neuroimage, 194, 136-148.

Baird B, Smallwood J, Gorgolewski KJ, Margulies DS (2013a) Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. The Journal of Neuroscience 33:16657-16665.

Baird B, Smallwood J, Gorgolewski KJ, Margulies DS (2013b) Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. J Neurosci 33:16657-16665.

Baird B, Mrazek MD, Phillips DT, Schooler JW (2014) Domain-specific enhancement of metacognitive ability following meditation training. Journal of Experimental Psychology: General 143:1972.

Baranski JV, Petrusic WM (1994) The calibration and resolution of confidence in perceptual judgments. Perception & Psychophysics 55:412.

Baumann MA, Fluet M-C, Scherberger H (2009) Context-specific grasp movement representation in the macaque anterior intraparietal area. Journal of Neuroscience 29:6436-6448.

Bègue I, Vaessen M, Hofmeister J, Pereira M, Schwartz S, Vuilleumier P (2019) Confidence of emotion expression recognition recruits brain regions outside the face perception network. Social cognitive and affective neuroscience 14:81-95.

Boldt A, De Gardelle V, Yeung N (2017) The impact of evidence reliability on sensitivity and bias in decision confidence. Journal of experimental psychology: human perception and performance 43:1520.

Brown EK, Templer VL, Hampton RR (2017) An assessment of domain-general metacognitive responding in rhesus monkeys. Behav Processes 135:132-144.

Brown R, Lau H, LeDoux JE (2019) Understanding the higher-order approach to consciousness. Trends in cognitive sciences 23:754-768.

Carpenter J, Sherman MT, Kievit RA, Seth AK, Lau H, Fleming SM (2019) Domain-general enhancements of metacognitive ability through adaptive training. Journal of Experimental Psychology: General 148:51-64.

Cortese A, Amano K, Koizumi A, Kawato M, Lau H (2016) Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. Nature Communications 7.

Crystal JD (2019) Domain-general enhancements of metacognitive ability through adaptive training. Animal Behavior and Cognition 6:254-261.

Desender, K., Donner, T. H., & Verguts, T. (2021). Dynamic expressions of confidence within an evidence accumulation framework. Cognition, 207, 104522.

Deng, Z. D., Lisanby, S. H., & Peterchev, A. V. (2013). Electric field depth–focality tradeoff in transcranial magnetic stimulation: simulation comparison of 50 coil designs. Brain stimulation, 6(1), 1-13.

Dienes Z, Seth A (2010) Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. Consciousness and Cognition 19:674-681.

Faivre N, Filevich E, Solovey G, Kühn S, Blanke O (2017) Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. Journal of Neuroscience:0322-0317.

Ferrigno S, Kornell N, Cantlon JF (2017) A metacognitive illusion in monkeys. Proceedings of the Royal Society B: Biological Sciences 284:20171541.

Fetsch CR, Kiani R, Newsome WT, Shadlen MN (2014) Effects of cortical microstimulation on confidence in a perceptual decision. Neuron 83:797-804.

Fischer H, Amelung D, Said N (2019) The accuracy of German citizens' confidence in their climate change knowledge. Nature Climate Change 9.

FleckMS, Daselaar SM, Dobbins IG,Cabeza R (2006) Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and non-memory tasks. Cereb Cortex 16:1623–1630.

Fleming S, Maniscalco B, Ko Y, Amendi N, Ro T, Lau H (2014a) Action-specific disruption of perceptual confidence. Psychological Science 26.

Fleming SM, Lau H (2014) How to measure metacognition. Frontiers in Human Neuroscience 8:443.

Fleming SM, Daw ND (2017) Self-evaluation of decision-making: a general Bayesian framework for metacognitive computation. Psychological Review 124:91-114.

Fleming SM, Huijgen J, Dolan RJ (2012) Prefrontal contributions to metacognition in perceptual decision making. Journal of Neuroscience 32:6117-6125.

Fleming SM, Ryu J, Golfinos JG, Blackmon KE (2014b) Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. Brain 137:2811-2822.

Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G (2010) Relating introspective accuracy to individual differences in brain structure. Science 329:1541-1543.

Gilbert SJ, Bird A, Carpenter JM, Fleming SM, Sachdeva C, Tsai P-C (2020) Optimal use of reminders: metacognition, effort, and cognitive offloading. Journal of Experimental Psychology: General 149:501.

Hampton RR (2009) Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? Comp Cogn Behav Rev 4:17-28.

Hampton RR, Zivin A, Murray EA (2004) Rhesus monkeys (Macaca mulatta) discriminate between knowing and not knowing and collect information as needed before acting. Animal Cognition 7:239-246.

Hangya, B., Sanders, J. I., & Kepecs, A. (2016). A mathematical framework for statistical decision confidence. Neural Computation, 28(9), 1840-1858.

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction: Springer Science & Business Media.

Higham PA (2007) No Special K! A signal detection framework for the strategic regulation of memory accuracy. Journal of Experimental Psychology: General 136:1.

J.Tunney R (2005) Sources of confidence judgments in implicit cognition. Psychonomic Bulletin & Review 12:367-373.

J.Tunney R, Shanks DR (2003) Subjective measures of awareness and implicit cognition. memory & cognition 31:1060-1071.

Janssen P, Shadlen MN (2005) A representation of the hazard rate of elapsed time in macaque area LIP. Nature 8:234-241.

Juslin P, Olsson N, Winman A (1996) Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence–accuracy correlation. Journal of Experimental Psychology Learning Memory & Cognition 22:1304-1316.

Kelemen WL, Frost PJ, Weaver CA (2000) Individual differences in metacognition: Evidence against a general metacognitive ability. Memory & Cognition 28:92-107.

Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. Science 324:759-764.

Kiani R, Corthell L, Shadlen MN (2014) Choice certainty is informed by both evidence and decision time. Neuron 84:1329-1342.

Koriat A (1993) How do we know that we know? The accessibility model of the feeling of knowing. Psychological review 100:609.

Kornell L, K. Son. and Herbert S. Terrace (2007) Transfer of metacognitive skills and hint seeking in monkeys. Psychological Science 18:64-70.

Kwok SC, Macaluso E (2015) Scale invariance of temporal order discrimination using complex, naturalistic events. Cognition 140:111-121.

Kwok SC, Cai Y, Buckley MJ (2019) Mnemonic introspection in macaques is dependent on superior dorsolateral prefrontal cortex but not orbitofrontal cortex. The Journal of Neuroscience 39:5922-5934.

Lak A, Costa GM, Romberg E, Koulakov AA, Mainen ZF, Kepecs A (2014) Orbitofrontal cortex is required for optimal waiting based on decision confidence. Neuron 84:190-201.

Lapate RC, Samaha J, Rokers B, Postle BR, Davidson RJ (2020) Perceptual metacognition of human faces is causally supported by function of the lateral prefrontal cortex. Communications Biology 3.

Lau HC, Passingham RE (2006) Relative blindsight in normal observers and the neural correlate of visual consciousness. Proceedings of the National Academy of Sciences 103:18763-18768.

Maniscalco B, Lau H (2012) A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Consciousness and Cognition 21:422-430.

Maniscalco B, Mccurdy LY, Odegaard B, Lau H (2017) Limited cognitive resources explain a trade-off between perceptual and metacognitive vigilance. Journal of Neuroscience 37:1213-1224.

Masset, P., Ott, T., Lak, A., Hirokawa, J., & Kepecs, A. (2020). Behavior-and modality-general representation of confidence in orbitofrontal cortex. Cell, 182(1), 112-126.

McCurdy LY, Maniscalco B, Metcalfe J, Liu KY, de Lange FP, Lau H (2013) Anatomical coupling between distinct metacognitive systems for memory and visual perception. The Journal of Neuroscience 33:1897-1906.

Middlebrooks PG, Sommer MA (2011) Metacognition in monkeys during an oculomotor task. Journal of Experimental Psychology: Learning Memory and Cognition 37:325-337.

Middlebrooks PG, Sommer MA (2012) Neuronal correlates of metacognition in primate frontal cortex. Neuron 75:517-530.

Miyamoto K, Setsuie R, Osada T, Miyashita Y (2018) Reversible silencing of the frontopolar cortex selectively impairs metacognitive judgment on non-experience in primates. Neuron 97:980-989.e986.

Miyamoto K, Osada T, Setsuie R, Takeda M, Tamura K, Adachi Y, Miyashita Y (2017) Causal neural network of metamemory for retrospection in primates. Science 355:188-193.

Morales J, Lau H, Fleming SM (2018) Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. The Journal of Neuroscience 38:3534-3546.

Nelson TO, Gerler D, Narens L (1984) Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. Journal of Experimental Psychology: General 113:282.

Nichols T, Brett M, Andersson J, Wager T, Poline J-B (2005) Valid conjunction inference with the minimum statistic. Neuroimage 25:653-660.

Norman E, Price MC (2015) Measuring consciousness with confidence ratings.159-180.

Odegaard, B., Knight, R. T., & Lau, H. (2017). Should a few null findings falsify prefrontal theories of conscious perception? Journal of Neuroscience, 37(40), 9593-9602.

Odegaard B, Grimaldi P, Cho SH, Peters MAK, Lau H, Basso MA (2018) Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. Proc Natl Acad Sci U S A 115:E1588-E1597.

Peters MA, Lau H (2015) Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. Elife 4:e09651.

Petrides M, Pandya D (1999) Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. European Journal of Neuroscience 11:1011-1036.

Pleskac TJ, Busemeyer JR (2010) Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. Psychological Review 117:864-901.

Proust Jl (2019) From comparative studies to interdisciplinary research on metacognition. Animal Behavior and Cognition 6:309-328.

Rahnev D, Koizumi A, McCurdy LY, D'Esposito M, Lau H (2015) Confidence Leak in Perceptual Decision Making. Psychol Sci 26:1664-1680.

Rahnev, D., Nee, D. E., Riddle, J., Larson, A. S., & D'Esposito, M. (2016). Causal evidence for frontal cortex organization for perceptual decision making. Proceedings of the National Academy of Sciences, 113(21), 6059-6064.

Reveley C, Gruslys A, Ye FQ, Samaha J, Glen D, Russ B, Saad Z,Seth A, Leopold DA, Saleem KS (2017). Three-dimensional digital template atlas of the macaque brain. Cerebral Cortex 27: 4463-4477

Redford JS (2010) Evidence of metacognitive control by humans and monkeys in a perceptual categorization task. J Exp Psychol Learn Mem Cogn 36:248-254.

Rosati AG, Santos LR (2016) Spontaneous metacognition in rhesus monkeys. Psychol Sci 27:1181-1191.

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & Safety of TMS Consensus Group. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clinical neurophysiology, 120(12), 2008-2039.

Rouault M, McWilliams A, Allen MG, Fleming SM (2018) Human metacognition across domains: insights from individual differences and neuroimaging. Personality neuroscience 1.

Rounis E, Maniscalco B, Rothwell JC, Passingham RE, Lau H (2010) Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. Cognitive Neuroscience 1:165-175.

Samaha J, Postle BR (2017) Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. Proc Biol 284:20172035.

Shekhar M, Rahnev D (2018) Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. The Journal of Neuroscience 38:5078-5087.

Shields WE, Smith JD, Guttmannova K, Washburn DA (2005) Confidence judgments by humans and rhesus monkeys. Journal of General Psychology 132:165-186.

Smith DM, Mizumori SJ (2006) Learning-related development of context-specific neuronal responses to places and events: the hippocampal role in context processing. Journal of Neuroscience 26:3154-3163.

Smith JD, Shields WE, Allendoerfer KR, Washburn DA (1998) Memory monitoring by animals and humans. Journal of Experimental Psychology General 127:227.

Sole LM, Shettleworth SJ, Bennett PJ (2003) Uncertainty in pigeons. Psychonomic Bulletin & Review 10:738-745.

Son LK, Kornell N (2005) Metacognitive judgments in rhesus macaques: Explicit versus implicit mechanisms. Missing Link in Cognition.

Stolyarova A, Rakhshan M, Hart. EE, O'Dell TJ, Peters MAK, Lau H, Soltani AI, A. (2019) Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. Nat Commun.

Sumie Iwasaki HK, and Kazuo Fujita (2019) Pigeons Show Metamemory by Requesting Reduced Working Memory Loads. animal Behavior and Cognition 6:247-253.

Tenney ER, Spellman BA, Maccoun RJ (2008) The benefits of knowing what you know (and what you don't): How calibration affects credibility. Journal of Experimental Social Psychology 44:0-1375.

Valero-Cabré, A., Amengual, J. L., Stengel, C., Pascual-Leone, A., & Coubard, O. A. (2017). Transcranial magnetic stimulation in basic and clinical neuroscience: A comprehensive review of fundamental principles and novel insights. Neuroscience & Biobehavioral Reviews, 83, 381-404.

Van Den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. Elife, 5, e12192.

Vo VA, Li R, Kornell N, Pouget A, Cantlon JF (2014) Young children bet on their numerical skills: metacognition in the numerical domain. Psychological Science 25:1712-1721.

Weidemann CT, Kahana MJ (2016) Assessing recognition memory using confidence ratings and response times. Royal Society open science 3:150670.

Ye Q, Zou F, Lau H, Hu Y, Kwok SC (2018) Causal evidence for mnemonic metacognition in human precuneus. The Journal of Neuroscience 38:6379-6387.

Zuo S, Wang L, Shin JH, Cai Y, Zhang B, Lee SW, Appiah K, Kwok SC (2020) Behavioral evidence for memory replay of video episodes in the macaque. Elife 9: e54519.

Table 1. Percentiles of each monkey's meta-scores compared with the simulated data. Inferential statistics performed using a minimum statistics method show that all monkey's meta-scores are significantly higher than chance level.

	Memory		Perception	
Monkey	Phi	II-model meta d'/d'	Phi	II-model meta d'/d'
Mars	99	79	99	99
Saturn	97	83	99	98
Uranus	98	86	99	94
Neptune	99	80	99	99
Statistics	$0.03^{4} < 0.001$	$0.17^{4} < 0.001$	$0.01^{4} < 0.001$	$0.06^{4} < 0.001$

Table 2. Logistic regression individual fitting result in TMS experiment. Logistic regression on response (correct / incorrect) with WT, TMS (TMS-46d / TMS-Sham), and cross-product item as factors to test whether TMS on BA 46d will affect the WT response-tracking ability. Logistic regression was performed for on-judgement phase and on-wagering phases separately for each monkey.

Coefficients	Estimate	Standard Error	Odds Ratio	z	p
	Monkey = Neptune	TMS phase = on judgement			
(Intercept)	-1.929	0.405	0.145	-4.767	<0.001
WT	0.149	0.030	1.160	4.947	<0.001
TMS	1.989	0.532	7.309	3.735	<0.001
WT * TMS	-0.146	0.040	0.864	-3.643	<0.001
	Monkey = Uranus	TMS phase = on judgement			
(Intercept)	-1.930	0.328	0.145	-5.881	<0.001
WT	0.174	0.031	1.190	5.541	<0.001
TMS	1.905	0.492	6.719	3.875	<0.001
WT * TMS	-0.175	0.048	0.83	-3.670	<0.001
	Monkey = Neptune	TMS phase = on wagering			
(Intercept)	-1.816	0.400	0.163	-4.539	<0.001
WT	0.147	0.028	1.158	5.206	<0.001
TMS	-0.138	0.579	0.871	-0.239	0.811
WT * TMS	0.008	0.041	1.008	0.184	0.854
	Monkey = Uranus	TMS phase = on wagering			
(Intercept)	-1.867	0.336	0.155	-5.551	<0.001
WT	0.175	0.032	1.191	5.439	<0.001
TMS	-0.345	0.541	0.708	-0.638	0.524
WT * TMS	0.048	0.054	1.049	0.883	0.377

Table 3. Logistic regression individual fitting result in domain-comparison experiment. Logistic regression on response (correct / incorrect) with WT, Task (memory / perception), and cross-product item as factors to test whether monkeys can track their response with WT. The results show the monkeys were able to track their response outcome with WT. Logistic regression was performed separately for each monkey.

Coefficients	Estimate	Standard Error	Odds Ratio	z	p
		Monkey = Mars			
(Intercept)	-1.082	0.225	0.339	-4.800	<0.001
Task	0.365	0.328	1.440	1.112	0.266
WT	0.087	0.021	1.091	4.207	<0.001
Task * WT	-0.027	0.028	0.973	-0.974	0.330
		Monkey = Saturn			
(Intercept)	-1.127	0.196	0.324	-5.746	<0.001
Task	-0.053	0.347	0.948	-0.153	0.879
WT	0.102	0.025	1.107	4.081	<0.001
Task * WT	-0.002	0.037	0.998	-0.046	0.964
		Monkey = Uranus			
(Intercept)	-1.435	0.178	0.238	-0.8060	<0.001
Task	0.530	0.279	1.699	1.898	0.058
WT	0.071	0.018	1.074	3.914	< .001
Task * WT	-0.016	0.023	0.985	-0.668	0.504
		Monkey = Neptune			
(Intercept)	-1.428	0.166	0.240	-8.596	<0.001
Task	0.685	0.274	1.984	2.499	0.012
WT	0.031	0.011	1.032	2.825	0.005
Task * WT	0.003	0.018	1.003	0.187	0.851