Time-sensitive prefrontal involvement in associating confidence with task performance illustrates metacognitive introspection in monkeys

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.

***Corresponding authors**

E-mail: sze-chai.kwok@st-hughs.oxon.org

E-mail: yytang0522@gmail.com

Abstract

Metacognition refers to the ability to be aware of one's own cognition. Ample evidence indicates 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 macaque monkeys. In the present study, we investigated the functionality of macaque metacognition by combining a confidence proxy, hierarchical Bayesian meta-d′ computational modelling, and a single-pulse transcranial magnetic stimulation technique. We found that Brodmann area 46d (BA46d) played a critical role in supporting metacognition independent of task performance; we also found that the critical role of this region in meta-calculation was time-sensitive. Additionally, we report 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 to humans. Ample evidence indicates that neural underpinnings supporting metacognitive abilities are different from cognitive processes[1-9]. A number of human transcranial magnetic stimulation (TMS) studies have implicated the dorsolateral prefrontal cortex (dlPFC) in meta-perceptual judgements more than in perceptual judgements[10-12]. This evidence indicates that the prefrontal cortex, especially the lateral prefrontal cortex (lPFC), is a key region in the metacognitive mechanism[8,13,14].

Less understood, however, is whether the importance of dlPFC is conserved in other species, such as nonhuman primates. Only one extant study has investigated the role of macaques' dlPFC in meta-perceptual processes. That study found that in a visual-oculomotor task, single neurons in the dlPFC encode metacognitive components of decision-making[15]. We sought to expand on the findings of that study; our first aim was to test for any functional role of the monkey dlPFC in meta-perception independent of perception itself. To achieve this goal, we applied single-pulse transcranial magnetic stimulation to the dlPFC (BA46d) of monkeys while they performed a perceptual resolution judgement task. We adopted a temporal wagering paradigm to measure the animals' trial confidence in each trial[16-18]. Following each perceptual decision, the animals were required to wait for an unknown and variable period by keeping their hand on the screen before they qualified for any possible reward. The amount of time wagered on their decision in a given trial was used as a proxy for confidence in the decision.

Taking advantage of single-pulse 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[19]. However, single-pulse TMS of the dorsal premotor cortex (PMd) impairs confidence reports in both the pre-response and post-response windows[20], suggesting that late-stage evidence accumulation might also be required for metacognitive processes. To more precisely determine the critical phase in which meta-calculation takes place, we included two time-sensitive TMS conditions: on-judgement and on-wagering stimulation. Specifically, we applied TMS either 100 ms after stimulus onset (on-judgement phase) or 100 ms after the animal's decision (on-wagering phase). 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., concurrent 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[21-26]. For example, monkeys' ability to transfer their metacognitive judgement from a perceptual test to a memory test shows that they can employ domain-general signals to monitor the status of cognitive processes and knowledge levels[27,28], suggesting that metacognition is generalized across domains. However, mounting anatomical[3,29], functional[6], and neuropsychological[4,30,31] evidence in the human research literature increasingly points to the domain specificity of

metacognition, indicating that humans possess specialized metacognitive abilities for different domains[6,21,31,32,33]. Here, we posed the question of whether macaques show domain-specific components of metacognition[27]. To this end, we trained two additional monkeys to perform a temporal-memory task in combination with the wagering task. Making use of the data collected in both experiments, we assessed both the covariation and the divergence between metacognitive abilities in the two domains.

Results

Metacognition in monkeys in both the memory and perception domains

To show that macaques are capable of metacognition, we quantified this capacity using bias-free metacognitive efficiency (H-model *meta-d′/d′*). We compared animals' scores to zero using one-sample *t* tests and found that the meta-index values of all monkeys were above zero for both tasks (Figure 2c & d; 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 the 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 on each monkey. By randomly shuffling all the pairings between "responses" (correct/incorrect) and their corresponding "confidence levels" (high/low) within each subject, we generated 2,000 random pairings for each animal and simulated 4,000 metacognitive scores per animal (both the H-model *meta-d′/d′* and the phi coefficient). These scores represent cases in which the animals had no metacognitive ability. We then tested these simulated scores against animals' actual scores using a minimum statistic method[34]; we found that the animals indeed performed significantly above chance metacognitive ability in both tasks (all p values < 0.001; Table 1).

As a control to rule out any possible contribution of training effects, we compared the animals' metacognition scores between the first ten days and the second ten days of testing. We found no difference between the first ten days and the second ten days of metacognitive performance in either perception (H-model *meta-d′/d′*: ($t_{(39)}$ = -0.314, p = 0.755) or memory (H-model *meta-d′/d′*: ($t_{(39)}$ = 0.89, p = 0.378). These results show that the metacognitive ability of the animals 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).

TMS of BA46d impairs metacognitive performance but not cognitive performance

We then turned to our main question. We tested whether TMS of BA46d would affect metacognition on perceptual decision-making. 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 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 of BA46d than following sham treatment 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 Figure 3a and b. We replicated the metacognition deficit in the on-judgement phase with the 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 and reflect how meaningful a subject's confidence (reflected here by time wagering) is in distinguishing between correct and incorrect responses. Accordingly, we performed a three-way ANOVA (TMS phase: on-judgement/on-wagering × TMS: TMS-46d/TMS-sham × Confidence: unreached/reached) on task performance (accuracy) and observed 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 stronger in the on-judgement TMS phase (TMS × Confidence interaction: Neptune, $F_{(1,1167)} = 10.672$, $p = 0.001$; Uranus $F_{(1,1160)} = 10.404$, $p < 0.001$, Figure 3c) than in the on-wagering TMS phase (TMS × Confidence interaction: Neptune, ($F_{(1,1146)} = 0.003$, $p = 0.954$; Uranus $F_{(1,1135)} = 0.309$, $p = 0.579$; Figure 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 U tests: Neptune, $p = 0.001$; Uranus, $p < 0.001$) but not in the reached trials (Mann–Whitney U tests: Neptune, $p = 0.235$; Uranus, $p = 0.192$). These findings confirmed that TMS targeting BA46d impairs metacognitive ability on a trial-by-trial level.

We further verified that type 1 task performance and mean wagered time were not affected by TMS. As expected, task performance (daily accuracy), reaction time (RT), and wagered time (WT) were not different between the two TMS conditions in either the on-judgement phase (paired *t* test, all *p* values > 0.1 for accuracy, RT, and WT in both monkeys) or the on-wagering phase (paired *t* test, all *p* values > 0.1 for accuracy, RT, and WT in both monkeys). These findings confirmed our first hypothesis that the monkey dlPFC is critical for meta-perception and that such effects are independent of perception processes.

Instantiation of TMS-induced impairment: Reduced accuracy-tracking ability of wagered time, altered reaction time–wagered time association, and altered trial-difficulty psychometric curve

We examined whether TMS would affect the ability of WT to track task performance in the two TMS phases (on-judgement/on-wagering). We focused our analysis on catch trials and incorrect trials, since we could not measure 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 of BA46d affected the response-tracking precision

of WT. We found a significant interaction between TMS and WT in the on-judgement TMS phase (both monkeys: β_3 = -0.149, standard error = 0.029, odds ratio = 0.862, z= -5.115, p < 0.001, Figure 3e) but not during the on-wagering phase (both monkeys: β_3 = 0.010, standard error = 0.030, odds ratio = 1.010, z= 0.321, p = 0.748, Figure 3f). This effect in the on-judgement phase was driven by higher WT in correct trials than in incorrect trials in the TMS-sham condition (Mann–Whitney U tests: Neptune, p < 0.001; Uranus, p < 0.001, Figure 3i and j) but not in the TMS-46d condition (Mann–Whitney U tests: Neptune, p = 0.98; Uranus, p = 0.45, Figure 3g). We also confirmed that WT can predict the trial outcomes in a graded manner in the 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 of BA46d, when administered during the on-judgement phase, affects metacognitive performance. We obtained the same results when we performed these logistic regressions on the two monkeys separately (Table 2).

Second, metacognitive abilities in animals are often confounded by behavioural association[35]. For example, animals are believed to make use of cues (environmental cues such as stimulus conditions and self-generated cues such as response time) to determine 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 whether the monkeys relied on RT as an associative cue to determine confidence. The results showed no correlation between RT and WT correlation in the domain-comparison experiment (Figure 4a), indicating that the macaques did not rely on RT as an associative cue to determine their WT. We then utilized this phenomenon to verify the effect of TMS. WT was significantly negatively correlated with RT during the

on-judgement TMS phase only in the TMS-46d condition (r = -0.195, p < 0.001) and not in the TMS-sham condition (Figure 4b). We found a significant difference in correlation coefficients between TMS-46d and TMS-sham in the 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 (Figure 4c).

Moreover, as seen in the rodent literature, WT can be expressed as a function of the strength of evidence (e.g., odour mixture ratio in their task) and response outcome (correct/incorrect)[18]; 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 phase), 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). Trial-difficulty psychometric curves of these results illustrated that the effects were driven by a strengthened correctness × resolution difference interaction in the TMS-sham condition (including trials in both the on-judgement TMS phase and the 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, Figure 5c).

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

Qualities of monkey metacognition: Wagered time (WT) is diagnostic of the animals' performance

To further substantiate these results, we expected that monkeys could indicate their confidence using their trial-by-trial wagered time. We showed that wagered time is diagnostic of the animals' performance using a number of analyses. First, we compared the accuracy in reached (high confidence) and unreached (low confidence) trials; chi-square tests revealed that monkeys had higher accuracy in higher-confidence trials in both meta-perception (all four monkeys: $\chi^2_{(1)} = 31.88$, $p < 0.001$; for individual monkeys: all p values < 0.05, Figure 6a) and meta-memory (all four monkeys: $\chi^2_{(1)} = 13.41$, $p <$

0.001; for individual monkeys: all p values < 0.05, Figure 6b). To test whether the WT tracked the response outcomes, we performed logistic regression on response outcomes with WT, task (memory/perception), and the cross-product as factors. We confirmed that the WT could accurately predict the trial outcome (β_1 = 0.033, standard error = 0.007, odds ratio = 1.033, z = 4.586, p < 0.001; Figure 6e). We found no interaction between task and WT (β_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 tracked the response outcomes. These results showed that the trial-wise wagered time was diagnostic of 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).

Qualities of monkey metacognition: Evidence regarding domain specificity

While we found a positive correlation between the perception and memory domains in daily individual accuracy ($r_{(80)}$ = 0.271; p = 0.0151; Figure 7a), their respective metacognitive efficiency scores did not correlate ($r_{(80)}$ = 0.1134; p = 0.3164; right panel in Figure 7b). This prompted us to examine the domain specificity with bias-free metacognitive efficiency (H-model *meta-d′/d′*). To assess the potential covariation between metacognitive abilities, we calculated a domain-generality index (DGI) for each subject. We quantified each monkey's domain generality as well as the mean across the two tasks (Figure 7c and d). Specifically, we shuffled the task types (memory/perception) across all 40 days (20 days of memory and 20 days of perception) within each subject. This procedure was shuffled 1,000 times, and we obtained 40,000

simulated DGI values for each monkey. We found that all monkeys' DGIs were above the simulated values, as confirmed by Mann–Whitney U tests against the mean of the simulated data (Mars: 0.167; Saturn: 0.182; Uranus: 0.350; Neptune: 0.260; Mann–Whitney U test results: all p values < 0.001, Figure 7e). Additionally, we employed pairwise correlation to assess the similarity of the two tasks across and within subjects (Figure 7g). The matrix of pairwise correlation was hierarchically clustered (Figure 7h), revealing two distinct clusters in which data from the same domain in multiple monkeys grouped together (whereas within-monkey data did not). This indicates that the within-task similarity of metacognitive efficiency was stronger than the within-subjects similarity. Together, these results suggest domain-specific constraints on metacognitive ability that transcend the individual animal level.

Discussion

Our findings on deficits following TMS of BA46d demonstrate functional and biological dissociation of cognition and metacognition in animals[16,36]. Together with evidence of metacognitive domain specificity, our results characterize the specialization of metacognition in primates.

The TMS-induced metacognitive deficit revealed here is specific to the correspondence between accuracy and confidence (cf. criteria for producing subjective ratings[10]) rather than to the animals' task performance (RT or accuracy). Mechanistically, TMS affects neural functioning by inducing a short-lasting electric field at suprathreshold intensities via electromagnetic induction[37]. By combining T1-weighted imaging with a stereotaxic system, we reliably confined the focus of the stimulation to

BA46d (with some stimulation possibly reaching subregions in the dlPFC, e.g., 9m, 9d, 46v, and 46f). Our results corroborate the human literature. The human lateral PFC has been associated with a unique type of metacognitive process—the feeling of knowing[14]. Studies inactivating the dlPFC to diminish metacognitive ability without altering perceptual discrimination performance and confidence criteria[10], as well as decoded multivariate patterns in the lPFC pertaining to metacognitive judgements, indicate the lPFC's involvement in conscious experiences[6]. Our results confirmed that the dorsal part of the lPFC in monkeys plays a critical role in mediating perceptual experiences. We should note that the metacognitive functions of the lPFC are distinct from the neuronal activity in the LIP[19], supplementary eye field (SEF)[15], and middle temporal visual area (MT)[38], which have been shown to carry information that correlates with both perceptual decisions and metacognition. Our results are in line with the view that the general role of the dlPFC lies in information monitoring and maintenance[39,40]. It is possible that the neural signal changes status from first-order representations to higher-order representations[8], which enables the perceptual content to enter consciousness. In terms of the temporal window of meta-computation, by applying high-temporal-resolution TMS to the monkey dlPFC in the on-judgement and on-wagering phases, we revealed that meta-calculation processes were carried out in the relatively early stage. This is in line with findings that the LIP in monkeys computes perceptual evidence at the time of judgement[19]. However, interestingly, the human aPFC[41]and dorsal premotor cortex[20] along with the rodent OFC[16,18] support late-stage meta-calculation. For example, single neurons in the OFC of rodents showed neural activity that predicted the trial-difficulty psychometric curve during wagering[18], indicating the role of the OFC in

late-stage meta-calculation. Some computational models have also proposed that post-decisional (late-stage) processes are essential for meta-calculation[42,43]. To tap further into these issues, a recent study applied online TMS pulses (three consecutive pulses: 250, 350, and 450 ms after stimulus onset) to the human dlPFC and showed that TMS alters subjective confidence but not metacognitive ability[12]. By comparing their TMS timing with ours, it can be inferred that processes necessary for meta-calculation might have happened earlier than those required for confidence calculation (TMS at 250 ms led to deficits in confidence calculation, whereas TMS at 100 ms led to deficits in meta-calculation in our study). In this case, the dlPFC performs meta-calculation at approximately 100 - 250 ms and permits the confidence expression at a later stage. The very short duration (100 – 250 ms) during which meta-calculation could be affected seems to suggest that meta-calculation is heuristic[44]. In contrast to humans, whose metacognitive ability can be assessed by quantifying trial-by-trial correspondence between objective performance and subjective confidence[45-48], studies on animals have used binary means of confidence expression such as betting[15,28,36,44,49-51], opt-out [19,49,52-54], or some secondary metrics such as reaction times[55,56] and saccadic endpoints[53]. However, binary reports have several shortcomings. For example, we cannot preclude the possibility that information is integrated before reporting, merging various putative processes underlying metacognitive control[57] and monitoring[58,59]. Since the relationship between response and confidence is affected by distribution assessments[60], binary or even scaled confidence reports will make it impossible to obtain a confidence distribution[22]. As a result, information falling within the intermediate confidence range in the calibration of confidence and accuracy will also be missed[61-63]. For these

considerations, we therefore adopted Lak et al.'s[16] 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 have established that several other species are capable of monitoring their own behaviour[19,27,52,54,64-68]. However, due to the extensive training that is often required, animals' metacognitive ability can be confounded by various types of cue associations[35]. Importantly, with the temporal wagering paradigm, the monkeys' introspective knowledge of their memory/perception state in our studies is unlikely to be confounded by these associative factors. The observation that their RT is not associated with WT under normal circumstances shows that monkeys did not use RT as a behavioural cue for wagering decisions[36,66]. Only when area 46d was perturbed did the monkeys rely on trialwise RT as an associative cue to determine confidence, potentially as a means to compensate for 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 changed their strategy to rely on external information (e.g., behavioural cues such as RT) when their introspective ability was suppressed[35], 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 that the domain specificity is more robust than the within-individual correlation. Behavioural studies have found that efficient metacognition in one task predicts good metacognition in another task[21-24,26,69]. The co-existence of domain-general and domain-

specific BOLD signals has been reported in humans[6]. Here, we found that monkeys successfully generalized their metacognitive ability from memory to perception (or vice versa). Such generalization suggests that monkeys are capable of using domain-general cues to monitor the status of cognitive processes and assess knowledge states[28,49], carrying theoretical implications for how metacognition and decision confidence are formed in animals.

In summary, we provided evidence for a high-level cognitive faculty in a nonhuman primate species. We pinpointed the critical functional role of BA46d in supporting metacognition independent of task performance, and we 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 y; mean weight: 8.2 ± 0.4 kg) took part in this study. They were initially housed in a group of 4 in a spacious, specially designed enclosure (maximum 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 individual rations 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 outskirts of Beijing, where the monkeys were bred and reared. The room in which they were housed was illuminated on a 12/12-hour light-dark cycle and was 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.

Behavioural tasks

Perception task

We used resolution difference judgement as our perceptual task[30]; see Figure 1b. The monkeys began a perceptual trial by touching a blue rectangle in the centre 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 were shrunken in both length and width) were displayed on opposite sides of the screen. The monkeys were trained to choose and hold onto the target picture (either higher or lower resolution; counterbalanced across monkeys). 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[70]. Monkeys initiated each memory trial by touching a red rectangle in the centre of the screen, and following a 4-s video clip and a variable delay duration (1–6 s), two frames extracted from the clip were displayed on opposite sides of the screen. Monkeys were trained to choose and hold onto the frame that was shown earlier in the clip. The memory and perception tasks drew from the same pool of pictures, which enabled us to avoid interference from stimulus context, allowing a matched comparison of the memory and perception tasks.

TMS experimental design (perceptual test only), 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: either TMS was administered to the right BA46d, or sham TMS was performed at the same anatomical site. The second factor was the timing of TMS: in the on-judgement condition, the monkeys received a single pulse 100 ms after stimulus onset, whereas in the on-wagering condition, the monkeys received a single pulse 100 ms after they made their decision (see Figure 1b). The timing conditions were completed in two within-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 (Figure 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 performed the meta-memory task followed by the meta-perception task, whereas Mars and Uranus performed the tasks in the opposite order. Each daily session required the animals to complete 120 trials. All monkeys completed the testing in the allotted time except for Mars, who did not complete enough trials of the meta-memory task on some days. Accordingly, we conducted an extra 10 days of testing on Mars to obtain the number of trials required.

TMS protocol

Single-pulse TMS (monophasic pulses, 100 μs rise time, 1 ms duration) was applied using a Magventure X100 (Magventure, Denmark) and an MC-B35 butterfly coil with 35-mm circular components. Based on feasibility analysis of cross-species TMS comparison[71,72], we made use of smaller coils to induce more focal electromagnetic fields to compensate for the small head size of monkeys relative to humans[73]. The pulse intensity was at 120% of the resting motor threshold (rMT), which was defined as the lowest TMS intensity that would elicit visible twitches in at least 5 of 10 consecutive pulses when delivered over the right motor cortex[74]. For the stability of the TMS setup, a headpost (Crist Instruments) was affixed to the monkey's skull with screws made of nonmagnetic material. The TMS coil was held in place by an adjustable metal arm. In the sham condition, we rotated the coil 90 degrees and still placed it over BA46d,

thereby ensuring that the sound and vibration (by-products) of the stimulation were identical between the TMS-46d and TMS-sham conditions.

Stimulation sites and localization procedure

Structural T1-weighted images from post-training MRI scanning were used to enable subject-specific neuronavigation. Brainsight 2.0, a computerized frameless stereotaxic system (Rogue Research), was used to localize the target brain regions. To determine the area of BA46d in each monkey, we first performed nonlinear registration of the T1W images to the D99 atlas and resampled the D99 macaque atlas in native space[75]. Then, the same atlas was used to define each monkey's BA46d. We uploaded each monkey's BA46d mask into the system along with the T1-weighted images for navigation. The stimulated site was located in BA46d (coordinates in monkey atlas: x = 13, y = 16, z = 12) for each monkey (Figure 1d). To align each monkey's head with the MRI scans, information on the location of each monkey's head was obtained individually by touching three fiducial points, namely, 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).

Requirements for reward delivery and post-decision confidence measured by wagered time (WT)

Our study measured monkeys' confidence via a post-decision, time-based wagering paradigm. Following a monkey's perceptual or mnemonic decision, the animal

needed to continue pressing the target (instead of merely tapping and releasing) to initiate a waiting process. The monkey would receive a reward (2 mL water) if it chose the correct picture *and* waited until 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[16], and it differed from trial to trial, ranging from 5250 ms to 11250 ms (with a new value selected every 500 ms) (Figure 1c). We did not impose additional punishment measures such as a blank screen, considering that the WT itself served as an effective means of 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 their trialwise decision confidence. We included catch trials (approximately 20% of correct trials) to reflect the maximum amount of wagered time, similar to a previous study[16]. In catch trials, we delivered the liquid reward after the monkeys released their hand off the screen.

Training

The preliminary training consisted of three main stages. First, we trained naïve monkeys to perform the perception and memory tasks separately. Note that the perceptual and mnemonic tasks require only brief touches as responses; thus, we avoided any preliminary training in confidence expression (no sustained contact required). Second, we introduced the requirement of sustained contact with the 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 rectangle and a yellow rectangle). The monkeys learned to

keep their hand on the target for 3 s in this stage. Third, we introduced a contingency of random WTs, in which the maximum WT gradually increased 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 sustained-contact wagering requirement from its outset.

Data analysis

In total, we 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 to trials with WT < 30 s (99.7% and 98.5% of trials were included in the TMS and domain-comparison experiments, respectively).

Meta-index with hierarchical Bayesian estimation (hierarchical model meta-d′/d′)

Here, we calculated meta-d′/d′, a metric for estimating metacognitive efficiency (the 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[76]. Meta-d′ is a measure of metacognitive accuracy from the empirical Type II receiver operating characteristic curve, which reflects the link between the subject's confidence and 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 the optimal choice[49,50] 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

In order to compute meta-d′/d′ and the phi coefficient, it is necessary to find the distribution of four trial types: high confidence/correct, low confidence/incorrect, low confidence/correct, and low confidence/incorrect. We used the trial-specific required waiting time to classify every trial as high confidence or low confidence, similar to the way confidence is binarized into high and low in human studies[4,6,77]. Specifically, we designated the unreached trials (where the actual wagered time was shorter than the required wagered time, in which case the monkeys would not receive a reward) as low-confidence trials. We designated the reached trials (where the actual wagered time was longer than or equal to the required wagered time, in which case the monkeys would receive 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 the response-tracking precision of wagered time (WT)

By running logistic regression to capture how well WT might align with accuracy at the trial level, we tested for differences between tasks in the domain-comparison experiment (memory/perception) and between the two conditions in the TMS experiment (TMS-sham/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 responses with a logistic function containing WT, task (memory/perception), and the cross-product of WT as items 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 in accuracy between two tasks, and β_3 reflects the difference in WT response-tracking precision between tasks (memory/perception).

In the TMS experiment, we fit the percentage of correct responses to a logistic function with WT, TMS condition (TMS-46d/sham), and the cross-product of WT and TMS as terms:

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

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

Generalized linear models (GLMs)

We used GLMs to examine how WTs might vary as a function of task difficulty levels (see trial-difficulty psychometric curves in Figure 5c-f). We used the "*Enter*" method to include several variables and their cross-products as items in 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. The DGI* quantifies the similarity between scores in each domain[4] 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 greater similarity in metacognitive efficiency between domains (DGI = 0 indicates identical scores).

In terms of pairwise correlation matrices, we built a matrix in which each entry E (task, monkey) represents the meta-efficiency correlation between a particular monkey and a particular task over a period of 20 days. For example, (M_Mars, P_Mars) represents the correlation between the meta-efficiency of the 20-day memory task and the 20-day perception task for Mars (Figure 7f). A single-linkage clustering method[78] was employed to compute the minimum pairwise distance and generate a hierarchical cluster. These allowed us to test whether the within-task similarity exceeded the within-subjects 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 the monkeys were allowed to move their hands to press and hold 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 a dark experimental cubicle. The stimulus display and data collection were controlled by Python programs on a computer with millisecond precision. An infrared camera and a video recording system (EZVIZ-C2C, Hangzhou Ezviz Network Co., Ltd, China) were used to monitor the subjects.,

Material

Documentary films on wild animals were 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). In total, we collected 36 hours of video. We used Video Studio X8 (Core Corporation) to split the film into smaller clips (2 s each), and we used the CV2 package in Python to eliminate any blank frames. We chose 800 2-s clips that did not contain snakes, blank screens, or altered components such as typefaces as the video pool. We extracted 1600 still frames (two frames per video: 10^{th} and 10^{th} last frames) from these 800 clips.

Data availability. Data is available on request.

Code availability. Data is available on GitHub.

References

1. Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J. & Rees, G. Relating introspective accuracy to individual differences in brain structure. *Science* **329**, 1541-1543 (2010).
2. Fleming, S. M., Huijgen, J. & Dolan, R. J. Prefrontal contributions to metacognition in perceptual decision making. *J. Neurosci.* **32**, 6117-6125 (2012).
3. McCurdy, L. Y. *et al.* Anatomical coupling between distinct metacognitive systems for memory and visual perception. *J. Neurosci.* **33**, 1897-1906 (2013).
4. Fleming, S. M., Ryu, J., Golfinos, J. G. & Blackmon, K. E. Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain* **137**, 2811-2822 (2014).
5. Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M. & Lau, H. Confidence leak in perceptual decision making. *Psychol. Sci.* **26**, 1664-1680 (2015).
6. Morales, J., Lau, H. & Fleming, S. M. Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *J. Neurosci.* **38**, 3534-3546 (2018).
7. Rouault, M., McWilliams, A., Allen, M. G. & Fleming, S. M. Human metacognition across domains: insights from individual differences and neuroimaging. *Pers. Neurosci.* **1**, e17 (2018).

8. Brown, R., Lau, H. & LeDoux, J. E. Understanding the higher-order approach to consciousness. *Trends Cogn. Sci.* **23**, 754-768 (2019).

9. Gilbert, S. *et al.* Optimal use of reminders: metacognition, effort, and cognitive offloading. *J. Exp. Psychol. Gen.* **149**, 501-517 (2020).

10. Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E. & Lau, H. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn. Neurosci.* **1**, 165-175 (2010).

11. Rahnev, D., Nee, D. E., Riddle, J., Larson, A. S. & D'Esposito, M. Causal evidence for frontal cortex organization for perceptual decision making. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6059-6064 (2016).

12. Shekhar, M. & Rahnev, D. Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. *J. Neurosci.* **38**, 5078-5087 (2018).

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

14. Lapate, R. C., Samaha, J., Rokers, B., Postle, B. R. & Davidson, R. J. Perceptual metacognition of human faces is causally supported by function of the lateral prefrontal cortex. *Commun. Biol.* **3**, 360 (2020).

15. Middlebrooks, P. G. & Sommer, M. A. Neuronal correlates of metacognition in primate frontal cortex. *Neuron* **75**, 517-530 (2012).

16. Lak, A. *et al.* Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron* **84**, 190-201 (2014).

17. Stolyarova, A. *et al.* Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. *Nat. Commun.* **10**, 4704 (2019).

18. Masset, P., Ott, T., Lak, A., Hirokawa, J. & Kepecs, A. Behavior- and modality-general representation of confidence in orbitofrontal cortex. *Cell* **182**, 112-126.e18 (2020).

19. Kiani, R. & Shadlen, M. N. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* **324**, 759-764 (2009).

20. Fleming, S. M. et al. Action-Specific Disruption of Perceptual Confidence. *Psychol Sci* **26**, 89–98 (2015).

21. Baird, B., Smallwood, J., Gorgolewski, K. J. & Margulies, D. S. Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *J. Neurosci.* **33**, 16657-16665 (2013).

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

23. Faivre, N., Filevich, E., Solovey, G., Kühn, S. & Blanke, O. Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. *J. Neurosci.* **38**, 263-277 (2017).

24. Samaha, J. & Postle, B. R. Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proc. Biol. Sci.* **284**, 20172035 (2017).

25. Lee, A. L. F., Ruby, E., Giles, N. & Lau, H. Cross-domain association in metacognitive efficiency depends on first-order task types. *Front. Psychol.* **9**, 2464 (2018).

26. Carpenter, J. *et al.* Domain-general enhancements of metacognitive ability through adaptive training. *J. Exp. Psychol. Gen.* **148**, 51-64 (2019).

27. Kornell, N., Son, L. K. & Terrace, H. S. Transfer of metacognitive skills and hint seeking in monkeys. *Psychol. Sci.* **18**, 64-71 (2007).

28. Brown, E. K., Templer, V. L. & Hampton, R. R. An assessment of domain-general metacognitive responding in rhesus monkeys. *Behav. Process.* **135**, 132-144 (2017).

29. Maniscalco, B., McCurdy, L. Y., Odegaard, B. & Lau, H. Limited cognitive resources explain a trade-off between perceptual and metacognitive vigilance. J. Neurosci. 37, 1213-1224 (2017).

30. Ye, Q. et al. Individual susceptibility to TMS affirms the precuneal role in meta-memory upon recollection. *Brain Struct Funct* **224**, 2407–2419 (2019).

31. Ye, Q., Zou, F., Lau, H., Hu, Y. & Kwok, S. C. Causal evidence for mnemonic metacognition in human precuneus. *J. Neurosci*. **38**, 6379-6387 (2018).

32. Kelemen, W. L., Frost, P. J. & Weaver, C. A. Individual differences in metacognition: evidence against a general metacognitive ability. *Mem. Cogn.* **28**, 92-107 (2000).

33. Vo, V. A., Li, R., Kornell, N., Pouget, A. & Cantlon, J. F. Young children bet on their numerical skills: metacognition in the numerical domain. *Psychol. Sci.* **25**, 1712-1721 (2014).

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

35. Hampton, R. R. Multiple demonstrations of metacognition in nonhumans: converging evidence or multiple mechanisms? *Comp. Cogn. Behav. Rev.* **4**, 17-28 (2009).

36. Miyamoto, K. *et al.* Causal neural network of metamemory for retrospection in primates. *Science* **355**, 188-193 (2017).

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

38. Fetsch, C. R., Kiani, R., Newsome, W. T. & Shadlen, M. N. Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* **83**, 797-804 (2014).

39. Fleck, M. S., Daselaar, S. M., Dobbins, I. G. & Cabeza, R. Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and non- memory tasks. *Cereb. Cortex* **16**, 1623-1630 (2006).

40. Fleming, S. M. & Dolan, R. J. The neural basis of metacognitive ability. *Phil. Trans. R. Soc. B* **367**, 1338–1349 (2012).

41. Pereira, M. et al. Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proc Natl Acad Sci USA* **117**, 8382–8390 (2020).

42. van den Berg, R. *et al.* Author response: a common mechanism underlies changes of mind about decisions and confidence. *Elife* **5**, e12192 (2016).

43. Pleskac, T. J. & Busemeyer, J. R. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol. Rev.* **117**, 864-901 (2010).

44. Ferrigno, S., Kornell, N. & Cantlon, J. F. A metacognitive illusion in monkeys. *Proc. R. Soc. B Biol. Sci.* **284**, 20171541 (2017).

45. Tunney, R. J. & Shanks, D. R. Subjective measures of awareness and implicit cognition. *Mem. Cogn.* **31**, 1060-1071 (2003).

46. Tunney, R. J. Sources of confidence judgments in implicit cognition. *Psychon. Bull. Rev.* **12**, 367-373 (2005).

47. Maniscalco, B. & Lau, H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* **21**, 422-430 (2012).

48. Fleming, S. M. & Lau, H. C. How to measure metacognition. *Front. Hum. Neurosci.* **8**, 443 (2014).

49. Kornell, N., Son, L. K. & Terrace, H. S. Transfer of Metacognitive Skills and Hint Seeking in Monkeys. *Psychol Sci* **18**, 64–71 (2007).

50. Middlebrooks, P. G. & Sommer, M. A. Metacognition in monkeys during an oculomotor task. *J. Exp. Psychol. Learn. Mem. Cogn.* **37**, 325-337 (2011).

51. Miyamoto, K., Setsuie, R., Osada, T. & Miyashita, Y. Reversible silencing of the frontopolar cortex selectively impairs metacognitive judgment on non-experience in primates. *Neuron* **97**, 980-989.e6 (2018).

52. Janssen, P. & Shadlen, M. N. A representation of the hazard rate of elapsed time in macaque area LIP. *Nat. Neurosci.* **8**, 234-241 (2005).

53. Kiani, R., Corthell, L. & Shadlen, Michael N. Choice certainty is informed by both evidence and decision time. *Neuron* **84**, 1329-1342 (2014).

54. Odegaard, B. *et al.* Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E1588-E1597 (2018).

55. Weidemann, C. T. & Kahana, M. J. Assessing recognition memory using confidence ratings and response times. *R. Soc. Open Sci.* **3**, 150670 (2016).

56. Kwok, S. C., Cai, Y. & Buckley, M. J. Mnemonic introspection in macaques is dependent on superior dorsolateral prefrontal cortex but not orbitofrontal cortex. *J. Neurosci.* **39**, 5922-5934 (2019).

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

58. Shields, W. E., Smith, J. D., Guttmannova, K. & Washburn, D. A. Confidence judgments by humans and rhesus monkeys. *J. Gen. Psychol.* **132**, 165-186 (2005).

59. Son, L. K. & Kornell, N. Metacognitive judgments in rhesus macaques: explicit versus implicit mechanisms in *The missing link in cognition: origins of self-reflective consciousness* (eds. Terrace, H. S. & Metcalfe, J.) 296-320 (Oxford University Press, 2005).

60. Juslin, P., Olsson, N. & Winman, A. Calibration and diagnosticity of confidence in eyewitness identification: comments on what can be inferred from the low confidence–accuracy correlation. *J. Exp. Psychol. Learn. Mem. Cogn.* **22**, 1304-1316 (1996).

61. Baranski, J. V. & Petrusic, W. M. The calibration and resolution of confidence in perceptual judgments. *Percept. Psychophys.* **55**, 412-428 (1994).

62. Tenney, E. R., Spellman, B. A. & MacCoun, R. J. The benefits of knowing what you know (and what you don't): how calibration affects credibility. *J. Exp. Soc. Psychol.* **44**, 1368-1375 (2008).

63. Fischer, H., Amelung, D. & Said, N. The accuracy of German citizens' confidence in their climate change knowledge. *Nat. Clim. Change* **9**, 776-780 (2019).

64. Smith, J. D., Shields, W. E., Allendoerfer, K. R. & Washburn, D. A. Memory monitoring by animals and humans. *J. Exp. Psychol. Gen.* **127**, 227-250 (1998).

65. Sole, L. M., Shettleworth, S. J. & Bennett, P. J. Uncertainty in pigeons. *Psychon. Bull. Rev.* **10**, 738-745 (2003).

66. Hampton, R. R., Zivin, A. & Murray, E. A. Rhesus monkeys (Macaca mulatta) discriminate between knowing and not knowing and collect information as needed before acting. *Anim. Cogn.* **7**, 239-246 (2004).

67. Rosati, A. G. & Santos, L. R. Spontaneous metacognition in rhesus monkeys. *Psychol. Sci.* **27**, 1181-1191 (2016).

68. Iwasaki, S., Kuroshima, H. & Fujita, K. Pigeons show metamemory by requesting reduced working memory loads. *Anim. Behav. Cogn*. **6**, 247-253 (2019).

69. Lee, A. L. F., Ruby, E., Giles, N. & Lau, H. Cross-Domain Association in Metacognitive Efficiency Depends on First-Order Task Types. *Front. Psychol*. **9**, 2464 (2018).

70. Zuo, S. et al. Behavioral evidence for memory replay of video episodes in the macaque. *Elife* **9**, e54519 (2020).

71. Rossi, S., Hallett, M., Rossini, P. M. & Pascual-Leone, A. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin. Neurophysiol*. **120**, 2008-2039 (2009).

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

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

74. Rossini, P. M. et al. Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N. Committee. *Clinical Neurophysiology* **126**, 1071–1107 (2015).

75. Reveley, C. *et al.* Three-dimensional digital template atlas of the macaque brain. *Cereb. Cortex* **27**, 4463-4477 (2017).

76. Fleming, S. M. & Daw, N. D. Self-evaluation of decision-making: a general Bayesian framework for metacognitive computation. *Psychol. Rev.* **124**, 91-114 (2017).

77. Crystal, J. Comparative approaches to metacognition: prospects, problems, and the future. *Anim. Behav. Cogn.* **6**, 254-261 (2019).

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

Acknowledgements

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

Author contributions

YC and SCK conceived the study. YC, ZJ, and CZ conducted the experiments. JW and YT provided the TMS equipment. YC performed the data analysis. All authors contributed to the interpretation of the results. YC wrote the first draft of the manuscript with input from ZJ, CZ, HW, and YT. YC and SCK produced the final manuscript. SCK supervised the project.

Competing interests

The authors have no conflicts of interest to declare.

Tables

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

	Memory		Perception	
Monkey	Phi	H-model meta d'/d'	Phi	H-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. Individual fitting of data from the TMS experiment by logistic regression.

Logistic regression of response (correct/incorrect) with WT, TMS (TMS-46d/TMS-sham), and a cross-product item as factors to test whether TMS of BA46d affects the ability of WT to track responses. Logistic regression was performed for the on-judgement 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. Individual fitting of data from the domain-comparison experiment by logistic regression. Logistic regression of response (correct/incorrect) with WT, task (memory/perception), and a cross-product item as factors to test whether WT tracks responses. The results show that the response outcomes were tracked by 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

Figure legends

Figure 1. Temporal structure of the TMS experiment. TMS experiment schedule with TMS-46d/sham conditions counterbalanced between monkeys (Uranus and Neptune) (**a**). Perceptual judgement task with temporal wagering. Each trial consisted of a starting (blue) cue, a delay lasting 1 ~ 6 s, and two simultaneously presented pictures. The monkeys needed to choose the picture with lower resolution (or higher resolution, counterbalanced across monkeys) by holding their hand on the touchscreen. The waiting process was initiated as soon as they laid their hand on the picture. Their confidence in the decision was measured by temporal wagering; that is, they could wait for a reward if they were confident or opt out to abort the current trial. There were two TMS conditions, which differed in the timing of stimulation. In each trial, the monkeys received a single TMS pulse either immediately after the onset of the picture stimulus (on-judgement phase) or 100 ms after they made their perceptual decision (on-wagering phase) (**b**). The required WT distribution and the actual WT distribution (only catch trials and incorrect trials) with WT bin size set to 1 s. The table depicts the classification of low-confidence trials (unreached trials) and high-confidence trials (reached trials) (**c**). An illustration of the TMS site, as indicated by the green arrows. Bottom: The green area indicates BA46d on a rendering of a macaque brain; the red disc indicates the target area (**d**).

Figure 2. Task performance and metacognitive capability remained steady across days. Plots depict daily accuracy (**a** & **c**) and metacognitive efficiency (**b** & **d**) across 20 days for four monkeys performing two tasks. Dots represent individual data points; their colours represent individual monkeys. Error bars indicate ± one standard error.

Figure 3. TMS during the on-judgement phase disrupts metacognition and the response outcome tracking ability of wagered time (WT). The monkeys demonstrated an impairment in metacognitive efficiency in the TMS-46d condition during the on-judgement phase but not during the on-wagering phase (**a**). TMS of area 46d does not affect task accuracy (**b**). Difference in accuracy between unreached trials (low confidence) and reached trials (high confidence) in the on-judgement phase and the on-wagering phase (**c** & **d**, respectively). The trendlines are fitted to accuracy by logistic regression with WT as a factor for the TMS-sham and TMS-46d conditions separately. WT reliably tracks response outcomes in the TMS-sham condition but not in the TMS-46d condition during the on-judgement phase. WT tracks response outcomes in both the TMS-sham and TMS-46d conditions during the TMS on-wagering phase (**e** & **f**). Distributional differences between correct and incorrect WT. The largest effects were observed in the TMS-sham condition, in which the BA46d was not perturbed (**g**-**j**). The WT bin size was set to 1 s; coloured lines indicate kernel density estimation. Error bars indicate ± one standard error; * indicates $p < 0.05$. $\otimes$ indicates a significant interaction effect ($p < 0.05$) of WT and TMS (TMS-46d/sham). Shaded areas indicate bootstrap-estimated 95% confidence intervals for the regression estimates.

Figure 4. On-judgement TMS alters the correlation between reaction time (RT) and wagered time (WT). No correlation was found between RT and WT in the domain-comparison experiment (**a**). The Pearson correlation between RT and WT during the on-judgement phase was statistically significant for the TMS-46d condition ($p < 0.001$) but not significant for the TMS-sham condition (**b**). The correlations during the on-wagering phase were not significant for either TMS condition (**c**).

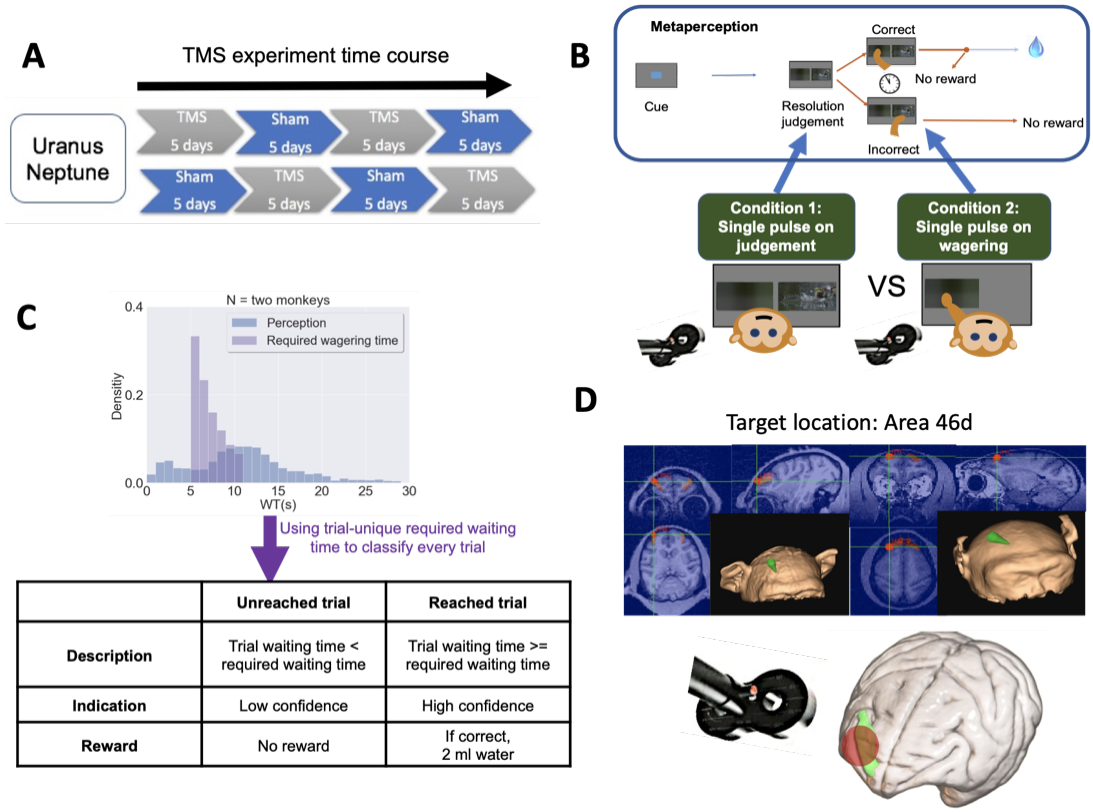
Figure 5. On-judgement TMS distorts the trial-difficulty psychometric curve. Accuracy decreases with task difficulty (resolution difference; higher values indicate lower task difficulty). The lines are logistic regression fits for accuracy with resolution difference as a factor, calculated separately for the TMS-sham and TMS-46d conditions in the on-judgement phase (**a**) and on-wagering phase (**b**). WT decreased with task difficulty in correct trials and increased with task difficulty in incorrect trials in all control conditions (**d**-**f**), but this pattern was absent in the on-judgement phase of the TMS-46d condition (**c**). Shaded areas indicate bootstrap-estimated 95% confidence intervals for the regression estimates.

Figure 6. Wagered time reflects monkeys' task performance (correctness) in both memory and perception tasks. Difference in accuracy between unreached trials and reached trials in the perception (**a**) and memory tasks (**b**). Differences between the WTs of correct and incorrect trials for each monkey in the perception (**c**) and memory tasks (**d**). WT tracks response outcome (correct/incorrect) in both memory and perception tasks. The lines are logistic regression fits for accuracy with WT as a factor. The WT bin size was set to 1 s; coloured lines indicate kernel density estimation (**e**). Error bars indicate ± one standard error; * indicates $p < 0.05$. Shaded areas indicate bootstrap-estimated 95% confidence intervals for the regression estimates.

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**). The DGI quantifies the similarity between their metacognitive efficiency scores in each domain. Greater DGI scores indicate less metacognitive

consistency across domains. Darker colours indicate lower metacognitive generality across domains, and the red area indicates the simulated DGI values. The 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 for pairwise correlation analysis are described (**f**). The pairwise correlation matrix indicates a pairwise correlation between each monkey and each domain (**g**). Cluster results from the pairwise correlation matrix, revealing two distinct clusters in which data from the same domain grouped together (**h**). Error bars indicate ± one standard error; * indicates $p < 0.05$.

Figures



	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

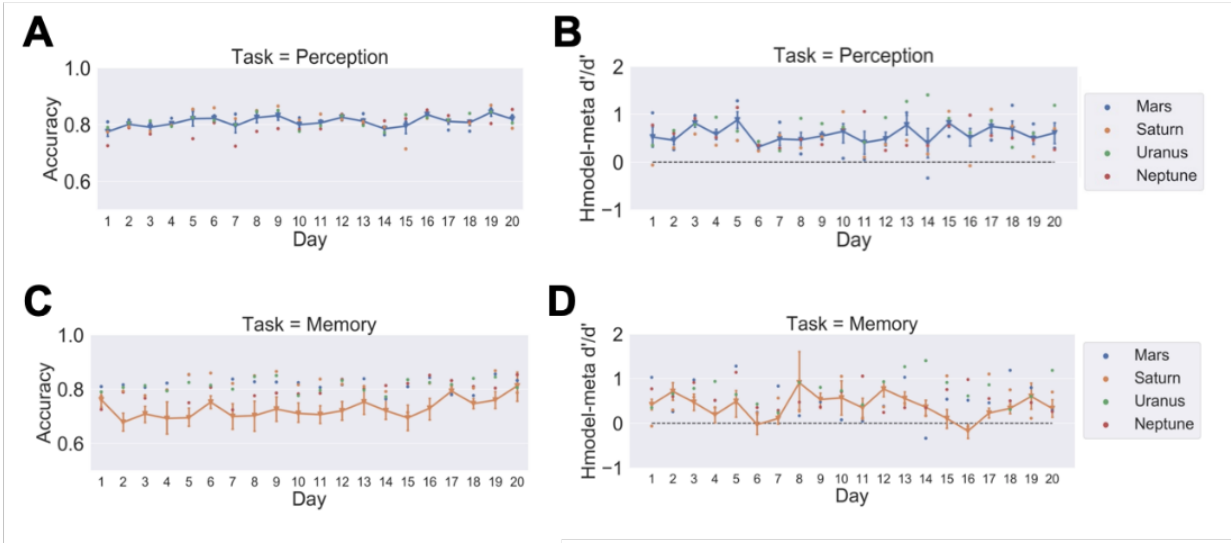


Figure 2

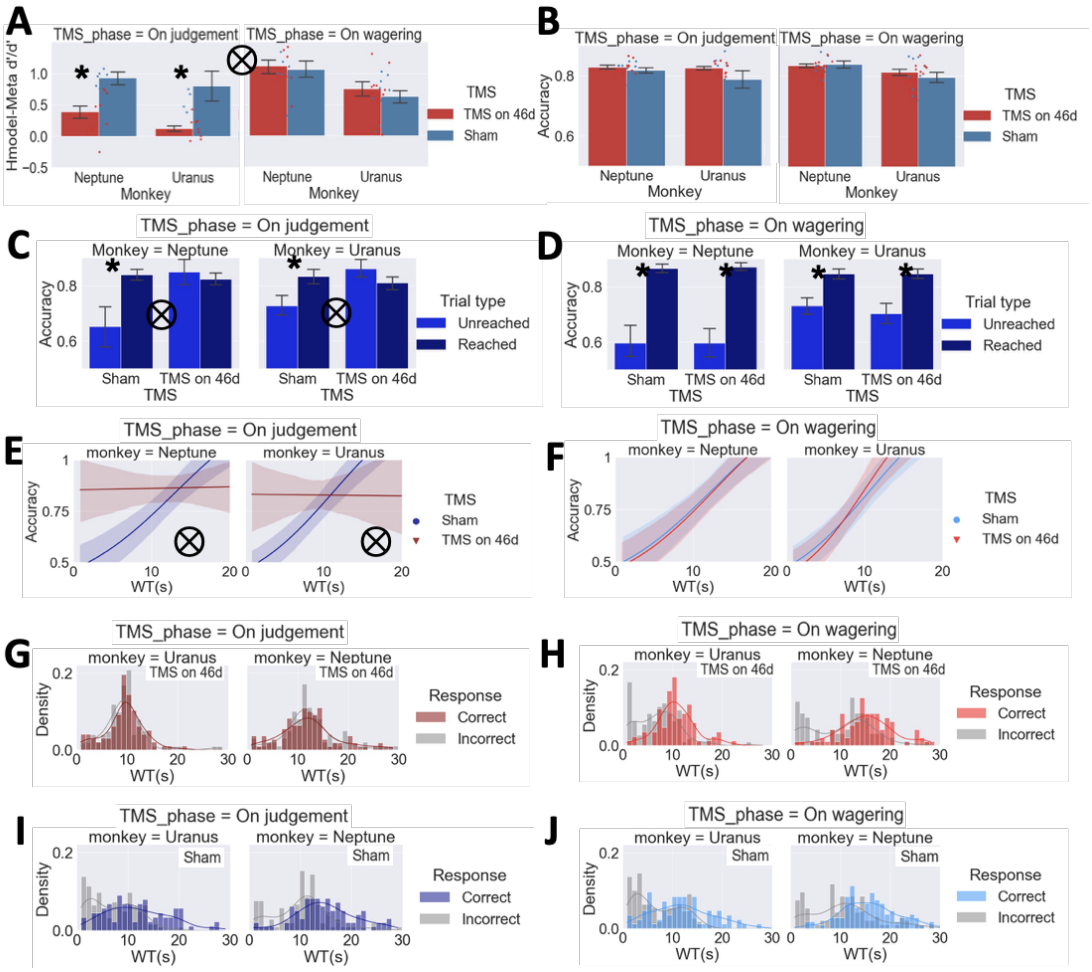


Figure 3

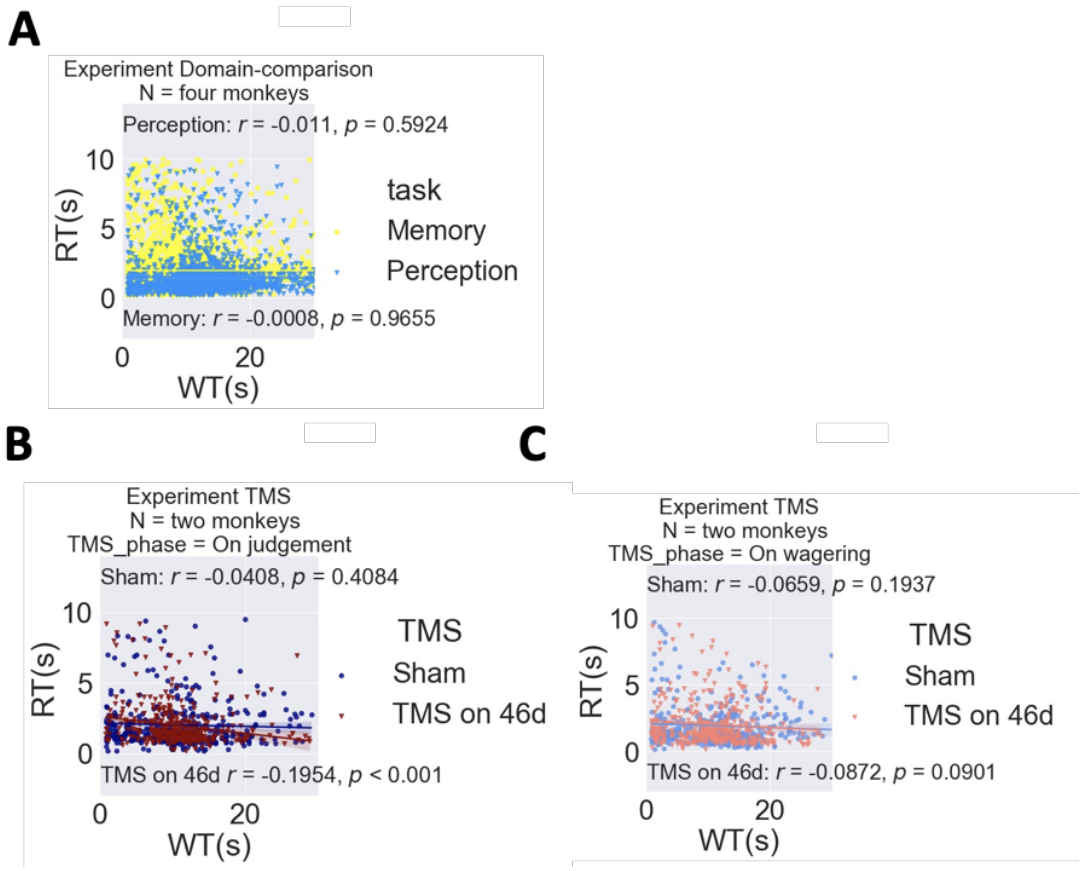


Figure 4

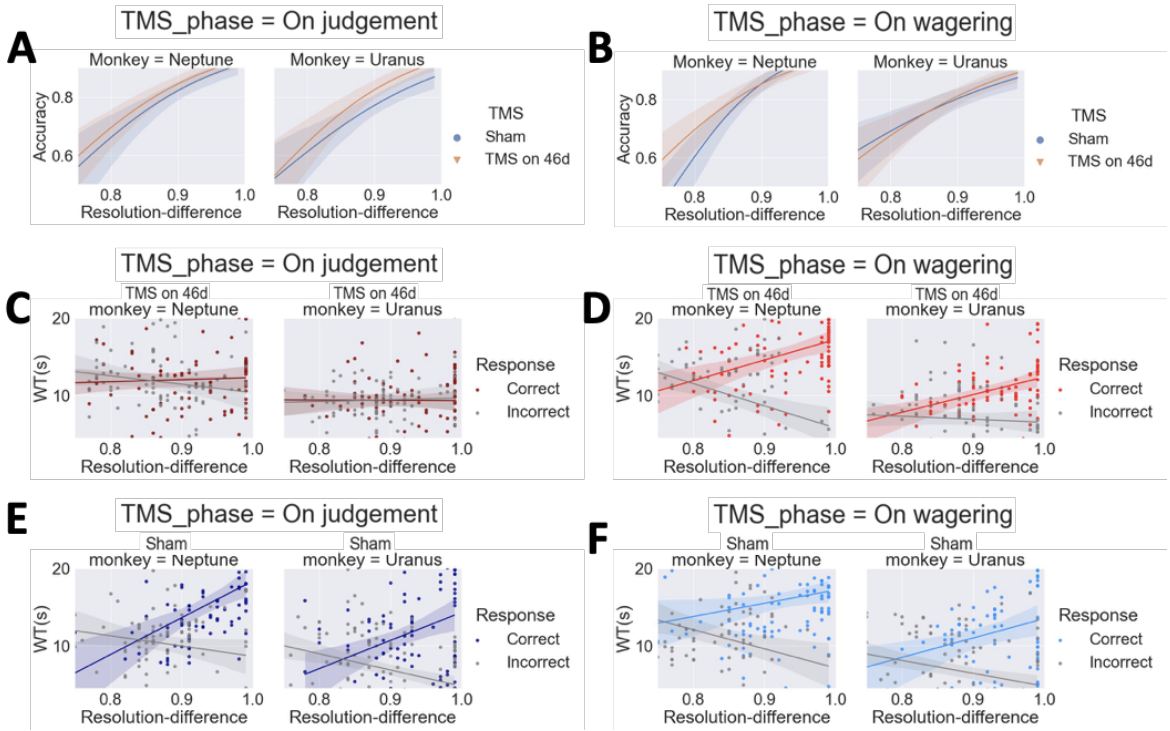


Figure 5

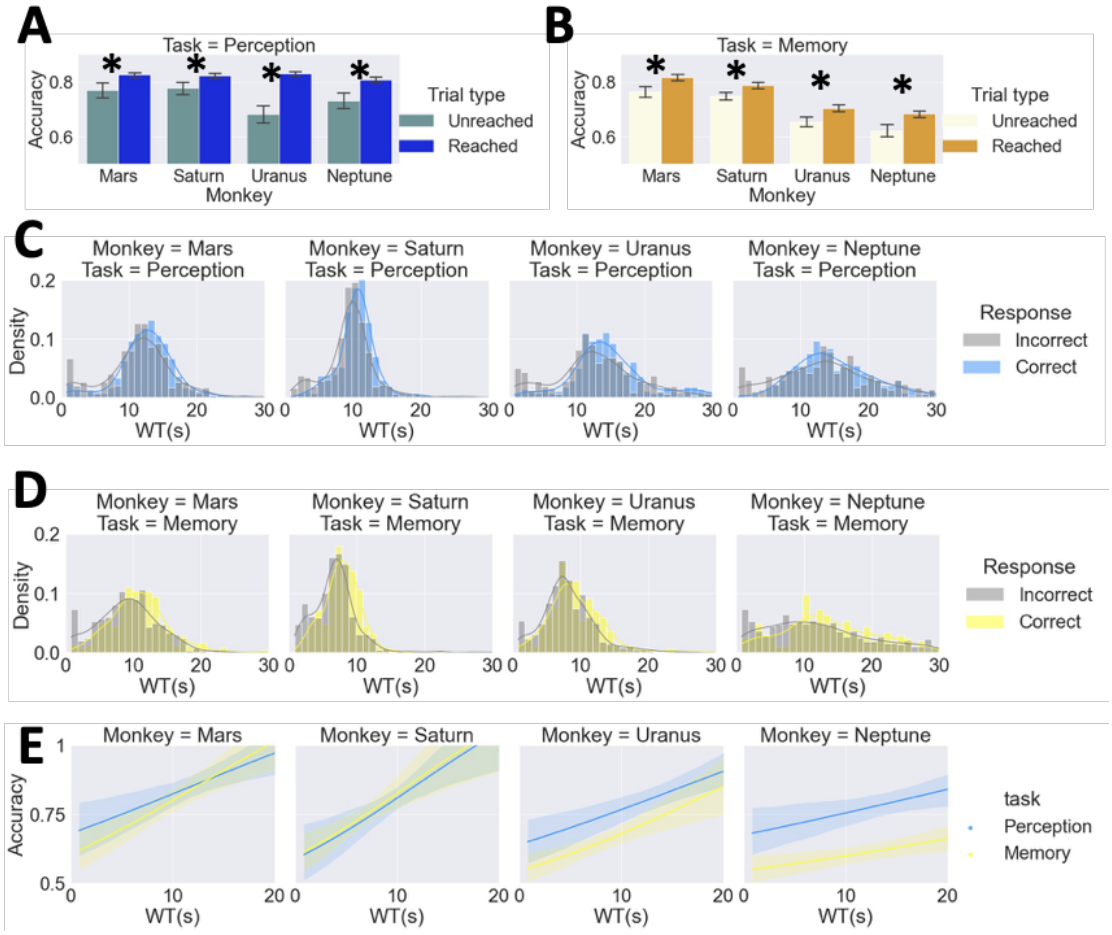
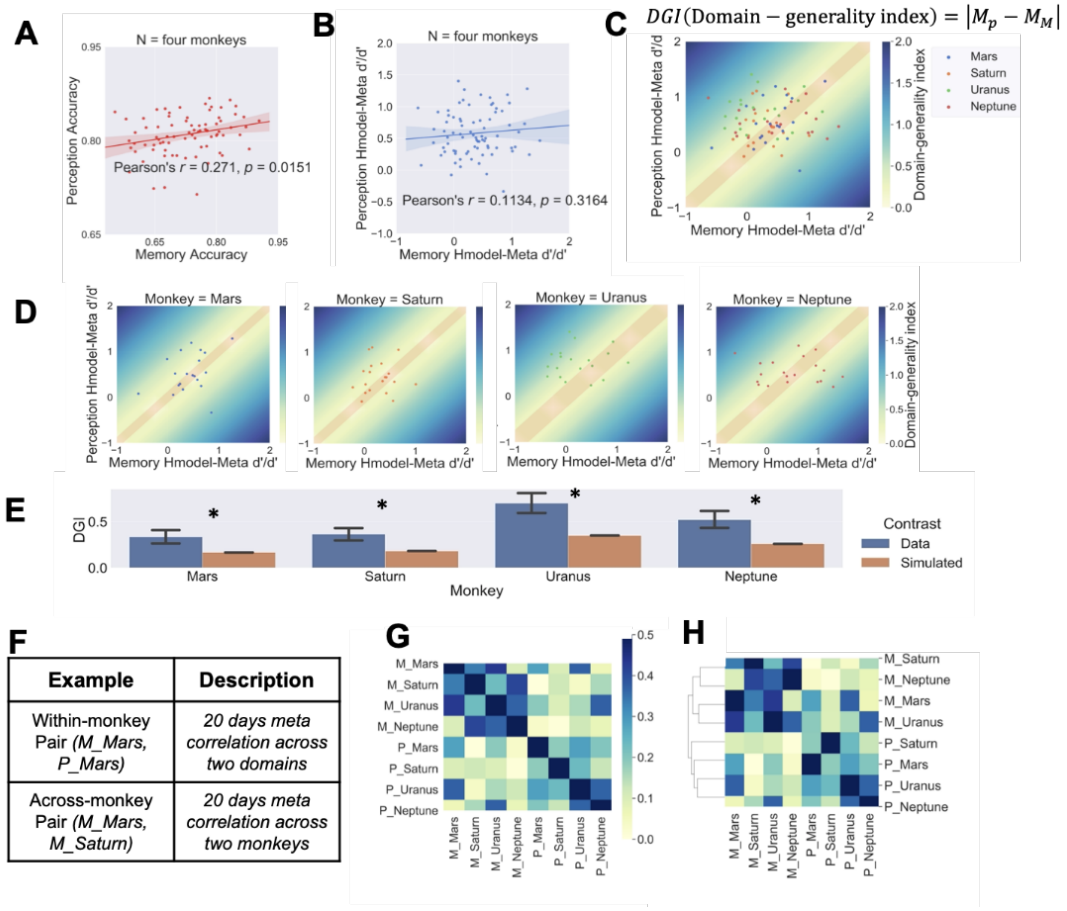


Figure 6



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