Neural substrates of metacognitive processes under sequential perceptual decision-making

Nanjo, Yoshinori

Doctor of Philosophy

Department of Physiological Sciences

School of Life Science

The Graduate University for Advanced Studies, SOKENDAI

Division of Visual Information Processing

National Institute for Physiological Sciences

September 2023

Declaration

I, Yoshinori Nanjo, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in this thesis.

June 30, 2023

Summary

The higher-level cognition works to monitor and control one's own cognition, which is known as metacognition. For instance, most people might have an experience of the feeling-of-knowing state as a consequence of monitoring our memory (metacognitive monitoring), even though we cannot recall it at that time. One strategy to recall is to control retrieval time (metacognitive control). This type of metacognition has functional roles in improving learning and decisions.

In neural substrates of metacognition, studies first focused on metacognitive monitoring. To assess metacognitive monitoring, confidence has been measured as a metric in which subjects rate their confidence on how accurate their choice would be. Based on confidence to the accuracy of choices correctly reflecting external sensory states, the neural activity and structure of the lateral prefrontal cortex (lPFC) have been shown to be related to the metacognitive sensitivity using the perceptual decision-making paradigm. However, this focus on metacognitive monitoring overlooks the inherent role of metacognition in sequential decision-making processes.

The other parts of prefrontal cortex (PFC) is also thought to play a crucial role in this higher cognitive function, especially its control functions. In neural mechanisms underlying control, the anterior cingulate cortex (ACC) in the medial PFC (mPFC) is

related to cognitive control and executive functions. It is thought that the mPFC has an important role in controlling our cognition and its implementation process. Moreover, damage to the PFC due to disease or injury results in the dissociation of knowledge (monitoring) and execution (control). Hence, the different views on the neural substrates of metacognition have not been completely resolved yet, considering both metacognitive monitoring and control.

This study then investigates 1) the role of metacognition in influencing subsequent decision-making, 2) the representation of metacognitive monitoring in the PFC, 3) the neural distinction of subsequent control from metacognitive monitoring, and 4) the neural interaction of these processes. To address these issues, a task-based functional magnetic resonance imaging (fMRI) experiment was conducted with 34 Japanese general subjects. They underwent visual size discrimination task with two-alternative forced choice paradigm, which involved sequential decision-making and rating confidence. This experimental task was designed to investigate how initial confidence influences subsequent decision-making, whether to change their choice. Each subject was required to discriminate which of the two visual stimuli presented (stimulus size was composed of four steps) was bigger and to rate how confident they were in their judgement at the time subsequently. The same stimuli were presented twice in each trial.

This task allows me to investigate whether initial decision was switched or stayed in subsequent decision-making processes. However, confidence is known to be influenced by external evidence. To examine the influences of internally-driven components of metacognition on subsequent decision-making, I classified all trials into high/low confidence trials in each stimulus difference according to median of initial confidence within each subject. I then calculated the proportion of change of mind in each confidence level.

I first, analyzed the behavioral data to elucidate whether initial confidence influences on subsequent decision-making in comparison with the proportion of change of mind between low and high-confidence trials. I found that when people were highly confident on their initial choice, they persisted initial choice. Second, I examined whether initial confidence has a functional role in error detection and correction. I also revealed that when they made an error on their initial decision given low confidence, they more frequently changed their initial choice compared to an error decision given high confidence. Both effects showed that internally-driven components of metacognition have a functional role in influencing subsequent control.

Next, regression analysis was performed on the fMRI data, comparing behaviorally identified variables against the BOLD signals recorded across the whole

brain. To investigate the metacognitive monitoring-related activity on the initial decision, I compared the activity of the high confidence with the low confidence trial and found that the anterior part of the mPFC was highly activated. Subsequently, to identify brain regions associated with change of mind, I compared the activity of the switch with the stay trial on the subsequent decision. This analysis revealed significantly higher activation mainly in the dorsal ACC. These neural results suggest that there is a dissociation representing initial confidence generation and subsequent control process. However, behavioral results showed that there was a relationship between metacognitive monitoring and subsequent control. To investigate whether a connection existed between the two processes, I finally performed a conjunction analysis between the change of mind and metacognitive monitoring related activity. Through this analysis, I uncovered a common brain region between metacognitive monitoring and change of mind in the perigenual ACC. These regions work as a coordinated system, providing new insights into human metacognition.

Impact statement

The special type of cognition, which is known as “cognition about cognition” and “thinking about thinking” is metacognition. This is the higher-order cognition, which works to monitor one’s own cognition (metacognitive monitoring) and control our subsequent behavior and mental state (metacognitive control). This is an overarching process, making us to regulate cognition and behavior. For instance, we might have an experience of tip of tongue (feeling of knowing) situations, when people recall something. In this situation, we are able to recognize our cognition about knowing it, in spite of not recalling it. This is an example of metacognitive monitoring, in that, you know what you know the event. One strategy to recall it is to extend the retrieval time. This process of using information obtained through metacognitive monitoring to regulate subsequent behavior and mental state is called metacognitive control. As one of the sequential cognitive processes, these are helpful in many aspects of our daily life, such as improving learning and business strategies.

Higher cognitive functions such as metacognition have long been thought to be performed by our prefrontal cortex (PFC). The PFC is a brain region associated with various cognitive functions such as self-other distinction, thinking, and memory, and its damage is known to interfere with behavioral control. Previous studies in neuroscience

have shown the neural substrates of metacognitive monitoring in the lateral part of the PFC (lPFC) associated with the individual difference between the behavioral metrics of metacognitive monitoring and its neural activity and structure (intensity of gray matter).

Previous studies might have overlooked two crucial aspects. Firstly, they have neglected to investigate the internal components of metacognition. The original concept of metacognition focused on internally generated components, such as self-reflection, which individuals employ to monitor their own cognition, particularly in learning situations. This internal component has been somewhat overlooked in prior research. Secondly, neuroscientific studies on metacognition have focused only on examining metacognitive monitoring within single-trial settings, rather than within the broader framework of a sequential decision-making process. However, it is important to consider metacognitive processes within the context of a sequential decision-making process, as the purpose of monitoring is to improve subsequent behavior. In my doctoral research, I aimed to address these issues and elucidate neural substrates of metacognition based on internally-driven components. To achieve this, I designed a novel experimental approach by designing a sequential perceptual decision-making task and classifying all trials into high/low confidence trials from psychophysical perspectives that minimized the influence of externally-driven components, such as stimulus differences on confidence.

I have succeeded in identifying the neural substrates of metacognitive monitoring, change of mind, and their common regions contributing to metacognitive control by using functional magnetic resonance imaging (fMRI). Further, I applied recent preprocessing of the Human Connectome Project (HCP) to my MRI data. This technique enables me to reveal the neural representations in human frontal regions because the preprocessing was performed for correcting distortion and denoising, since the frontal regions are prone to MR image distortion due to sinus cavities.

To summarize, my work has investigated behavioral and neural underpinnings, which have usually used a single decision-making context. I believe that the present findings have yielded unique insights into our understanding of metacognition.

CONTENTS

List of Figures and Tables

2	**MNI coordinates of neural correlates of change of mind**
3	**MNI coordinates of common brain regions between metacognitive monitoring and change of mind**
4	**MNI coordinates of Stay (persisting initial choice) related neural activity**

List of Abbreviations

Abbreviation	Details
ACC	**anterior cingulate cortex**
aIC	**anterior insular cortex**
amPFC	**anterior part of the medial prefrontal cortex**
ANOVA	**analysis of variance**
AP	**anterior to posterior**
BOLD signal	**blood oxygenation-level dependent signal**
BW	**band width**
dACC	**dorsal anterior cingulate cortex**
deg	**degree**
EPI	**echo planner imaging**
ES	**echo-spacing**
FA	**flip angle**
fMRI	**functional magnetic resonance imaging**
FOV	**Field of View**
FSL	**FMRIB Software Library**
FWE	**Family-wise error**

FWHM	**Full width at half maximum**
GLM	**General linear model**
HCP	**Human connectome project**
HRF	**Hemodynamic response function**
IPL	**inferior parietal lobe**
lPFC	**lateral prefrontal cortex**
mPFC	**medial prefrontal cortex**
MNI coordinate	**Montreal Neurological Institute coordinate**
PA	**Posterior to anterior**
s.d.	**standard deviation**
SEM	**standard error of mean**
SPM	**statistical parametric mapping**
PFC	**prefrontal cortex**
pTFCE	**probabilistic threshold-free cluster enhancement**
TE	**echo time**
TR	**repetition time**
T1w	**T1 weighted**
T2w	**T2 weighted**

2AFC	**Two alternative forced choice**

1. Background

1.1. Conceptual overview of metacognition

In our daily lives, we engage in certain actions, evaluate their accuracy, and optimize our subsequent behavior to achieve better outcomes. The higher-order cognitive functions that monitor current cognition and regulate subsequent actions based on self-awareness are known as metacognition (H, 1976 p.196; Flavell, 1979). Originally developed in the fields of educational psychology and cognitive psychology, metacognition is based on the idea that individuals are able to enhance their own learning efficiency by accurately recognizing their own cognitive processes. For instance, imagine a situation where a student is learning about differential calculus. If the student monitors their understanding of ordinary derivatives and determines that it is sufficient, it would be more efficient for them to transition to learning another topic, such as partial derivatives, rather than continuing to study ordinary derivatives, which they already understand well. Metacognition comprises the processes of metacognitive monitoring, which involves the overarching awareness of one's own cognition, and metacognitive control, which involves using that information to regulate subsequent actions (Nelson and Narens, 1990). As illustrated above, accurately perceiving one's own cognition can lead us to improved learning efficiency and effective decision-making.

One salient aspect pertains to the fundamental nature of original metacognition, which centers around the monitoring of one's own cognition based on internal information. Unlike perceptual cognition, which is influenced by external stimuli, our memories are inherently stored within the our brains, rendering them less susceptible to external perturbations. Conversely, when monitoring self-awareness of perceptual stimuli, it is plausible that metacognition encompasses externally-driven components, such as the strength of stimuli or their impact on decision-making. In contrast to memory tasks, perceptual decision-making tasks offer the advantage of administering a substantial number of trials within a limited timeframe, making them prevalent in experimental investigations. However, both aspects share the common objective of promoting a thorough comprehension of one's own cognition, with the overarching belief that this understanding plays a functional role in regulating subsequent behaviors.

Metacognition has functional roles on both negative and positive sides. Flexibility in cognition, a defining characteristic, involves the ability to modify (change) one's perspective in response to new information (i.e., change of mind). To effectively adjust prior confidence, it is crucial to consistently incorporate emerging evidence, even after committing to a particular course of action. This capacity holds particular significance within the infection context of the Covid-19 pandemic. Given the rapidly evolving

scientific understanding surrounding this issue, each individual must carefully consider the latest facts to make informed decisions. Consequently, previously held beliefs might need to be corrected as new information becomes available. This function of belief (confidence) is required to adapt people to new environments and information, and in guiding flexible decisions. A notable example of this can be observed in the case of vaccination, in which the effects were not believed initially. However, with the new evidence, the people's behavior was subsequently revised to adapt their beliefs and behaviors accordingly to align with the updated scientific reports. While it is crucial to integrate new evidence after making decisions for the formation of accurate beliefs, it has been consistently observed that humans exhibit biases when processing new information. Specifically, individuals tend to process new information in a way that aligns with their existing confidence on their decisions, showing a preference for information that supports their prior perspectives. This biased information processing is commonly referred to as "confirmation bias" (Tversky and Kahneman, 1974; Nickerson, 1998; Rollwage et al., 2020). In contrast, our beliefs can work to optimize our decision-making. Beliefs are recognized to serve a functionality in identifying errors in our judgments, and individuals can utilize these beliefs (confidence) to adapt their behaviors even in the absence of explicit feedback on the accuracy of their decisions. This phenomenon is known as "error

detection" (Yeung and Summerfield, 2012). Common to both of them is beliefs (confidence) about current decisions that influence subsequent decision-making.

Its neural substrates have primarily focused only on monitoring within single-trial contexts. Furthermore, studies have explored the relationship between estimated metacognitive ability and individual differences in brain structure (Fleming et al., 2010; McCurdy et al., 2013) and function (Fleming et al., 2012). However, there are two critical aspects. First, metacognition is an intra-individual phenomenon, and its comprehensive examination necessitates a deeper exploration within the individual. Second, metacognition is not a simplistic intra-individual process; rather, it encompasses multifaceted dimensions that extend beyond individual differences. As metacognition exerts functional influence not only on immediate cognitive processes but also on subsequent behavior, relying solely on single-trial settings may limit our understanding of its intricacies. Therefore, a broader investigation encompassing diverse perspectives is essential for a thorough comprehension of metacognition.

1.2. Open Questions

Metacognition has been thought that it has functional roles in influencing subsequent behavior, such as confirmation bias, error detection and correction. However, there are

some questions to understand neural and behavioral mechanisms of metacognition.

1) Internally-driven components of metacognition have functional roles in influencing subsequent decision-making: The previous works have investigated functional roles of confidence in metacognitive monitoring on subsequent decision-making. However, subjective confidence is known to be influenced by external evidence, such as stimulus strength and difference (Koriat, 2011; Koriat and Adiv, 2015). Although perceptual decision-making task was applied to examine metacognitive effects by measuring subjective confidence, they might reveal externally-driven components of metacognition. Considering original concept of metacognition in learning situation, it is important to investigate functional roles of internally-driven components of metacognition.
2) Neural substrates of metacognitive process in sequential decision settings: This is also important because previous literatures have examined neural substrates within a single trial setting (Fleming et al., 2012; Mazor et al., 2020). Recent two studies have use sequential task paradigm (Qiu et al., 2018; Boldt and Gilbert, 2022), but such a sequential paradigm may not fully investigate internally-driven components of metacognition.

2. Neural substrates of metacognitive processes under sequential decision-making

2.1. Introduction

The role of the prefrontal cortex (PFC) is associated with the transformation from knowledge to action. As one of cardinal function of the higher order cognitive function in the PFC, metacognition is thought to reflect this link between knowledge and action (Flavell, 1979; Nelson and Narens, 1990): metacognition has been thought to work to monitor our own cognitive processes and control our behavior and mental state. For instance, people have an experience of feeling-of-knowing state as a consequence of monitoring their memory (metacognitive monitoring), despite being unable to recall specific information at that moment and use this knowledge to control retrieval time of memory (metacognitive control) (Nelson and Narens, 1990; Nelson, 1996). This endogenous type of metacognition has functional roles in improving learning and decisions in daily life, such as education and business strategy (Frith, 2012; Hargis et al., 2017).

In studies on metacognitive monitoring, confidence has been measured as a metric in a situation in which subjects rate their confidence on how much their choice would be accurate (Dunlosky and Lipko, 2007; Fleming et al., 2010, 2012; McCurdy et

al., 2013; Dunlosky et al., 2016; Fleming, 2017; Mazancieux et al., 2020). Attempts have also been made to build computational models to describe the metacognitive ability (e.g., metacognitive sensitivity and meta-*d*') to precisely estimate the accuracy of their choice (Fleming and Dolan, 2012; Maniscalco and Lau, 2012; Fleming and Lau, 2014). In the neural substrates of the sensitivity of the metacognition to the accuracy of their choices correctly reflecting external sensory states, the neural activity and structure of the lateral part of the PFC (lPFC) have been shown to be related to the metacognitive ability across individuals (Fleming et al., 2010, 2012; McCurdy et al., 2013; Morales et al., 2018; Qiu et al., 2018). However, this focus on metacognitive monitoring overlooks the inherent role of metacognition in transforming knowledge to action. The initial conception of metacognition involves the influences of confidence in one's own behavioral and mental state on subsequent control (Flavell, 1979).

Earlier studies on neural substrates of metacognition have considered the medial PFC (mPFC) rather than the lPFC to be crucial for control aspects of metacognition, which is closely related to executive function (Fernandez-Duque et al., 2000; Shimamura, 2000). The anterior cingulate cortex (ACC) in the mPFC (Johansen-Berg et al., 2004; Beckmann et al., 2009; Vogt, 2009) is related to cognitive control and executive functions (Botvinick et al., 2001; Shenhav et al., 2016; Musslick and Cohen, 2021; Mansouri et al.,

2022). Researchers have also investigated neural mechanisms underlying control of our cognition, especially in temporally extended contexts, such as foraging decision-making and cognitive control (Shenhav et al., 2013). From these studies, it is thought that the medial PFC has important roles for controlling our cognition and its implementation process. Thus, the differing views on the neural substrates of metacognition have not been completely resolved yet when considering metacognitive control.

Previous studies have addressed parts of the metacognitive process, especially externally driven components in a single trial setting. With one single behavioral context, the consequence of the metacognitive process especially control of action might not be apparent. To fully investigate this control process of metacognition, an examination of sequential behavioral contexts is needed.

The metacognitive process of the subsequent control has been investigated in such sequential decision contexts. Studies have suggested the importance of metacognition to control our subsequent decision-making: 1) Metacognition may play a role in detecting errors in our decision (monitoring) and correcting these errors (control) (Butterfield and Mangels, 2003; Yeung and Summerfield, 2012; Boldt and Yeung, 2015; Harty et al., 2017; Kononowicz and van Wassenhove, 2019), and 2) Metacognition in some other contexts does not always lead us to optimal behavior. Once people commit to

a decision with high confidence, this would bias us toward persisting with the initial choice, making us reluctant to incorporate evidence and correct initial beliefs. This cognitive tendency is known as confirmation bias (Tversky and Kahneman, 1974; Beattie and Baron, 1988; Nickerson, 1998; Talluri et al., 2018; Kappes et al., 2020; Rollwage et al., 2020). Thus, metacognitive process has various influences on subsequent processes even when this is not optimal effects.

Only a few studies have explicitly examined metacognitive monitoring and control (Qiu et al., 2018; Boldt and Gilbert, 2022). Qiu et al. (2018) used a sequential perceptual decision task, in which subjects reported confidence in their judgements. However, the design of their task did not completely dissociate the effects of the stimulus strength and the internally generated assessment of their decision. On the other hand, Boldt and Gilbert (2022) utilized the cognitive offloading task, which did not require the sequential adjustment of the internal cognitive process in the control aspect of metacognition. Instead, the task externalized the control process to the objects in the environment for a reminder of the memory item. Thus, although metacognition has both a monitoring and a more direct control function on our behavior, studies have only focused on metacognitive monitoring based on external information or metacognitive control is outsourced external to the brain. However, there may be internal components

of metacognitive process that are independent of the momentary external information and persists across multiple decision contexts to influence the internal regulatory processes. Thus, it is important to examine whether internally-driven components of metacognition have a functionality in sequential decision-making for internal control process and which brain regions are related to these metacognitive processes.

The present study then investigates 1) the role of internally-driven components of metacognition in influencing subsequent decision-making, 2) the neural representation of metacognitive monitoring in the PFC, 3) the neural distinction of subsequent control from metacognitive monitoring, and 4) the neural interaction of these processes.

To address these issues, I designed and conducted a functional magnetic resonance imaging (fMRI) experiment using the sequential perceptual decision-making task with confidence ratings. This experimental design allows me to examine the influences of initial confidence (metacognitive monitoring) on subsequent decision-making, such as whether to stay or switch their choice in subsequent decision-making according to initial confidence levels, and to identify their neural substrates. To estimate the endogenous component of metacognition, all trials were classified into high/low internally-driven components of metacognition according to the median of initial confidence within each stimulus difference, as subjective confidence would be correlated

with task difficulty and stimulus strength (Lebreton et al., 2015). My findings show that initial confidence influences change of mind at subsequent decision-making and metacognitive monitoring and control are represented in separate yet partially overlapping regions of the medial PFC. Together these findings suggest that metacognition functions as an integral component of sequential decision-making.

2.2. Materials and methods

2.2.1. Subjects

This study was conducted with 34 right-handed Japanese general volunteers [24 females and 10 males, age: 23.56 ± 3.47 years (Mean ± s.d.); ranging from 20 to 34 years]. This number of subjects is consistent with recent criteria on sample size of neuroimaging study (Poldrack et al., 2017). All subjects reported normal visual acuity and no prior history of major medical or neurological illnesses. Written informed consent was obtained from all subjects in accordance with the Declaration of Helsinki. The experimental procedures were approved by the local medical ethics committee at the National Institute for Physiological Sciences in Japan.

2.2.2. Experimental setup, task design and procedures

The experimental task was implemented by the Psychtoolbox-3 added on MATLAB R_2018b (MathWorks Inc., Natick, MA, USA). I equidistantly presented the visual stimuli consisted of two white circles on a black background on a screen, which was viewed through a mirror from a liquid-crystal display projector (resolution = 1280 × 1024, distance = 190.8 cm, visual angle = 1.02 deg). The half-transparent viewing screen was located behind the MRI head coil, and visual cues were projected through the projector.

The subjects held an optical response button-box (HHSC1 × 4-D; Current Designs Inc., Philadelphia, PA, USA) on their right hand to record their decisions. During the inter-trial intervals, the subjects were instructed to look at a white crosshair placed at the center of a screen. Before performing the main experiment, subjects practiced the experimental task to familiarize the task and general procedures inside the scanner for approximately 2 minutes.

I designed and conducted the novel experimental task with two-alternative forced choice (2AFC) paradigm, which involved sequential perceptual decision-making and rating confidence (Fig. 2A). This experimental task was designed to investigate how the initial confidence influences subsequent decision-making, whether to change their choice. In each trial, there were following five phases: (1) presentation of first stimuli (1st evidence), (2) first decision and rating confidence, (3) inter-decision interval, (4) presentation of second stimuli (2nd evidence), and (5) second decision and rating confidence. Within a single trial, each subject viewed two circular stimuli on the screen (1st evidence: 1 s) and was required to indicate which circular stimulus was larger. The circular stimuli were prepared in four different sizes (0.75, 1.5, 3.0, and 6.0 % difference) and were presented on the left and right sides of the center of the screen. The same pair of stimuli with the same size difference was presented for the 1st and 2nd

stimulus presentations. Subsequently, each subject had to respond with the button press as to which circular stimulus they felt larger and with rating their confidence using a fifty-point scale (1 = minimum confidence, 50 = maximum confidence) of the visual analogue scale (VAS). The subjects made their 1st decision and rating confidence within 2 seconds. The 1st decision was followed by an inter-decision interval of 1.5 seconds, during which the first decision was presented as a position at either left or right side of the center of the screen. Next, they viewed same circular stimuli again (2nd evidence: 1 s), and then they had to make their decision and rate confidence within 2 seconds. After 2nd decision and rating confidence, a white crosshair was presented during an inter-trial interval (2.5 s).

The current experiment used an event-related fMRI design consisting of four scanning runs with 48 trials [four stimulus difference levels × two sides with a larger stimulus (left or right side) × two experimental conditions × three repetitions (repetition of the above combination)]. The experimental recording lasted approximately 50 minutes. To improve the efficiency of fMRI data, I utilized a genetic algorithm (Wager and Nichols, 2003) to arrange the orders of stimulus pairs with different stimulus sizes across trials. Within each run, jittering inter-trial intervals were set at 2.5 or 7.5 seconds the fixation crosshair was presented for the first and last 10 seconds of each run. Consequently, 4

different trial order sequences were created that were then randomly assigned to 4 scanning runs.

In addition, the same experiment was conducted simultaneously on two MRI scanners. In half of all trials, only one's own 1st decision was presented (Solo condition), and in the other half of the trials, the 1st decision of another subject performing the task at the same time was presented during an inter-decision interval (Dyad condition). In this study, I focused only on the Solo condition to investigate the neural substrates of metacognitive process occurring within the individuals.

2.2.3. Behavioral data analyses

2.2.3.1. Psychometric Fittings

Perceptual decision-making is known to be influenced by external evidence, such as evidence based on stimulus differences and across decisions (Shadlen and Newsome, 2001; Gold and Shadlen, 2007; Krueger et al., 2017; Rangelov and Mattingley, 2020). In order to examine whether subjects based their decisions on external evidence and engaged in perceptual decision-making during this task, I first compared sensitivity between the first and second decisions. Psychometric functions were constructed for each subject and each decision (first decision and second decision) by plotting the proportion of trials in

which the right stimulus was chosen as a larger stimulus (y-axis) against the stimulus difference levels (x-axis) (the positive value means that the right stimulus is larger and vice versa). I fitted the psychometric function to the proportion of the right responses using logistic function (Fig. 2B&C). To estimate the parameters for perceptual sensitivity, a logistic regression model was employed using the following equation:

$$p = \frac{1}{1 + e^{-s(x-x_0)}}$$

where p is the proportion of the right response, x is the size difference of the right versus left stimulus, x_0 is the influence point (the size difference of the stimuli such that the proportion of choosing the right side is the same as the proportion of choosing the left side), and s is the slope of the point 0.5 of y–axis, which becomes steeper as s increases. Therefore, s reflects the sensitivity of size discrimination, which we refer to as the Weber Ratio. In my analysis, I also used parameter s (slope) as a sensitivity index (Berkson, 1944, 1953; Aldrich and Nelson, 1984). I examined the slope of first and second decision to determine whether (1) each slope is positive and (2) whether the slope on the 2nd decision is steeper than the 1st one using one-tailed one sample and two-tailed paired *t*-tests.

2.2.3.2. Confidence ratings

To examine whether subjective confidence is influenced by external evidence, I investigated the effects of the stimulus size difference of each decision and the effects of evidence across decisions on confidence ratings (Fig. 2D). Subject's confidence ratings were averaged across trials within each stimulus size difference and evidence across decisions (decision on the first stimulus and decision on the second stimulus) in each subject. I conducted a two-way analysis of variance (ANOVA) with repeated measures in the design of within-subject factors of stimulus size difference (0.75, 1.5, 3.0, and 6.0 %) and evidence across decisions (1st and 2nd evidence). I applied the Greenhouse Geiser correction when the assumption of sphericity was violated as determined by the results of Mauchly's test of sphericity.

2.2.3.3. Change of mind

To examine how initial confidence influences subsequent decision-making, the proportion of change of mind was compared (Fig. 3). I focused on confirmation bias, a cognitive tendency that biases us toward persisting with the initial decision and being reluctant to revise initial beliefs in light of evidence (Tversky and Kahneman, 1974; Beattie and Baron, 1988; Nickerson, 1998; Talluri et al., 2018; Kappes et al., 2020;

Rollwage et al., 2020). To assess how initial confidence would influence subsequent decision-making, I classified all trials into high and low confidence trial according to the median of initial confidence within each stimulus difference level (median-split). I then calculated the proportion of change of mind (1st and 2nd decisions being different). This allows us to examine internally-driven metacognition, while controlling for the influence of external factors such as stimulus difference level on subjective confidence (Koriat and Adiv, 2015). To analyze the data, I then performed two-tailed paired *t* - test between high and low confidence trials. I also investigated the role of confidence in error detection because confidence has been shown to play a functional role in error detection (Butterfield and Mangels, 2003; Yeung and Summerfield, 2012; Boldt and Yeung, 2015; Harty et al., 2017; Kononowicz and van Wassenhove, 2019). To examine the effect of the initial confidence on error detection, I calculated the proportion of change of mind for trials where the initial decision was correct or incorrect. I then performed a two-way ANOVA with repeated measures, with within-subject factors of initial confidence (high and low confidence) and accuracy of initial response (correct and incorrect response). I applied the Greenhouse Geiser correction for the cases where the results of Mauchly's test of sphericity was significant.

2.2.4. MRI data acquisition

I acquired MR images of each subject using the 3T MRI scanner (MAGNETOM Verio 3T; Siemens Healthineers, Erlangen, Germany) using a standard 32-channel phased-array coil (Siemens Healthineers, Erlangen, Germany). Whole-brain functional images were obtained using multi-band gradient-echo echo-planar image (EPI) sequences (TR = 1,000 ms, TE = 35 ms, flip angle = 65 deg, FOV = 192 × 192 mm^2, matrix size = 96 × 96, slices thickness = 2.5 mm, 60 slices to the transversal brain, voxel size = 2.0 × 2.0 × 2.5 mm^3, multi-band acceleration factor = 6, phase-encoding direction = anterior to posterior). Each run consisted of 520 volumes, and four runs were administered. Before the main functional scan session, spin-echo EPI images were acquired with the opposite phase-encoding direction for subsequent susceptibility distortion correction in preprocessing for functional images (TR = 7,560 ms, TE = 64 ms, in plane FOV = 192 × 192, matrix = 96 × 96, 60 transversal slices, FA = 90 deg, refocus FA = 180 deg, BW = 1,860 Hz/pixel, and ES = 0.76 ms). In another session, T1- and T2-weighted structural images with 0.8-mm isotropic resolution were obtained in the same way as a previous study (Yamamoto et al., 2021).

2.2.5. MRI data analyses

In the following fMRI analyses, I excluded a total of four subjects [1 female and 3 males] because I did not have their T2-weighted and resting-state images available to perform the same preprocessing as the other subjects (Glasser et al., 2013, 2016; Smith et al., 2013).

2.2.5.1. Preprocessing

I used Human Connectome Project (HCP) Pipelines (v4.0.1; Glasser et al., 2013) for MRI data preprocessing for correction for gradient nonlinearity-induced distortion, head motion correction, susceptibility-induced distortion correction, and normalization same as previous studies (Glasser et al., 2013, 2016; Yamamoto et al., 2021). I performed FMRIB's ICA (Independent component analysis)-based Xnoiseifier (FIX) for multiple-run-concatenated fMRI time series data (MultiRunFIX) that effectively extracted and removed structured noise components observed through a session in each subject, which was implemented as described by previous studies (Smith et al., 2013; Glasser et al., 2016, 2018). Prior to concatenation, time series data of each run was demeaned and variance-normalized and high-pass temporal filtered (FWHM period = 2,000 s). Then, the normalized time-series were concatenated across runs. Melodic ICA implemented on FSL

(v6.0.1, Centre for Functional MRI of the Brain, Oxford University, UK, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslInstallation) was applied to the concatenated time-series data to produce component spatial maps and time-series. Using a classifier trained with the resting state fMRI data of HCP, these components were automatically classified into signal or noise (Griffanti et al., 2014). The noise components were non-aggressively regressed out of the EPI time-series (Glasser et al., 2016). The time-series data after denoising were split back into the individual runs, and the spatial mean and variance profiles were restored to the individual runs. I then performed spatial smoothing to improve signal to noise ratio using an isotropic Gaussian kernel of 6 mm full width at half maximum (FWHM = 6 mm) using FSL v6.0.1.

2.2.5.2. Analysis of neural activity for the metacognitive processes

I used statistical parametric mapping (SPM) 12 software (Welcome Trust Centre for Neuroimaging University College London, UK; https://www.fil.ion.ucl.ac.uk/spm/software/download/) in MATLAB R_2018b. At single subject level, I fitted a general linear model (GLM) to the fMRI data of each subject. I modeled the neural activity with functional data convolved with a canonical hemodynamic response function (HRF). The design matrix included six regressors of

interest for each decision-making process: Two regressors specified high and low confidence trials on the 1st decision and four regressors specified the 2nd decision trials based on the combinations of initial confidence (high and low confidence) and change of mind (change and no change from initial decision). Each regressor was defined by the onset of the visual stimuli and a duration of 3 s, covering the period of presentation of visual stimuli and decision and confidence ratings. The data were high-pass filtered with a cut-off period of 104 s to remove low-frequency signal drifts and the autocorrelation model [AR(1)] was globally applied over the brain. This GLM also included regressors for conditions with presentation of other's initial decision in inter-decision interval, which were convolved with canonical HRF as regressors of no interest. Temporal derivatives of each regressors of interest were also included in the design matrix as a nuisance regressor.

I calculated the values of estimated parameters using the least-squares estimation on the high-pass-filtered data. The parameter estimates in the individual analysis consisted of contrast images, which were to be used for the second-level analysis of random effects to estimate brain activation at the group-level. I entered the contrast images comparing the high confidence with low confidence on the 1st decision trials of each subject into the one-sample *t*-test. I also entered the contrast estimates on the 2nd decision trials of each subject into the two initial confidence levels (low/high confidence)

x two conditions of change of mind (no change (stay)/change (switch) of mind) full factorial design. In group-level analysis, I performed Probabilistic Threshold-free Cluster Enhancement to group level data (Smith and Nichols, 2009; Spisák et al., 2019).

I then performed a conjunction analysis to investigate common regions bet*w*een metacognitive monitoring (high > low confidence on the 1st decision) and change of mind (switch > stay on the subsequent decision-making process) applied a binary mask. This binary mask was created for a conjunction analysis by metacognitive monitoring related activity obtained from high > low confidence [$p < 0.05$ FWE corrected at the cluster-level with height threshold of $p < 0.001$] to contrast image of change of mind.

2.2.5.3. Functional connectivity analysis

I conducted a partial correlation analysis to examine the relationship between the brain regions associated with metacognitive processes in the PFC. I identified the peak coordinates of the activated regions corresponding to metacognitive processes: the anterior part of the medial prefrontal cortex (amPFC) for confidence in the initial decision, the dorsal ACC (dACC) for changes of mind, and the perigenual ACC as a common region. The BOLD time series from separate regions of interest (ROIs), defined by spheres (4mm) centered on the MNI coordinates of each region, were extracted after

removing the effects of no interest using F-contrasts [Left coordinates: amPFC (-5 63 15), dACC (-1 35 21), and perigenual ACC (-3 45 19); Right coordinates: amPFC (5 63 15), dACC (1 35 21), and perigenual ACC (3 45 19)].

Next, I estimated the neural time series as follows (Smith et al., 2013). First, the MR signal from each ROI was extracted as an eigenvariate time series. Then, the extracted signal was deconvolved using the canonical HRF, resulting in an approximation of neural activity (Gitelman et al., 2003). Subsequently, I conducted partial correlation analyses between each ROI in the solo condition within each subject. The statistical analysis at the group level was performed using a one-sample *t-test*.

2.2.6. Statistical threshold and labelling brain regions

The statistical significance level for all behavioral data analyses was set at $p < 0.05$. The statistical threshold was set at $p < 0.05$ with family-wise error (FWE) correction at the cluster level for the whole brain, with the height threshold of $p < 0.001$ for controlling false positives in all fMRI data analyses (Eklund et al., 2016; Flandin and Friston, 2019).

To label brain regions, I used SPM - Anatomy Toolbox v3 (Eickhoff et al., 2005, 2006, 2007), anatomical automated labeling 3 (Tzourio-Mazoyer et al., 2002; Rolls et al., 2015, 2020), and a book atlas (Mai et al., 2015). All analyses were conducted on

MATLAB R_2018b.

2.3. Results

2.3.1 General questions and methods

In the present study, I examined behavioral and neural mechanisms of internally-driven components of metacognitive processes: monitoring and its effect on control process in a sequential decision-making process. To investigate these mechanisms, I designed a novel experimental task in which subjects made two sequential perceptual decisions with rating confidence for each decision while the brain activity of the subjects performing this task was recorded. I aimed to examine influences of initial confidence (metacognitive monitoring) on subsequent decision-making, whether to switch to another choice or stay (persist) with the initial decision (control) and to identify which brain areas support these endogenous metacognitive processes underlying sequential decision-making.

2.3.2. Basic behavioral results

I conducted basic behavioral analyses to examine whether subjects performed the behavior in response to external evidence in a manner required in this experimental task of perceptual decision-making. In this experimental task, evidence for perceptual decisions originated from stimulus size difference between two circular stimuli, which were presented twice in a single trial. I fitted psychometric functions to the behavioral

data to examine (1) whether the trends of these psychometric slopes were positive and (2) whether the repeated presentation of the stimuli enhanced accuracy of perceptual decisions. The slope on each decision was estimated based on maximum likelihood estimation at the choice proportion of 0.5 on the y-axis (Fig. 2B & C).

I first performed one-tailed one sample *t*-test to examine whether these psychometric slopes were positive, and the results confirmed that they were positive [1st decision: 0.436 ± 0.029 (Mean ± SEM); t (33) = 15.28, $p < 0.001$; 2nd decision: 0.503 ± 0.033 (Mean ± SEM); t (33) = 15.15, $p < 0.001$]. These psychometric functions showed that the proportion of right response increased as the right stimulus became relatively larger than the left stimulus. Subsequently, I compared these slopes using two-tailed paired *t*–test to examine whether the repeated presentation of the same stimuli increased the accuracy of responses. The results showed that the slope on the 2nd decision was significantly steeper than the 1st decision [t (33) = 2.60, $p = 0.013$]. These results suggest that evidence from external stimuli enhanced the accuracy of perceptual decisions (Shadlen and Newsome, 2001; Gold and Shadlen, 2007; Murphy et al., 2015; Rangelov and Mattingley, 2020).

I also tested whether confidence is influenced by stimulus evidence in this experimental task. Previous studies have reported that subjective confidence is influenced

by stimulus evidence (Koriat, 2011, 2012; Murphy et al., 2015; Rangelov and Mattingley, 2020). In a framework of an evidence accumulation model, repeated presentation of the same stimulus makes the responses of the subjects more accurate (Shadlen and Newsome, 2001; Gold and Shadlen, 2007). In a similar way, the repeated stimulus presentations would also enhance the levels of confidence on the perceptual decisions (Kiani and Shadlen, 2009; Koriat, 2011; Koriat and Adiv, 2015; Lebreton et al., 2015). I tested these ideas of the influences of stimulus evidence on confidence within a single decision timing and across decision timings (Fig. 2D). I implemented two-way ANOVA with repeated measures and within-subject factors of stimulus difference levels (0.75, 1.5, 3.0 and 6.0 %) and decision timings in the orders of presentation of stimulus (the 1st and 2nd decisions) to examine whether subjective confidence is influenced by the accumulation of evidence (Fig. 2D). To test the influences of evidence on subjective confidence within single decision timing, I confirmed the main effect of stimulus difference [F (3,99) = 22.18, $p < 0.001$]. The main effect of the accumulated evidence across decision timings was also significant [order of presentation of stimulus: F (1,33) = 19.61, $p < 0.001$]. However, there was no significant interaction between the two factors [F (3,99) = 2.36, $p = 0.076$].

Overall, these results suggest that the subjects in our study performed the task in

response to the external evidence, and their perceptual decisions and subjective confidence were influenced by the stimulus evidence.

2.3.3. Change of mind

In this perceptual decision-making task, it was evident that the external evidence influenced confidence reports and its underlying metacognition. However, there might also be internal components of metacognitive decision process that are independent of the momentary external information and persists across multiple decision contexts (Rahnev et al., 2015). Furthermore, these internally-driven and persisting metacognitive process can drive the decision processes across multiple timings. Here, I investigated the influences of internally-driven components of initial confidence on subsequent decision-making. To do this, I classified all trials within each stimulus difference defined by the stimulus size difference into high and low confidence on the 1st decision and examined the influence of these confidence on the second decisions. First, I investigated whether internally-driven metacognition influences subsequent decision-making by comparing the proportion of change of mind between high and low confidence trials. I observed that the proportion of change of mind on the high confidence trials was significantly lower than that of low confidence trial [high confidence: 0.120 ± 0.019 (Mean ± SEM); low

confidence: 0.205 ± 0.019 (Mean ± SEM); t (33) = 4.59, $p < 0.001$; Fig. 3A]. I then analyzed the proportion of change of mind for each stimulus difference level to test whether the initial confidence influences on subsequent decision-making independent of stimulus strength (Fig. 3B). I performed a two-way ANOVA with repeated measures and within-subject factors of initial confidence (high and low confidence) and stimulus difference level (0.75, 1.5, 3.0 and 6.0 %). I found significant main effects of initial confidence and stimulus difference level [initial confidence: F (1,33) = 14.67, $p < 0.001$; stimulus difference: F (3,99) = 5.81, p = 0.001]. However, there was no significant interaction between the two factors [F (3,99) =2.13, p = 0.101].

Next, I analyzed the proportion of change of mind in erroneous and correct choices with regard to the 1st decisions. In this analysis, I followed the logic of error detection with metacognition, which posits that metacognition should impact the processes of error detection and correction (Yeung and Summerfield, 2012). I conducted a two-way ANOVA with repeated measures, using initial confidence (high and low confidence) and accuracy of initial response (correct and incorrect) as within-subject factors, to examine whether initial confidence played a role of error detection and optimizing subsequent decision-making (Fig. 3C). The main effects of initial confidence and accuracy of initial choice were also significant [initial confidence: F (1,33) = 19.54,

$p < 0.001$; accuracy of initial choice: $F(1,33) = 66.02$, $p < 0.001$]. There was a significant interaction between initial confidence and accuracy of initial choice [$F(1,33) = 10.72$, $p = 0.002$]. To investigate influences of initial confidence on change of mind, the post-hoc test revealed that the proportion of change of mind after the low initial confidence reports in incorrect initial choice was significantly higher than that after high initial confidence reports in incorrect choice [low confidence: 0.369 ± 0.036 (Mean ± SEM); high confidence: 0.231 ± 0.032 (Mean ± SEM), $p < 0.001$, Bonferroni corrected]. The proportion of change of mind after high initial confidence reports in correct choice was significantly lower than that after low initial confidence reports in correct choice [high confidence: 0.082 ± 0.015 (Mean ± SEM); low confidence: 0.124 ± 0.015 (Mean ± SEM), $p = 0.008$, Bonferroni corrected]. These results suggest that the part of confidence derived from internally driven metacognitive component influences the subsequent decision-making process, over and above the effects of the objective accuracy of the initial choice.

2.3.4. Neural activity underlying initial confidence

The behavioral data analyses revealed that initial confidence plays a role in modulating the proportion of changing and sustaining choices across decision timings (Fig. 3). To investigate brain regions responsible for metacognitive monitoring, as identified by

behavioral data (Fig. 3), I next analyzed the fMRI data.

I conducted a regression analysis, comparing behaviorally identified variables against the blood oxygenation level-dependent (BOLD) signals recorded across the whole brain. To investigate the neural activity related to the metacognitive monitoring on the 1st decision, I performed a one-sample *t* – test within individuals (high > low confidence on the 1st decision) and found that the anterior part of the medial prefrontal cortex (amPFC) was significantly activated in high confidence trials [$p < 0.05$ FWE corrected at the cluster-level with height threshold of $p < 0.001$, Fig. 4].

Although I have obtained the neural substrates for metacognitive monitoring in the 1st decision, it is necessary to investigate the subsequent neural process where a change of mind occurs to examine the effects of the initial confidence on the subsequent control processes.

2.3.5. Neural substrates of change of mind

In this study, I demonstrated that the amPFC represents internally generated confidence levels. To further elucidate the neural substrates involved in controlling behavioral changes during the 2nd decision, I analyzed fMRI data related to change-of-mind occurrences. The change-of-mind metric corresponds to the variable displayed on the y-

axis of the behavioral data in Figure 3.

To identify the brain regions with increased neural activity associated with change of mind, I performed one sample t – test between the switch and stay in the full factorial design. This analysis revealed significantly higher activation mainly in the right dorsal ACC, (dACC) right inferior parietal lobe (IPL), and bilateral anterior insular cortex (aIC) [$p < 0.05$ FWE corrected at the cluster-level with height threshold of $p < 0.001$, Fig. 5]. These brain regions were located relatively posterior to the brain regions identified as the regions for metacognitive monitoring in the previous analysis.

To examine the brain regions actively engaged for the persisting the consistent choice sequence, I conducted the analysis of stay versus switch contrast. I did not find any PFC regions significantly activated in this contrast. The brain regions significantly activated in this contrast were primary visual cortex [$p < 0.05$ FWE corrected at the cluster-level with height threshold of $p < 0.001$, Fig. 8 & 9, Table 4].

These neural results suggest that there was a spatial difference in the neural activity representing initial confidence generation (monitoring) and subsequent decision-making process (control). However, behavioral results showed that there was a relationship between metacognitive monitoring and subsequent control (Fig. 3). To investigate whether a connection existed between the two processes, I performed a

following conjunction analysis.

2.3.6. Common brain regions between initial confidence and subsequent decision-making

Thus far, I have identified the neural activity corresponding to internally-driven confidence generation and subsequent control. Next, I conducted a conjunction analysis to examine common activated regions between initial confidence (metacognitive monitoring) and change of mind (control). I used a binary mask of metacognitive monitoring to conduct a conjunction analysis between the change of mind (switch > stay) and initial confidence (high > low confidence) related activity. Through this analysis, I uncovered a common brain region between initial confidence and change of mind in the right perigenual ACC ($p < 0.05$ FWE corrected at the cluster-level with height threshold of $p < 0.001$; Fig. 6, Table 3). However, because this conjunction analysis is influenced by the size of the masked regions, the relative timing and duration of the event regressors for the two processes would influence the results of the conjunction analysis. To examine the robustness of the initial conjunction analysis, I created a binary mask of brain regions related to change of mind and conducted a conjunction analysis to contrast image of initial confidence with this binary mask to confirm the above result. The same commonly

activated regions were observed (Fig. 6C).

This common region between the two neural activations might also depend on length and onset of each event regressor because these have overlapping task time series, which were created by convolution with the canonical HRF. I performed confirmation analyses of the common neural activation between pairs of regressors of initial confidence and change of mind using GLMs of events with different lengths and different onset timings (Fig. 7). These analyses consistently showed that there was still a common activation in the perigenual ACC for each pair of regressors with different event onsets and varying lengths.

2.3.7. Functional connectivity between confidence and control regions

So far, I identified the common brain region associated with metacognitive monitoring and control in the perigenual ACC. It shows spatially overlapping between metacognitive processes, although no demonstrated temporal communication in the PFC. Finally, I performed partial correlation analysis between the amPFC, perigenual ACC, and dACC defined by peak coordinates (Fig. 10). It shows that there are significant positive correlation between the amPFC and perigenual ACC (Left: $r(29) = 0.339 \pm 0.025$, $p < 0.001$; Right: $r(29) = 0.393 \pm 0.028$, $p < 0.001$; Mean ± SEM), and the perigenual ACC

and dACC (Left: r(29) = 0.506 ± 0.025, $p < 0.001$; Right: r(29) = 0.578 ± 0.028, $p < 0.001$; Mean ± SEM). However, there was no significant correlation between the amPFC and dACC (Left: r(29) = 0.048 ± 0.030, $p = 0.118$; Right: r(29) = 0.060 ± 0.031, $p = 0.065$; Mean ± SEM). This result shows that the amPFC and dACC were not directly connected although they communicated each other through perigenual ACC. Thus, it implicated that perigenual ACC related to metacognitive processes spatially as well as temporally has a role for the hub connecting the neural activity of metacognitive monitoring with control.

2.4. Discussion

The present study advances our understanding of metacognitive monitoring and control, especially on the process of how internally-driven and persisting confidence guides subsequent decision processes. By designing a sequential perceptual decision-making experiment with a size discrimination task, this study found that when initial confidence was low, the proportion of changing decisions was higher and vice versa. The amPFC was associated with initial confidence, while the dACC showed increased activity when subjects changed their minds during subsequent decision-making. The perigenual ACC was involved in both processes, highlighting the crucial connection from metacognitive monitoring to control in the medial PFC during sequential decision-making.

2.4.1. Limits in generalization of the current findings about metacognition

It is important to discuss my findings in the context of the original concept of metacognition (Flavell, 1979; Nelson and Narens, 1990), which focused on monitoring one's own cognition about memory in order to control and optimize subsequent learning strategies accordingly. This mnemonic aspect of metacognition known as metamemory is related to the feeling of knowing and tip-of-tongue experiences (Hart, 1965; Brown and McNeill, 1966; Nelson, 1984; Brown, 1991; Burke et al., 1991; Koriat, 1993). In this

context, metacognitive monitoring involves assessing memory access, while metacognitive control governs adjusting memory retrieval time. Similar functionality of metacognition has been observed in animal behavior for avoiding penalties and obtaining rewards in mnemonic and perceptual decision tasks (Miyamoto et al., 2018; Stolyarova et al., 2019).

It is possible that the differences between the original concept of metacognition and the present study lies in the source of cognitive information used for metacognitive monitoring. Monitoring one's own memory relies on internal sources of information because memory is retrieved from internal storage. In contrast, my study focused on confidence about decisions and its underlying metacognition in perceptual decision-making. This source of information for monitoring is partly external because perceptual decision-making relies on sensory input from the external world. However, perceptual decision-making in the present study also involves substantial internally-driven components of metacognition on decision process, which is estimated by experimentally controlling influences of external factors and focusing on the spontaneously fluctuating levels of confidence (Fig. 3). Though there is difference of modalities between my study and original studies, my findings on internally-driven metacognition on decision process echo the original concept and its functionality for optimizing our behavior.

It is inherently challenging to implement psychological concepts in experimental tasks, as any instantiation of metacognitive monitoring and control might be deemed insufficient or only cover the parts of the whole concept. I aim to compare previous studies used sequential decision paradigms (Qiu et al., 2018; Boldt and Gilbert, 2022) with the present study to highlight the novelty of my findings.

Qiu et al. (2018) used sequential perceptual decision task but employed the staircase procedure to control the difficulty of the task. While this might seem to control the contribution of the external information on confidence, the performance levels might not remain consistent within an experimental session for a given subject, and the amount of the external information would fluctuate within the experimental session. Consequently, the reported confidence levels may be influenced by the variation of the external sensory evidence. The current study potentially offers better control of the effects of external sensory evidence on confidence by using a median split procedure within the physically identical condition of external evidence to estimate the endogenous component of confidence.

On the other hand, Boldt and Gilbert (2022) utilized similar internal fluctuation of confidence. However, the authors used the cognitive offloading task, which did not require the sequential adjustment of the internal cognitive process in the control aspect of

metacognition. Instead, the task externalized the control process to the objects in the environment for a reminder of the memory item. In contrast, my experimental task involved an internal control process at the second decision without the aides of external stimulus, based on the internal monitoring process at the initial decision. This distinction underscores the originality of this study within the context of the neural and behavioral mechanisms of endogenous metacognition underlying sequential decision-making.

2.4.2. Functional roles of internally driven component of metacognition

The present study reveals that internally-driven metacognition plays functional roles in controlling subsequent behavior (Fig. 3). However, this finding could be influenced not only by confidence, but also by the accuracy of the initial decision: people tend to be more confident when making correct decisions and less confident when making erroneous ones (Butterfield and Mangels, 2003; Yeung and Summerfield, 2012; Boldt and Yeung, 2015; Harty et al., 2017; Kononowicz and van Wassenhove, 2019). I conducted analyses on the effects of both confidence and accuracy of initial decisions (Fig. 3C) and found that in trials with errors in initial decisions, people were more likely to stick to their decision when they were highly confident in their first decision though initial decisions with errors should be corrected. On the other hand, when the subjects were less confident

in their erroneous first decision, this low confidence helped people to correct their error decision (error detection). In the trials with correct initial decisions, people sustained their correct decision when they were highly confident in the initial decisions. Thus, the internally driven component of confidence exhibits both positive aspects (facilitating error detection and correction, sustaining correct decisions), and negative aspects (hindering error detection and correction). The positive side of metacognition is relevant to error detection while the negative side reflects confirmation bias. Importantly, the duality of these metacognitive processes may not be present if it is simply driven by external information. Hence, the behavioral findings in the present study suggest that the initial confidence, generated by internally driven metacognition, plays a functionally significant role in shaping the subsequent cognitive processes.

2.4.3. Functionality of metacognition

It has been debated whether metacognition has functionality. Previous studies on perceptual metacognition have primarily focused on externally-driven metacognition and metacognitive monitoring (Fleming et al., 2010, 2012; McCurdy et al., 2013; Morales et al., 2018; Qiu et al., 2018). The original concept of metacognition emphasized the optimization of learning in educational settings (H, 1976; Flavell, 1979; Nelson and

Narens, 1990; Nelson, 1996). My findings, demonstrating the impact of internally-driven metacognition on subsequent processes, underscore the need to consider metacognitive process as a part of functional systems, encompassing not only monitoring, but also controlling sequential decision-making. Metacognition in nature would have input from the external world, which is integrated by metacognition, and then an action is implemented. Considering foraging decision-making, where animals respond to external stimuli while monitoring the external state of the world, their own internal state, and the consequences of their actions on the environment (Hayden et al., 2011; Stolyarova et al., 2019), metacognition likely play a role in governing sequential interactions with the external world. Thus, the present study supports the view that metacognition as a part of the system coordinates the adaptive interaction bet*w*een the agent and the environment.

2.4.4. Neural substrates of metacognitive processes

In the previous studies, the neural substrates of the metacognitive monitoring and control have been thought to separately reside in the lateral and the medial surfaces of the PFC (Fleming et al., 2010, 2012; McCurdy et al., 2013) (Fernandez-Duque et al., 2000), without a clear understanding of the neural link between the two processes of metacognition. By focusing on the internally-driven component of metacognitive

monitoring, my study found that metacognitive monitoring was represented in the amPFC with activity correlated with initial confidence and the dACC represented the control by showing the correlated activity with change of mind. Anatomically, the brain regions with the observed neural activity are dissociated by fiber connectivity patterns of the respective brain regions (Passingham et al., 2002; Neubert et al., 2015). The amPFC represents metacognitive monitoring, which was defined by difference of confidence levels, the dACC is associated with changes of mind, which is reflected in subsequent behavior, and the perigenual ACC exhibits common activity for both processes. These findings suggest that the metacognition investigated in this study aligns with the original concepts of metacognition as an overarching and sequential cognitive processes (Flavell, 1979; Nelson and Narens, 1990).

2.4.5. Medial and lateral brain regions of metacognition

The PFC itself plays a central role in the higher-level organization, and damages to the PFC produce several deficits and disorders in goal directed cognition (Damasio et al., 1993; Miller and Cohen, 2001; Luria, 2012). Lateral PFC regions are associated with cognitive control (Koechlin et al., 2003; Badre and D'Esposito, 2007, 2009; Koechlin and Summerfield, 2007), including metacognitive ability (Fleming et al., 2010, 2012;

McCurdy et al., 2013), while medial PFC regions represent self-monitoring or estimation (Frith, 2012; Yoon et al., 2021).

It is important to consider the reasons for the divergence between the current and the previous results. Previous studies have mainly used the correlation methods between neural structure/activity and behavioral indices, such as metacognitive sensitivity, which is compared with neural parameters across individuals (Fleming et al., 2010, 2012; McCurdy et al., 2013; Goupil and Kouider, 2016; Maniscalco et al., 2016; Mazor et al., 2020). In contrast, this study analyzed neural activity based on trial-by-trial fluctuations of confidence within individuals. This methodological divergence might explain the observed location difference. Another possibility is that the previous studies focused on externally-driven metacognition and its neural substrates while the current study used the internal component of metacognitive monitoring by carefully controlling the effects of the external sensory evidence. In this study, metacognitive monitoring was estimated by confidence, which it provided simple overarching processes. Thus, amPFC might be related to metacognitive monitoring of abstract processes.

2.4.6. The dorsal-ventral gradient in medial PFC

The neural activity in the medial prefrontal area extends not only ventrally, but also

dorsally. The medial prefrontal area is known to have the neural representations related to the self on the ventral side and others on the dorsal side (Denny et al., 2012; Wittmann et al., 2016). Considering that metacognition is an overarching process of cognition about cognition, this functional gradient might suggest that each subject is cognizant of one's own cognition with a perspective of like others. Taken together, this finding indicate that the amPFC could be associated with metacognitive monitoring of original idea that it is cognition about cognition.

2.4.7. Neural substrates of subsequent control process

I hypothesized that the dACC is involved in a subsequent control because it is associated with cognitive control and executive functions in sequential situations. I demonstrated that the dACC was involved in change of mind, which was identified with switch v.s. stay contrast at the timing of the 2nd decision. The activated regions included the right dACC, bilateral aIC, and right IPL. These regions for change of mind are consistent with previous studies on switching behavior (Mobbs et al., 2013; Philipp et al., 2013; Fleming et al., 2018). In the situation of foraging decision under competition, such switching behavior is supported by the dACC and the aIC (Mobbs et al., 2013), and the regions including the IPL is al, so associated with switching category and response (Philipp et al., 2013). Thus,

my findings on change of mind are consistent with previous works.

2.4.8. The common brain region for metacognitive monitoring and control

It has not been clear what neural mechanism translates the metacognitive monitoring to the control process. A conjunction analysis between the neural activities of metacognitive monitoring and change of mind revealed that the perigenual ACC was the common brain region for both processes. This finding demonstrates that the perigenual ACC can serve as a hub between metacognitive monitoring and subsequent decision-making process or is involved in metacognitive control. Notably, this region is located between amPFC (metacognitive monitoring) and dACC (change of mind). Metacognitive control appears to be represented in this region, which is related to higher-order decision-making process (Wang et al., 2020), along with the dACC, which is considered as related to cognitive control and executive function (Shenhav et al., 2016; Musslick and Cohen, 2021; Mansouri et al., 2022). The analysis comparing the activity patterns of the metacognitive monitoring and control revealed the common brain region in the perigenual ACC, which can provide the insight into the potential link between the two processes. Thus, the perigenual ACC together with dACC and amPFC work as a coordinated system of metacognitive monitoring and control in the human brain.

2.4.9. Functional connectivity among amPFC, dACC, and perigenual ACC

It is possible that the behaviorally observed link between metacognitive monitoring and control can be realized through several different neural mechanisms. It can be done by the overlapping neural activity. Alternatively, it can be based on the interregional functional conectivity. However, I did not find the direct interaction between the amPFC and dACC. Still, I observed the interaction between the perigenual ACC and the two brain regions (Fig. 10). This may suggest that the perigenual ACC might play a role of mediating the interaction between the amPFC and the dACC. These patterns of the interaction could be influenced by the factor of spatial proximity. I did some control analysis with the neural time series from the neighboring regions of the amPFC but it could not find any significant functional connectivity. Hence, it is possible that the perigenual ACC genuinely play a role of a hub region in this medial prefrontal network. Whether as a hub region or as the nexus of the overlapping activity for the metacognitive monitoring and control, the perigenual ACC play a crucial role in bridging the metacognitive monitoring and control.

Conclusion

The present study elucidated that internally-driven components of metacognition have functionality and its neural substrates. Metacognitive monitoring is encoded in the amPFC. In subsequent decision-making process, the dACC is related to change of mind and the perigenual ACC has a role of hub connecting metacognitive monitoring with subsequent decision-making process and metacognitive control. I believe that this study will guide future research on metacognition by emphasizing its functionality within a framework of sequential process.

Acknowledgements

I would like to express my deepest appreciation to my supervisor, Prof. Yumiko Yoshimura, whose exceptional mentorship, and unwavering support have been instrumental in shaping the outcome of my graduate study. I am profoundly grateful for the invaluable opportunity she provided me to continue my research under her guidance. I extend my sincere gratitude to my previous supervisor, Dr. Norihiro Sadato. He provided diligent guidance on my research design and the design of experimental tasks and granted me the invaluable opportunity to conduct fMRI experiments. Special thanks are due to Dr. Tetsuya Yamamoto, whose extensive guidance and kind assistance proved indispensable in the successful execution of my MRI data analysis. Without his guidance, the findings of this thesis would not have been possible. I am deeply indebted to Dr. Kohei Miyata of Graduate School of Arts and Sciences, The University of Tokyo, who guided me in experimental design and programming skills in the early stages of my research. I am also honored to acknowledge the insightful comments and expertise of my thesis committee members, Prof. Keiichi Kitajo, Prof. Masaki Isoda, and Prof. Tetsuya Matsuda. Furthermore, I extend my heartfelt appreciation to the technical support staff, Mr. Yoshikuni Ito, and past members of the Sadato-lab, for their unwavering support, inspiring discussions, and invaluable contributions throughout my academic journey.

References

Aldrich JH, Nelson FD (1984) Linear probability, logit, and probit models. Sage.

Badre D, D'Esposito M (2007) Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. J Cogn Neurosci 19:2082–2099.

Badre D, D'Esposito M (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? Nat Rev Neurosci 10:659–669.

Beattie J, Baron J (1988) Confirmation and matching biases in hypothesis testing. The Quarterly Journal of Experimental Psychology 40:269–297.

Beckmann M, Johansen-Berg H, Rushworth MFS (2009) Connectivity-Based Parcellation of Human Cingulate Cortex and Its Relation to Functional Specialization. J Neurosci 29:1175–1190.

Berkson J (1944) Application of the Logistic Function to Bio-Assay. Journal of the American Statistical Association 39:357–365.

Berkson J (1953) A Statistically Precise and Relatively Simple Method of Estimating the Bio-Assay with Quantal Response, Based on the Logistic Function. Journal of the American Statistical Association 48:565–599.

Boldt A, Gilbert SJ (2022) Partially Overlapping Neural Correlates of Metacognitive Monitoring and Metacognitive Control. J Neurosci 42:3622–3635.

Boldt A, Yeung N (2015) Shared neural markers of decision confidence and error detection. Journal of Neuroscience 35:3478–3484.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624–652.

Brown AS (1991) A review of the tip-of-the-tongue experience. Psychological Bulletin 109:204–223.

Brown R, McNeill D (1966) The “tip of the tongue” phenomenon. Journal of Verbal Learning and Verbal Behavior 5:325–337.

Burke DM, MacKay DG, Worthley JS, Wade E (1991) On the tip of the tongue: What causes word finding failures in young and older adults? Journal of Memory and Language 30:542–579.

Butterfield B, Mangels JA (2003) Neural correlates of error detection and correction in a semantic retrieval task. Cognitive Brain Research 17:793–817.

Damasio AR, Anderson SW, Heitman KM, Valenstein E (1993) Clinical neuropsychology.

Denny BT, Kober H, Wager TD, Ochsner KN (2012) A Meta-analysis of Functional Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex. Journal of Cognitive Neuroscience 24:1742–1752.

Dunlosky J, Lipko AR (2007) Metacomprehension: A Brief History and How to Improve Its Accuracy. Curr Dir Psychol Sci 16:228–232.

Dunlosky J, Mueller ML, Thiede KW (2016) Methodology for investigating human metamemory: Problems and pitfalls.

Eickhoff SB, Heim S, Zilles K, Amunts K (2006) Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. Neuroimage 32:570–582.

Eickhoff SB, Paus T, Caspers S, Grosbras M-H, Evans AC, Zilles K, Amunts K (2007) Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. NeuroImage 36:511–521.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage 25:1325–1335.

Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proceedings of the National Academy of Sciences of the United States of America 113:7900–7905.

Fernandez-Duque D, Baird JA, Posner MI (2000) Executive attention and metacognitive regulation. Consciousness and cognition 9:288–307.

Flandin G, Friston KJ (2019) Analysis of family-wise error rates in statistical parametric mapping using random field theory. Hum Brain Mapp 40:2052–2054.

Flavell JH (1979) Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist 34:906–911.

Fleming SM (2017) HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. Neurosci Conscious 2017:nix007.

Fleming SM, Dolan RJ (2012) The neural basis of metacognitive ability. Philosophical Transactions of the Royal Society B: Biological Sciences 367:1338–1349.

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

Fleming SM, Lau HC (2014) How to measure metacognition. Frontiers in Human Neuroscience 8 Available at: https://www.frontiersin.org/articles/10.3389/fnhum.2014.00443 [Accessed August 15, 2022].

Fleming SM, Van Der Putten EJ, Daw ND (2018) Neural mediators of changes of mind about perceptual decisions. Nature Neuroscience 21:617–624.

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

Frith CD (2012) The role of metacognition in human social interactions. Philosophical Transactions of the Royal Society B: Biological Sciences 367:2213–2223.

Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. NeuroImage 19:200–207.

Glasser MF, Coalson TS, Bijsterbosch JD, Harrison SJ, Harms MP, Anticevic A, Van Essen DC, Smith SM (2018) Using Temporal ICA to Selectively Remove Global Noise While Preserving Global Signal in Functional MRI Data. Neuroimage 181:692–717.

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC (2016) A multi-modal parcellation of human cerebral cortex. Nature 536:171–178.

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Van Essen DC, Jenkinson M (2013) The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage 80:105–124.

Gold JI, Shadlen MN (2007) The neural basis of decision making. Annual Review of Neuroscience 30:535–574.

Goupil L, Kouider S (2016) Behavioral and neural indices of metacognitive sensitivity in preverbal infants. Current Biology 26:3038–3045.

Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, Moeller S, Xu J, Yacoub E, Baselli G, Ugurbil K, Miller KL, Smith SM (2014) ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage 95:232–247.

H FJ (1976) Metacognitive Aspects of Problem Solving. The nature of intelligence Available at: https://cir.nii.ac.jp/crid/1572543025488914944 [Accessed September 28, 2022].

Hargis MB, Yue CL, Kerr T, Ikeda K, Murayama K, Castel AD (2017) Metacognition and proofreading: the roles of aging, motivation, and interest. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 24:216–226.

Hart JT (1965) Memory and the feeling-of-knowing experience. Journal of Educational Psychology 56:208–216.

Harty S, Murphy PR, Robertson IH, O'Connell RG (2017) Parsing the neural signatures of reduced error detection in older age. Neuroimage 161:43–55.

Hayden BY, Pearson JM, Platt ML (2011) Neuronal basis of sequential foraging decisions in a patchy environment. Nat Neurosci 14:933–939.

Johansen-Berg H, Behrens TEJ, Robson MD, Drobnjak I, Rushworth MFS, Brady JM, Smith SM, Higham DJ, Matthews PM (2004) Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. Proceedings of the National Academy of Sciences 101:13335–13340.

Kappes A, Harvey AH, Lohrenz T, Montague PR, Sharot T (2020) Confirmation bias in the utilization of others' opinion strength. Nature neuroscience 23:130–137.

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

Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. Science 302:1181–1185.

Koechlin E, Summerfield C (2007) An information theoretical approach to prefrontal executive function. Trends Cogn Sci 11:229–235.

Kononowicz TW, van Wassenhove V (2019) Evaluation of Self-generated Behavior: Untangling Metacognitive Readout and Error Detection. J Cogn Neurosci 31:1641–1657.

Koriat A (1993) How do we know that we know? The accessibility model of the feeling of knowing. Psychological Review 100:609–639.

Koriat A (2011) Subjective Confidence in Perceptual Judgments: A Test of the Self-Consistency Model. Journal of Experimental Psychology: General 140:117–139.

Koriat A (2012) The self-consistency model of subjective confidence. Psychological Review 119:80–113.

Koriat A, Adiv S (2015) The Self-Consistency Theory of Subjective Confidence (Dunlosky J, Tauber S (Uma) K, eds). Oxford University Press. Available at: http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199336746.001.0001/oxfordhb-9780199336746-e-18.

Krueger PM, van Vugt MK, Simen P, Nystrom L, Holmes P, Cohen JD (2017) Evidence accumulation detected in BOLD signal using slow perceptual decision making. J Neurosci Methods 281:21–32.

Lebreton M, Abitbol R, Daunizeau J, Pessiglione M (2015) Automatic integration of confidence in the brain valuation signal. Nat Neurosci 18:1159–1167.

Luria AR (2012) Higher Cortical Functions in Man. Springer Science & Business Media.

Mai JK, Majtanik M, Paxinos G (2015) Atlas of the Human Brain. Academic Press.

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, Peters MAK, Lau H (2016) Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. Attention, Perception, and Psychophysics 78:923–937.

Mansouri FA, Buckley MJ, Tanaka K (2022) The neural substrate and underlying mechanisms of executive control fluctuations in primates. Prog Neurobiol 209:102216.

Mazancieux A, Fleming SM, Souchay C, Moulin CJA (2020) Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. J Exp Psychol Gen 149:1788–1799.

Mazor M, Friston KJ, Fleming SM (2020) Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. eLife 9:1–34.

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. J Neurosci 33:1897–1906.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.

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.e6.

Mobbs D, Hassabis D, Yu R, Chu C, Rushworth M, Boorman E, Dalgleish T (2013) Foraging under competition: the neural basis of input-matching in humans. J Neurosci 33:9866–9872.

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

Murphy PR, Robertson IH, Harty S, O RG (2015) Neural evidence accumulation persists after choice to inform metacognitive judgments.

Musslick S, Cohen JD (2021) Rationalizing constraints on the capacity for cognitive control. Trends Cogn Sci 25:757–775.

Nelson T 0, Narens L (1990) METAMEMORY: A THEORETICAL FRAMEWORK AND NEW FINDINGS.

Nelson TO (1984) A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychological Bulletin 95:109–133.

Nelson TO (1996) Consciousness and metacognition. American psychologist 51:102.

Neubert F-X, Mars RB, Sallet J, Rushworth MFS (2015) Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. Proceedings of the National Academy of Sciences 112:E2695–E2704.

Nickerson RS (1998) Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology 2:175–220.

Passingham RE, Stephan KE, Kötter R (2002) The anatomical basis of functional localization in the cortex. Nat Rev Neurosci 3:606–616.

Philipp AM, Weidner R, Koch I, Fink GR (2013) Differential roles of inferior frontal and inferior parietal cortex in task switching: evidence from stimulus-categorization switching and response-modality switching. Hum Brain Mapp 34:1910–1920.

Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafò MR, Nichols TE, Poline J-B, Vul E, Yarkoni T (2017) Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat Rev Neurosci 18:115–126.

Qiu L, Su J, Ni Y, Bai Y, Zhang X, Li X, Wan X (2018) The neural system of metacognition accompanying decision-making in the prefrontal cortex. PLoS Biology 16.

Rahnev D, Koizumi A, McCurdy LY, D'Esposito M, Lau H (2015) Confidence leak in perceptual decision making. Psychological science 26:1664–1680.

Rangelov D, Mattingley JB (2020) Evidence accumulation during perceptual decision-making is sensitive to the dynamics of attentional selection. Neuroimage 220:117093.

Rolls ET, Huang C-C, Lin C-P, Feng J, Joliot M (2020) Automated anatomical labelling atlas 3. NeuroImage 206:116189.

Rolls ET, Joliot M, Tzourio-Mazoyer N (2015) Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. NeuroImage 122:1–5.

Rollwage M, Loosen A, Hauser TU, Moran R, Dolan RJ, Fleming SM (2020) Confidence drives a neural confirmation bias. Nature Communications 11.

Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916–1936.

Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79:217–240.

Shenhav A, Cohen JD, Botvinick MM (2016) Dorsal anterior cingulate cortex and the value of control. Nat Neurosci 19:1286–1291.

Shimamura AP (2000) Toward a cognitive neuroscience of metacognition. Consciousness and cognition 9:313–323.

Smith SM et al. (2013) Resting-state fMRI in the Human Connectome Project. Neuroimage 80:144–168.

Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44:83–98.

Spisák T, Spisák Z, Zunhammer M, Bingel U, Smith S, Nichols T, Kincses T (2019) Probabilistic TFCE: A generalized combination of cluster size and voxel intensity to increase statistical power. NeuroImage 185:12–26.

Stolyarova A, Rakhshan M, Hart EE, O'Dell TJ, Peters MAK, Lau H, Soltani A, Izquierdo A (2019) Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. Nature Communications 10.

Talluri BC, Urai AE, Tsetsos K, Usher M, Donner TH (2018) Confirmation Bias through Selective Overweighting of Choice-Consistent Evidence. Current Biology 28:3128-3135.e8.

Tversky A, Kahneman D (1974) Judgment under Uncertainty: Heuristics and Biases. Science 185:1124–1131.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. NeuroImage 15:273–289.

Vogt B (2009) Cingulate Neurobiology and Disease. OUP Oxford.

Wager TD, Nichols TE (2003) Optimization of experimental design in fMRI: A general framework using a genetic algorithm. NeuroImage 18:293–309.

Wang S, Tepfer LJ, Taren AA, Smith DV (2020) Functional parcellation of the default mode network: a large-scale meta-analysis. Sci Rep 10:16096.

Wittmann MK, Kolling N, Faber NS, Scholl J, Nelissen N, Rushworth MFS (2016) Self-Other Mergence in the Frontal Cortex during Cooperation and Competition. Neuron 91:482–493.

Yamamoto T, Fukunaga M, Sugawara SK, Hamano YH, Sadato N (2021) Quantitative Evaluations of Geometrical Distortion Corrections in Cortical Surface-Based Analysis of High-Resolution Functional MRI Data at 7T. Journal of Magnetic Resonance Imaging 53:1220–1234.

Yeung N, Summerfield C (2012) Metacognition in human decision-making: confidence and error monitoring. Philosophical Transactions of the Royal Society B: Biological Sciences 367:1310–1321.

Yoon L, Kim K, Jung D, Kim H (2021) Roles of the MPFC and insula in impression management under social observation. Soc Cogn Affect Neurosci 16:474–483.

Figures

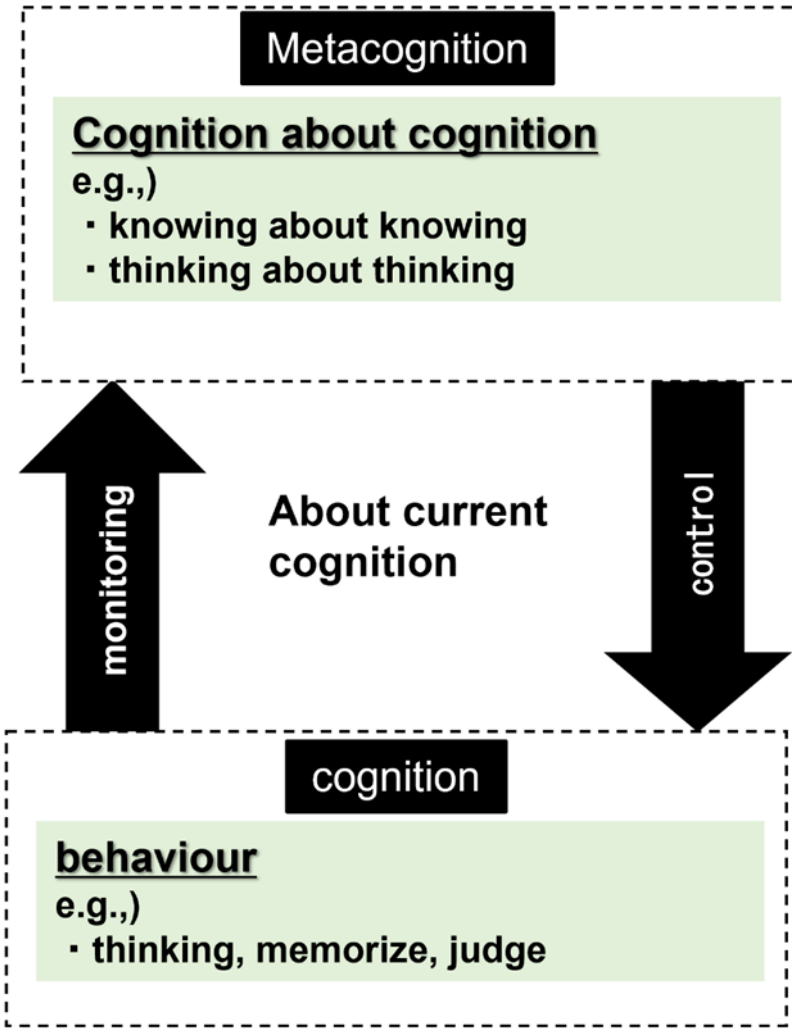

Fig. 1 Illustration of metacognitive processes.

This is the illustration of metacognitive processes. Metacognitive monitoring allows us to monitor one's own cognition, such as thinking about thinking. Metacognitive control works to regulate subsequent behavior using information obtained through metacognitive monitoring. These processes are sequential behavior.

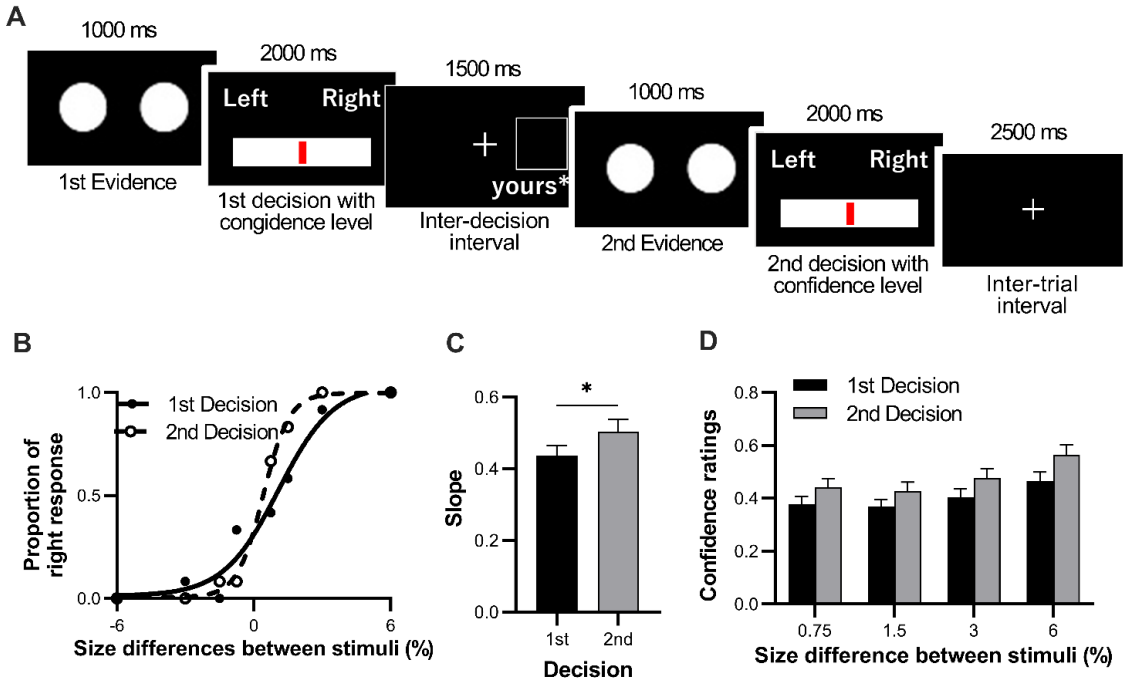

Fig. 2 Experimental task and basic behavioral results.

(A) Illustration of the experimental task within one trial. Subjects were required to discriminate which circular stimulus was larger and rate confidence on their decision. Subjects viewed the same stimuli at the timing of the first and second decisions. After each trial, any feedback on the accuracy of their decision was not presented.

(B) Psychometric function for a representative subject. The x-axis represents the standardized zero-centered difference of the size between the right and left stimuli. The y-axis is a proportion of right responses. The solid line shows the fitted psychometric curve of the 1st decision and the dotted line is the fitted one of the 2nd decision. The steepness of the slopes of the psychometric curves reflects the subject's sensitivity to changes of the size contrast between the two circular stimuli.

(C) Average values of psychometric slopes across subjects. This graph presents averaged values of slopes of the psychometric curves across subjects. The error bars are standard errors of the mean (SEM). * is $p < 0.05$.

(D) Average subjective confidence ratings across stimulus difference levels. This figure shows subjective confidence for each stimulus difference level and for the 1st and the 2nd decisions. Each bar shows the average value of confidence rating scores across subjects. The x-axis represents the stimulus difference (%), and the y-axis is the level of confidence. The error bars indicate SEM.

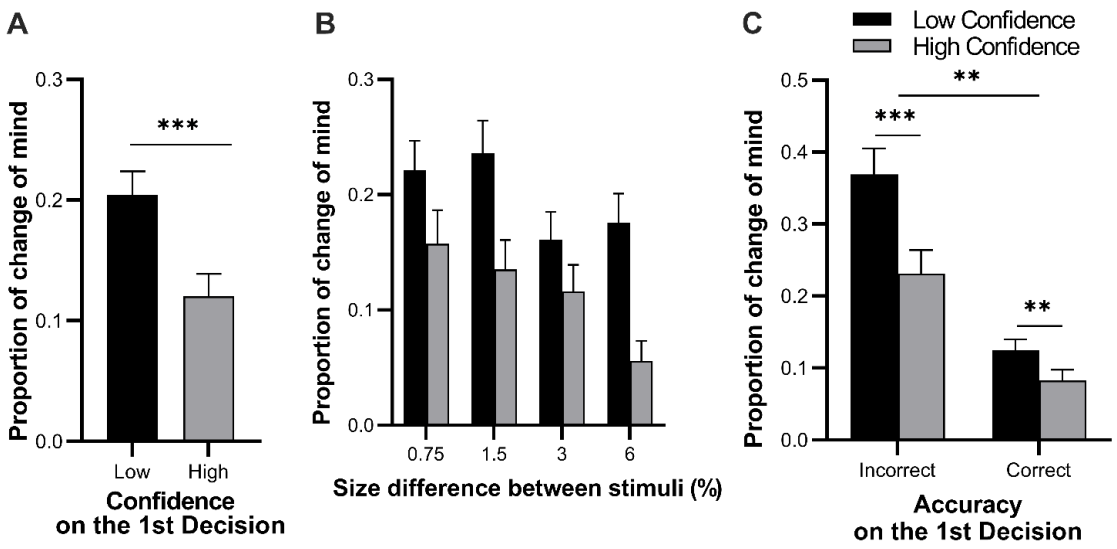

Fig. 3 Effects of confidence on change of mind and error detection.

(A) Confirmation bias. These bar graphs show the proportion of change of mind from 1st to 2nd decision (y-axis) for high and low confidence trials (x-axis). *** is $p < 0.001$.

(B) Influences of initial confidence on change of mind. This panel shows the influences of initial confidence on change of mind for each stimulus difference level. The x-axis is the stimulus difference level, and the y-axis is a proportion of change of mind.

(C) Error detection. In addition to the initial confidence, this graph includes the information of the initial response (correct/incorrect). This enables the examination of whether the initial confidence had a specific function of switching/staying choices in error and correct trials. The error bars are SEM. ** is $p < 0.01$. *** is $p < 0.001$.

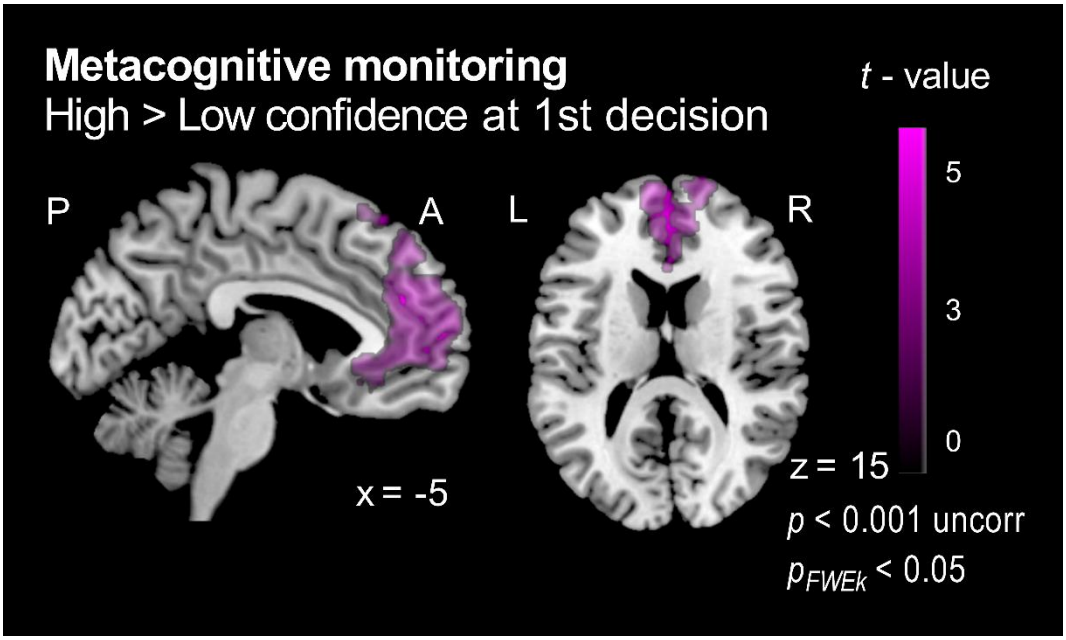

Fig. 4 Neural activity associated with metacognitive monitoring.

This figure displays group-level statistical maps depicting the neural activity related to metacognitive monitoring. To investigate the influences of initial confidence, I compared the neural activity in high confidence trials to that in low-confidence trials for each subject. A random effects model was used for group-level analysis, with significance set at $p < 0.05$ (corrected at the cluster level with a height threshold of $p < 0.001$). The color bar represents the *t*-value, while the "x" and "z" denote the x and z-coordinates in MNI space. The abbreviations used are as follows: P for posterior, A for anterior, L for left, and R for right.

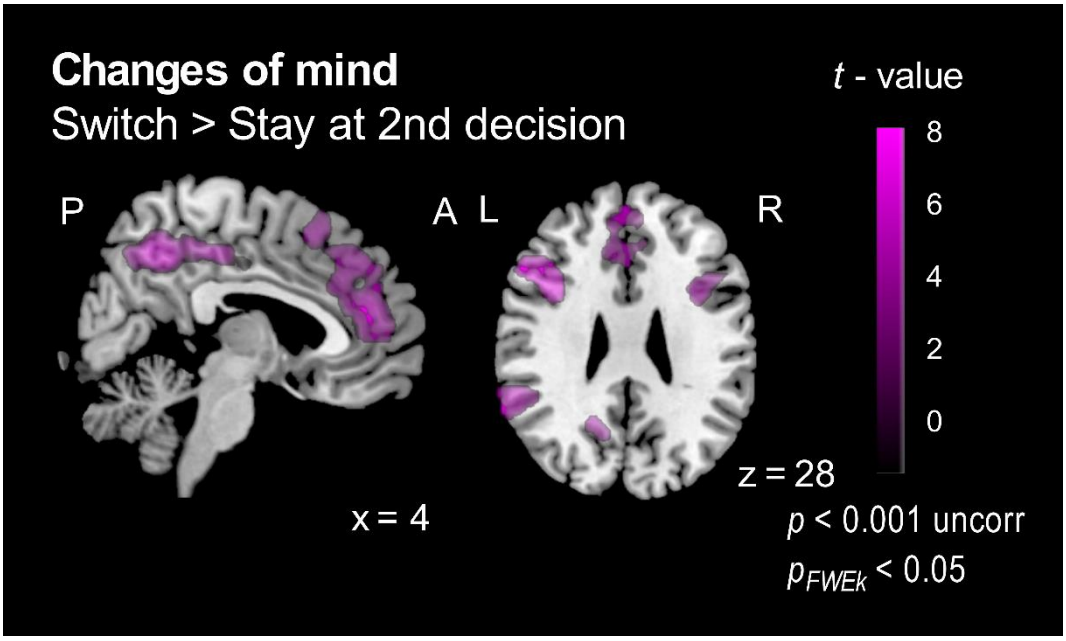

Fig. 5 Neural activity related to change of mind.

I compared the neural activation when subjects changed their initial decision with the activation when they did not change their initial decision during the second decision phase (change > no change), by comparing switch to stay activity [$p < 0.05$ FWE corrected at cluster level with height threshold of $p < 0.001$]. The color code represents t-values. The "x" and "z" denote the x and z-coordinates in MNI space. P, posterior; A, anterior; L, left; R, right.

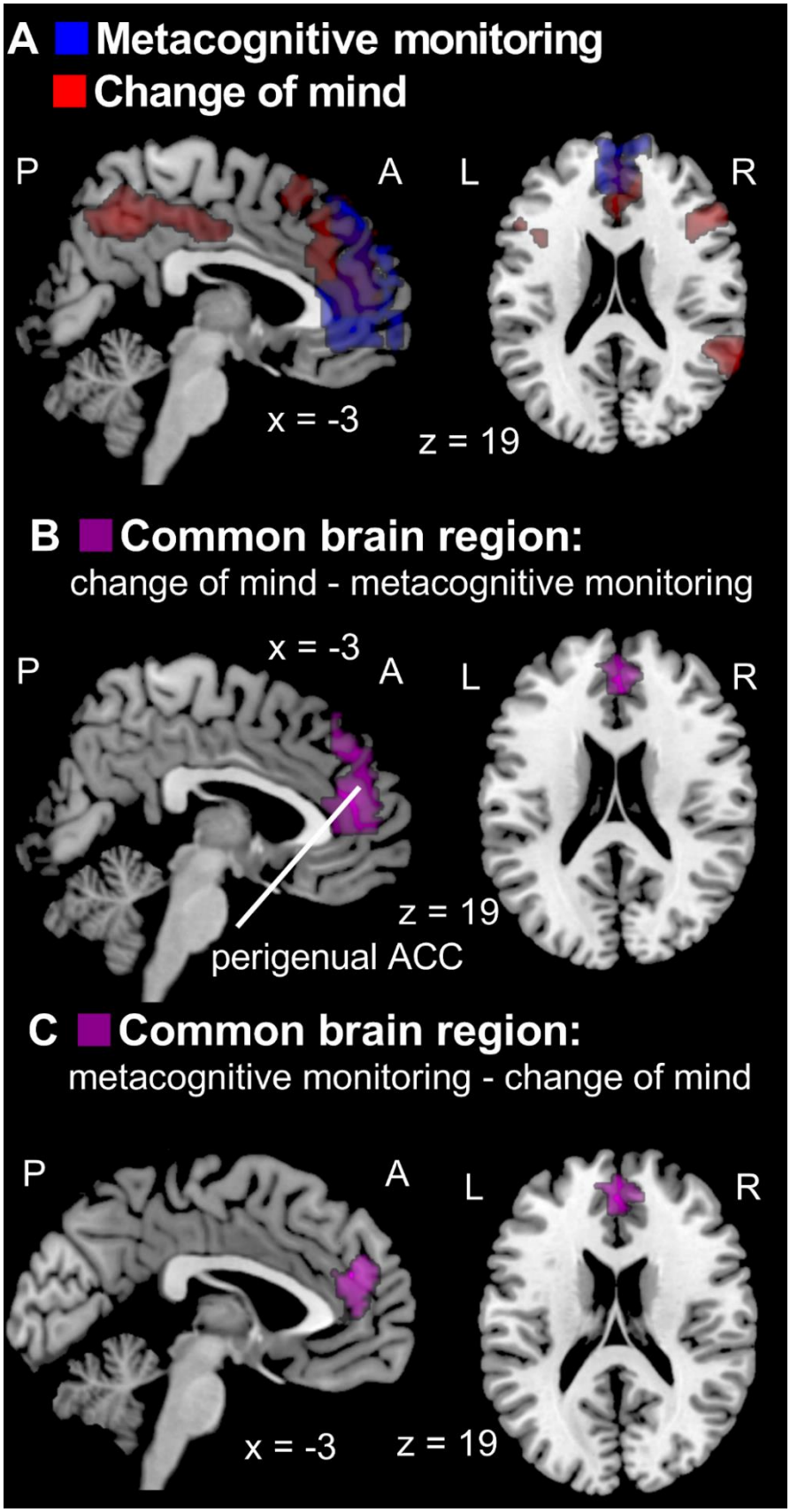
A Metacognitive monitoring
Change of mind
P
A
L
R
x = -3
z = 19
B Common brain region:
change of mind - metacognitive monitoring
x = -3
P
A
L
R
z = 19
perigenual ACC
C Common brain region:
metacognitive monitoring - change of mind
P
A
L
R
x = -3
z = 19

Fig. 6 Spatial distribution of activated regions for two processes.

This figure illustrates that spatial distribution of neural activity related to change of mind (red color) and its shared region with metacognitive monitoring (blue color) (panel A). The violet regions represent common activated regions between metacognitive monitoring and change of mind (panel B & C). The binary mask of metacognitive monitoring related regions [$p < 0.05$ F*I* corrected at cluster level with a height threshold of $p < 0.001$] was applied to change of mind related activity [$p < 0.05$ FWE corrected at cluster level with a height threshold of $p < 0.001$, panel B]. Conversely, the binary mask of change of mind related regions [$p < 0.05$ FWE corrected at cluster level with a height threshold of $p < 0.001$] was applied to metacognitive monitoring related activity to confirm reproducibility of the result of panel B [$p < 0.05$ F*I* corrected at cluster level with a height threshold of $p < 0.001$]. The “x” and “z” denote the x and z-coordinates in MNI space. P, posterior; A, anterior; L, left; R, right.

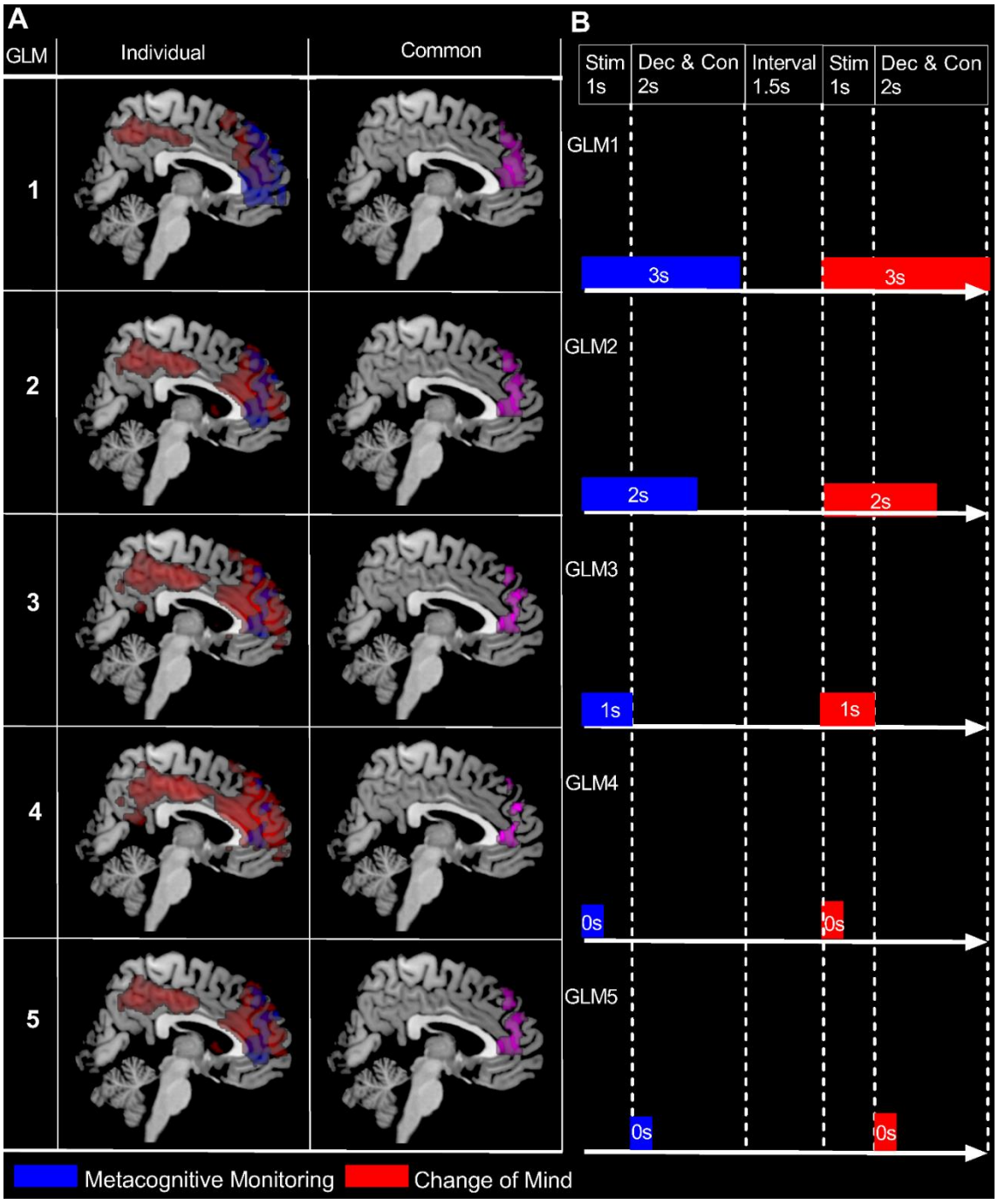

Fig. 7 Summary figure of neural substrates involved in metacognition and change of mind.

This figure illustrates that neural activity associated with metacognitive process and change of mind (A) and a graphical representation of its GLM (B). Although I have already found the neural activity related to each of the two processes, there is possibility that these results are an artifact of closer timing or time length of the two event regressors. To confirm whether these activities depend on similarity of the two regressors, I analyzed fMRI data with varying timings and durations, where the onset is either stimulus presentation (GLM1-4) or the beginning of decision making and confidence rating (GLM5). Each color represents neural activity related to metacognitive monitoring (blue), change of mind (red) and common regions (violet).

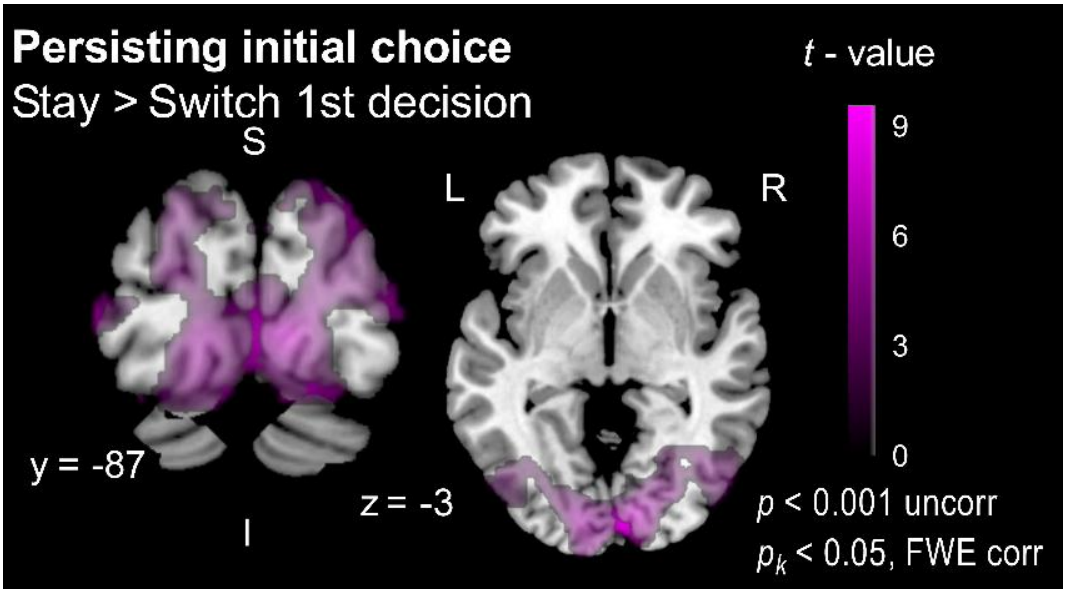

Fig. 8 Neural activity of persisting choices.

I compared the neural activation when subjects persisted their initial decision with the activation when they changed their initial decision during the second decision phase (stay > switch). By comparing activity of stay with switch, I found the neural representation in the primary visual cortex [$p < 0.05$ FWE corrected at cluster level with height threshold of $p < 0.001$]. This result, which showed neural activity mainly in the visual and motor areas, suggests that the behavior of repeating the same decision as the initial decision already devoted cognitive resources to the first decision, and that the subsequent decision activated brain regions associated with visual response and motor function to make decisions. The color code represents *t*-values. The “y” and “z” denote the x and z-coordinates in MNI space. S, superior; I, inferior; L, left; R, right.

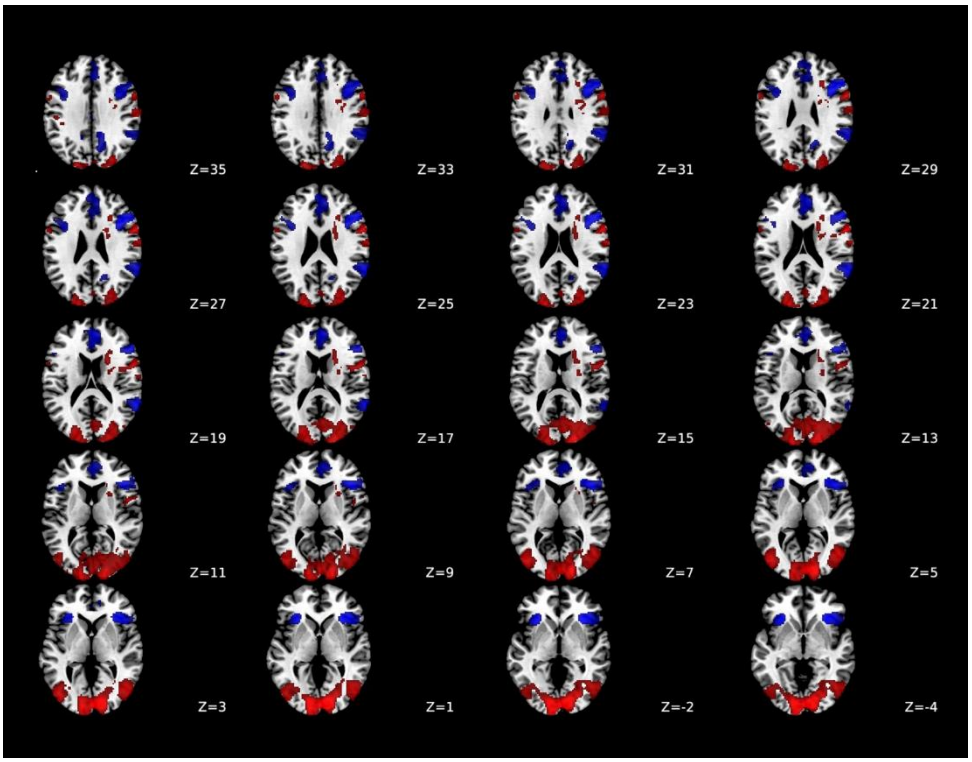

Fig. 9 Spatial distribution of the neural representation related to subsequent decision-making.

This figure shows that the spatial distribution of the neural activity of change of mind (switch > stay during 2nd decision, blue) and persisting initial choice (stay > switch during 2nd decision, red).

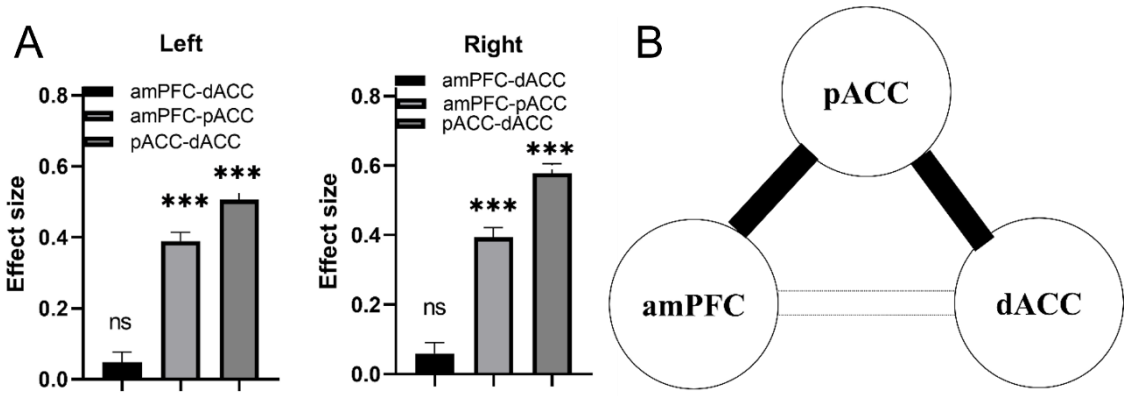

Fig. 10 Functional connectivity analysis.

These figures show the results of functional connectivity analysis between the amPFC, dACC, and perigenual ACC (pACC). (A) Results of partial correlation analysis. The x-axis is relationship between each ROI, and the y-axis is the mean of the standardized effect size across subjects. Each ROI was defined from the same hemisphere, "Left" and "Right" side. The error bars are SEM. "***" is $p < 0.001$, and "n.s." means no significant difference. (B) Summary of functional connectivity. The dotted line between the amPFC and dACC shows the no significant connectivity. The black lines between the amPFC and pACC, the pACC and dACC mean significant connectivity.

Tables

High > Low Confidence at the 1st Decision						
Cluster	*MNI coordinates*				*Peak*	
Size (mm^3)	*x*	*y*	*z*	*Hemi (L/R)*	*t* - value	*Region Label*
30510	-3	45	21	L	5.82	anterior medial Prefrontal
	-5	63	15	L	5.62	anterior part of the medial
	8	45	10	R	5.03	perigenual anterior cingulate

Table 1. MNI coordinates of neural correlates of metacognitive monitoring.

Switch > Stay at the 2nd Decision						
Cluster	*MNI coordinates*				*Peak*	
Size (mm^3)	*x*	*y*	*z*	*Hemi (L/R)*	*t* - value	*Region Label*
19450	31	26	-3	R	7.18	anterior insular cortex
	42	26	6	R	6.34	inferior frontal gyrus, pars triangularis
	44	19	26	R	6.32	inferior frontal gyrus, pars triangularis
4440	-29	26	-3	L	7.16	anterior insular cortex
	-42	19	8	L	4.37	inferior frontal gyrus, pars triangularis
3790	47	-20	-11	R	6.65	superior temporal sulcus
6090	64	-48	19	R	6.63	inferior parietal lobule
	47	-44	32	R	4.66	inferior parietal lobe
15890	8	-52	43	R	6.59	precuneus
	-10	-52	45	L	5.67	precuneus
	14	-48	37	R	5.58	precuneus
14930	3	45	19	R	5.75	perigenual anterior cingulate cortex
	1	35	21	R	4.85	dorsal anterior cingulate cortex
	8	19	54	R	4.71	superior frontal gyrus
3820	-36	11	28	L	5.02	inferior frontal gyrus, pars opercularis
	-51	22	23	L	3.51	inferior frontal gyrus, pars triangularis

Table 2. MNI coordinates of neural correlates of change of mind.

Common Brain Region: *Change of mind masked with metacognitive monitoring*						
Cluster	*MNI coordinates*				*Peak*	
Size (mm^3)	*x*	*y*	*z*	*Hemi* (*L*/*R*)	*t* - value	*Region Label*
7290	3	45	19	R	5.75	perigenual anterior cingulate cortex
	10	48	6	R	4.50	perigenual anterior cingulate cortex
	5	43	39	R	4.45	superior frontal gyrus
Common Brain Region: *Metacognitive monitoring masked with change of mind masked*						
Cluster	*MNI coordinates*				*Peak*	
Size (mm^3)	*x*	*y*	*z*	*Hemi* (*L*/*R*)	*t* - value	*Region Label*
7290	-3	45	21	L	5.82	anterior part of the medial prefrontal cortex
	8	45	10	R	5.03	perigenual anterior cingulate cortex
	-1	41	2	L	4.83	perigenual anterior cingulate cortex

Table 3. MNI coordinates of common brain regions between metacognitive monitoring and change of mind.

Stay > Switch on the 2nd Decision						
Cluster	*MNI coordinates*				*Peak*	
Size (mm^3)	*x*	*y*	*z*	*Hemi (L/R)*	*t* - value	*Region Label*
94860	12	-87	-3	R	9.70	primary visual cortex
	-12	-91	-3	L	8.04	primary visual cortex
	-12	-78	-16	L	7.50	primary visual cortex
4540	57	6	21	R	6.46	premotor cortex
	60	9	34	R	4.99	premotor cortex
	51	4	13	R	4.98	primary motor cortex
5350	-23	-63	58	L	6.19	intra parietal sulcus
	-29	-50	56	L	4.57	intra parietal sulcus
7320	23	-61	56	R	5.36	intra parietal sulcus
	31	-50	58	R	4.71	intra parietal sulcus
	21	-72	58	R	4.68	intra parietal sulcus
5180	21	13	19	R	4.73	caudate
	25	22	28	R	4.24	superior frontal gyrus
	23	-4	21	R	4.02	caudate
3590	60	-20	37	R	4.67	supra marginal gyrus
	64	-11	23	R	4.64	supra marginal gyrus
	53	-28	43	R	3.84	supra marginal gyrus
3590	-23	-11	63	L	4.46	dorsal premotor cortex
	-5	-7	58	L	4.09	supplementary motor cortex
	-14	-11	67	L	3.84	primary motor cotex
2420	-53	-22	41	L	4.44	premotor cortex
	-40	-31	37	L	3.50	premotor cortex

Table 4. MNI coordinates of Stay (persisting initial choice) related neural activity.