Metacognition in distinct perceptual tasks.

# 1. Abstract

A detection task, which requires one to identify the presence/absence of the target, is different from a discrimination task, which requires one to choose the target based on certain features. The differences between detection and discrimination was shown to extend to metacognitive judgments such that observers are more accurate at judging their performance in discrimination task than detection task. Neuroimaging data also contribute to the comparison of detection and discrimination by presenting a unique quadratic effect of confidence for detection in contrast to discrimination. These differences in metacognitive performance and neural activities between detection and discrimination were explained from two perspectives. A first-order perspective adopting the unequal variance model considers them as a result of quantitative differences in the distribution of sensory inputs in detection and discrimination whereas a higher-order perspective employing counterfactual reasoning model attributes those differences to the distinctive cognitive processes involved in these two tasks. This project tested these two models using a new discrimination task that entails unequally distributed sensory inputs for two choices, to resemble a detection task from the first-order perspective. With an incomplete sample (interrupted due to COVID-19), we confirmed the discrepancy in metacognitive sensitivity between two responses in detection in contrast to discrimination and furthermore, an even wider discrepancy was observed in the new tilt recognition task. However, the unique quadratic effect of confidence in detection was not replicated in our sample and we found no evidence to support task-related confidence effects in participants’ neuroimaging data, which could be due to underpowered sample or the previous finding being noise. We thus do not make conclusions from our results until data for a complete sample is collected.

# 2. Introduction

Understanding mechanisms supporting our daily cognitive functioning has been the central aim of scientific research across disciplines for centuries. In recent decades, beyond studying cognitive performances, another line of research concerns how we monitor, interpret, evaluate and reflect on those performances: the study of metacognition. Although metacognition deals with a subjective collection of judgments, feelings and predictions etc., it can have significant impacts on objective performances/behaviours. For example, when a student is setting a revision plan, the time allocated to each subject supposedly would depends on how this person judge the extent to which they have mastered that subject. Given their importance, the mechanisms supporting metacognition have been investigated extensively in both animals and humans, leading to developments of behavioural, neural and computational models.

The research of metacognition often rests on asking subjects to provide ratings of their subjective ‘confidence’, which can be defined as ‘a belief about the validity of our own thought, knowledge or performance that relies on a subjective feeling.’ (Grimaldi, Lau, and Basso 2015). Research investigating how confidence is formed with regards to perception commonly used discrimination tasks (sometimes labeled as categorization), which involves distinguishing a target stimulus from other distractors based on certain features (e.g. whether the line is Tilted to the left or the right). However, of equal importance, another type of perceptual task also makes up the very basics of our perceptual abilities– detection tasks. In detection tasks, participants identify the presence of a stimulus (e.g. whether a dot is present or absent). Evidence has suggested detection and discrimination can lead to differences in objective performance (e.g. Mack and Palmeri (2010)), different levels of metacognitive accuracy (Meuwese et al. 2014) and neuroimaging data suggested different neural representations associated with judging confidence in these two tasks (Mazor, Friston, and Fleming 2020). Various models have been put forward to explain the differences in metacognitive judgments in detection and discrimination, which can be divided into into two camps:

* **Higher order interpretation**: Detection and discrimination recruit different higher-order cognitive processes, such as self-monitoring and control of one’s internal state. This manifests as differences in confidence judgments for the two tasks.
* **First order interpretation**: Behavioural and neural activities differ during confidence formation process as a response to the physical features of detection and discrimination at the perceptual level such as amount of sensory evidence.

In the former interpretation, confidence judgments are a product of a distinctive cognitive process, divorced from basic sensory stimuli processing. Conversely, in the second interpretation such a division between first- and higher-order cognitive modules is not necessary, and confidence judgments arise from the same inferential process as perceptual processing.

The following section will start with describing the distinctive features of detection and discrimination and behavioural evidence showing different objective performances in them. Studies compared the metacognitive performance for these two tasks will then be discussed, followed by introduction of some models that have been put forward to explain differences in these two perceptual tasks, including both higher order interpretations and first order interpretations. These two different levels of explanations also motivate this project; we aim to compare the first level account based on SDT and higher order account based on counterfactual reasoning for the distinctive patterns found in forming confidence in detection and discrimination. By adopting a discrimination task with sensory features of a detection task, the design of this study allows us to untangle the qualitative and quantitative differences in detection and discrimination which will be introduced in the end of this section.

## 2.1. Detection vs. Discrimination

### 2.1.1. Dependence/independence between detection and discrimination

How do we detect and/or discriminate something? Does one first detect something is there (e.g. ‘I saw something is there’) and then further categorize it to certain type (e.g. ‘The thing I saw is a car.’), or do we know what it is immediately when we see something is there (e.g. ‘I saw a car there)? This is an important consideration for a wide range of research themes, as it affects the flexibility for researchers to use these two tasks in studies of other topics such as confidence formation. Is detection simply a specific case of discrimination between signal and noise? Is discrimination the same as the detection of distinguishing target features? Mack and Palmeri (2010) used a paradigm to assess the level of dependence of detection and discrimination at the basic-level (car vs. boats) and superordinate level (e.g. car vs. people). They found that detection and basic-level object categorisation is supported by single mechanism as their results revealed a dependence between the success for the detection task and the success for basic-level categorisation task. However, the success for one to correctly categorize a target at superordinate level is not dependent on the success at detection, which suggests distinctive mechanisms supporting detection and superordinate categorization. Their findings reflect the importance of distinguishing between different levels of discrimination (i.e. basic, ordinate and super-ordinate), challenging the view of one underlying perceptual ability for, or at least strong dependence between detection and discrimination suggested previously by Grill-Spector and Kanwisher (2005). Grill-Spector and Kanwisher (2005) concluded that “As soon as you know something is there, you know what it is.”, based on the finding that participants’ response time and accuracy was similar between object categorization and object identification, a deduction now seems to be crude and challenged by more recent findings such as Mack and colleagues’ work.

Another study compared detection and discrimination found that the time taken to make a response in detection is significantly different in two choices whereas it was similar in discrimination (Mazor, Friston & Fleming, 2020). Participants were faster in giving Yes responses than No responses, and in contrast, judging whether the line was tilted clockwise or anticlockwise took similar time. The shorter reaction time in Yes response was consistent with the long-established pattern that people are generally faster in giving positive response than negative responses, despite it being true or false (Wason 1959). The differences in reaction time between two responses suggests that detection and discrimination are two distinctive tasks that would lead participants to behave differently, which again highlighting the importance of differentiating the two tasks for future research.

### 2.2.2. Absence in detection

Another important difference between detection and discrimination tasks is that in discrimination tasks sensory evidence is available for both alternative choices whereas in detection, virtually no sensory evidence is available for participants to make an absence decision. This makes the “absence” condition in detection tasks a particularly interesting case: how do we detect nothing, based on nothing? The concept of nothing sparked strong interests from philosophers and scientists for centuries, who have discussed this from different perspectives. Neuroscientists are interested in how our brain encodes something that is not there and translate it into sensory experience which is later used for decisions. The absence of a stimulus is traditionally thought to be coded by the brain as baseline activity in contrast to more robust firing of neurons in the presence of a stimulus. Contrariwise, when stimulus with strength on a continuing spectrum was shown, only those evoking activities beyond the threshold were considered to be perceived as present (Victor De Lafuente and Ranulfo Romo 2005). However, despite the lack of sensory evidence in absence conditions, recent evidence has suggested that neurons encode both stimulus absence and presence information (e.g. Katharina Merten and Andreas Nieder (2012)).

The study by Merten and Nieder (2012) was designed to investigate the neural representations of absence in rhesus monkeys dissociated from motor actions. They suggested that the lack of evidence supporting absence activation in previous studies was due to the common action-based design, in which the animal was trained to carry out certain actions associated with the stimulus. Given previous finding that when no report was required, neurons encode information for both alternatives in a discrimination task (Bennur and Gold 2011) , thus in an action-independent framework, both alternatives for detection, presence and absence, might also be encoded by neurons. Using single neuron recordings, they found that in the prefrontal cortex (PFC) not only there are cells selectively encodes abstract Yes decisions, but there also exists a group of neurons that fired at an increased rate to No decisions. From their findings, it was shown that neurons in the PFC, known to be involved in high level cognitive functioning, is recruited in rhesus monkeys during stimulus present trials as well as absence trials despite the fact that no sensory evidence is available in absence trials.

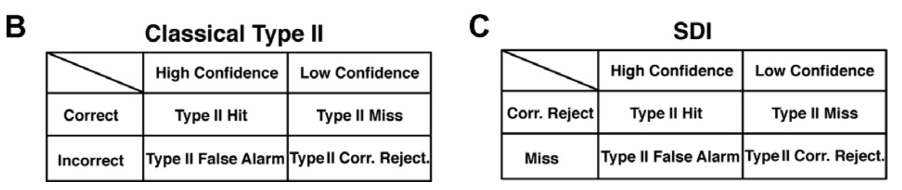
A similar line of research has investigated the representation of ‘empty set’, the precursor for zero, in the brain. As trivial as this concept might appear, it actually makes up a fundamental part of the mathematical continuum that has not been investigated until recent decades (Rinaldi and Girelli 2016). How does the brain react to an empty set, i.e. nothing in contrast to other numbers? It has been shown that a group of neurons in the posterior parietal cortex of monkeys were activated in response to numerosity ‘zero’ in a numerical operation task (Okuyama, Kuki, and Mushiake 2015). Furthermore, Okuyama (2015) and colleagues found two types of ‘zero’ neurons in the ventral intraparietal (VIP) area, that was shown to process inputs from visual, somatosensory and motor areas (Grefkes and Fink 2005) and known for its involvement in visual perception of motion (Field, Biagi, and Inman 2020). In the VIP area, they found an exclusive group that showed selective response to zero target in contrast to non-zero targets and a continuous group that encodes both zero and non-zero numerosities (from 0 to 4) with strength on a continuous spectrum. Their results showed the presence of active encoding of numerosity zero, i.e. empty set, both when treating it as a continuous and categorical variable. Similar results were found by (Ramirez-Cardenas, Moskaleva, and Nieder 2016). By training monkeys to perform a number matching task, they found that not only all recorded neurons in VIP area and prefrontal cortex responded to empty sets, some of them also discharged maximally to empty sets in comparison to other numerosities, i.e. zero selective, which constitutes an even higher proportion of VIP cells in comparison to findings from Okuyama, Kuki, and Mushiake (2015).

Together, this evidence suggests that the lack of sensory inputs in “absence” or “zero” does not necessarily imply inactivity of neurons in the brain. These findings challenge the belief that neural activity is proportional to the strength of sensory inputs, while it remains unclear where those representations of ‘nothing’ arises from and what consequences they lead to. One way to interpret the neural activities in response to ‘empty set’, for instance, is that they held explicit information of ‘nothing’ as working memory which is later used for the task as suggested by Okuyama, Kuki, and Mushiake (2015). Others have proposed that the lack of sensory signals leads brain to generate a quantitative representation of zero abstractly that is later positioned on a numerical continuum (Ramirez-Cardenas et al., 2016). Nonetheless, the uniqueness and level of cognitive demands required for representing nothing is unequivocal, which again potentially distinguishes detection tasks from discrimination tasks. This raises the interesting question of how subjective judgments are formed in detection, and in decisions about stimulus absence more specifically.

## 2.2. Forming confidence in detection and discrimination

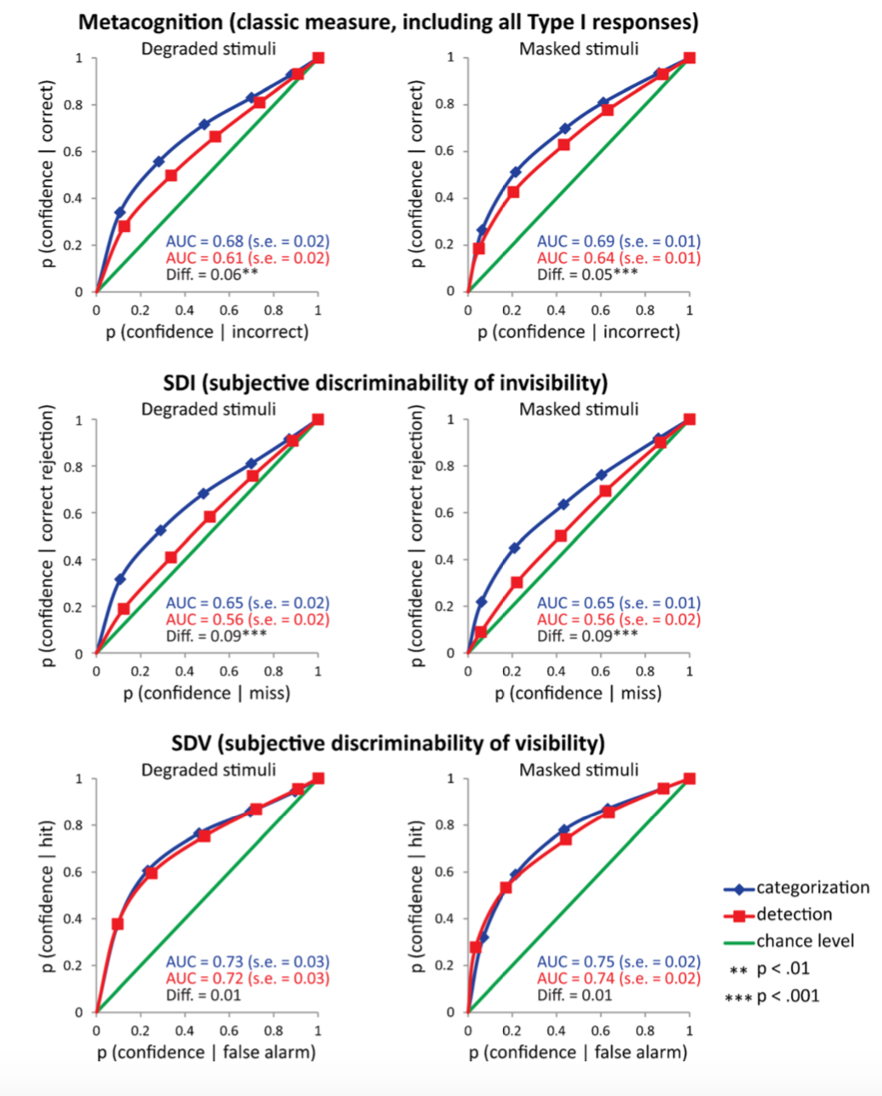
The first-order differences between detection and discrimination performance sparks interests in studying the differences in second-order subjective experiences in these two perceptual tasks such as confidence. In psychophysics, the level of confidence observers assign to their decisions is studied in relation to objective performance. Participants are asked to give high confidence ratings when they believed their choice was correct and low confidence ratings if they are less certain. To maximally incentivize this manner of behaviour, participants are often offered bonus for accurate judgments of confidence in relation to their objective performance. The correspondence between objective performance and subjective report reflects one’s ability to have insight into the objective correctness of their response, which is also known as one’s metacognitive sensitivity/accuracy. Metacognitive sensitivity is sometimes quantified as the area under the pseudo type-2 receiver operating characteristics curve (Type-II ROC), which is computed by combining confidence ratings with objective performances (Fleming and Lau 2014). A higher area under the curve (AUC) in Type-II ROC curves imply higher level of metacognitive sensitivity. Research has shown that individual’s metacognitive sensitivity can vary, depending on the domain of interests (e.g. memory vs. perception) and type of task (e.g. detection vs. discrimination). The following paragraphs in this section will present some studies showing distinct metacognitive performance found in detection and discrimination.

Meuwese et al. (2014) found that metacognitive ability is higher in categorization/ discrimination task than in the detection task for both masked and degraded stimulus. All participants performed the detection and discrimination task, which require them to identify the presence of an animal (e.g. Was there an animal present?) or identify the category (e.g. Was the animal a cat?). The stimuli were either masked by textured patterns or degraded by means of phase scrambling. Participants were then asked to rate how confident they were in the correctness of the choice they made on a scale of 1 to 6. The objective performance in detection and discrimination were also matched at the level of 71% correct to control for the potential confound of objective performance on confidence judgments. Three measures were included for metacognition: 1.) the classic AUC measure, which reflects the consistency between subjective confidence ratings of responses made and the actual, objective performance; 2.) Subjective Discriminability of Invisibility (SDI), which includes only trials of miss and correct rejections (i.e. trials in which participants reported a stimulus was absent developed by Kanai, Walsh, and Tseng (2010)); 3.) Subjective Discriminability of Visibility (SDV) which includes only trials of hits and false alarms (i.e. trials in which participants reported a stimulus was present, see figure below).



Classic Type 2 measure and SDI, taken from Kanai, Walsh, and Tseng (2010)

They found that metacognitive sensitivity, according to the classic measure, is higher for discrimination than detection task (Fig.3). Participants can more accurately evaluate the correctness of their choice when the judgement is about which category the target belongs to than when the it is about whether a target was present. To further unpack the elements driving the differences between detection and discrimination, the metacognitive sensitivity for hits, misses, correct rejections, and false alarms were calculated separately. They found that lower metacognitive accuracy in detection compared with discrimination is driven by lower metacognitive accuracy for correct rejections, i.e. participants subjective judgments of their choice being correct is less accurate for an absence decision when it is physically not there compared with a choice of the stimuli not belonging to the target group when it actually doesn’t. Their results revealed several interesting aspects of confidence formations in these two tasks; people tend to have worse insight for detection than discrimination task, but only in situations where they have correctly rejected the target.



Metacognition scores for detection and categorization tasks.Adapted from Meuwese et al. (2014).

Consistent with findings in Meuwese et al. (2014), Ariel Ezylberberg et al. (2012) found that observers were only influenced by sensory evidence in favor of their decision when forming confidence judgments, while the sensory evidence for the unselected choice did not impact such judgments. This positive evidence bias was observed in two discrimination tasks, luminance comparison and random dot motion task. This bias may provide an explanation for the poor metacognitive performance in detection ‘No’ responses. Consider the situation where participants have made a correct “absence” decision, their confidence rating would be in theory be strongly influenced by evidence in favour of this choice according to the findings from Zylberberg (2012). Hitherto, the lack of positive sensory evidence in these ‘absence’ trials might underlie a disturbance to a conventional confidence formation process that is engaged in present trials or discrimination trials. Although it is unclear what the consequences of such disturbance are, e.g. does this give rises to the employment of alternative processes such as counterfactual reasoning (discussed in section 2.4 below) specific for absent trials, the uniqueness of confidence formation in absent trials in detection is clear. Therefore, given the finding that participants were strongly influenced by the evidence for selected choice, the fact that in detection, virtually no sensory evidence is available in absence trials to support confidence judgments, positing a disadvantage for metacognitive judgments. However, research has shown that not all types of subjective judgments towards absence decisions are disadvantaged and by manipulating the source of uncertainty in the task, one’s metacognitive performance can vary significantly (Kanai, Walsh & Tseng, 2010).

The study by Kanai, Walsh, and Tseng (2010) demonstrated that failure of visual perception is not always reflected by one’s subjective reports. They used a detection task in which participants were asked to report the presence of the target stimuli. Stimulus visibility was manipulated in six different ways, which were categorized in two groups: attentional manipulations (dual task, attentional blink and spatial uncertainty) and sensory manipulations (low contrast, backward masking and flash suppression). Participants reported their confidence ratings of the correctness of their response after each trial. Beyond using the traditional Type II AUC measurements as shown in Fig.1A, they also developed a new measure termed Subjective Discriminability of Invisibility (SDI) based on computing the Type II performance but only for correct rejections and misses, i.e. trials in which participants reports absence of the stimulus correctly or incorrectly, respectively (Fig.1B). They found that SDI is significantly higher when stimulus visibility was degraded using attentional manipulations compared to sensory manipulations. In other words, participants can accurately judge whether their choice is a correct rejection or a miss when the difficulty arise as a result of increasing attentional demand, more than when the difficulty arise as a result of lacking sensory input. Therefore, the source of noise in perceptual tasks is an important consideration for observers to make confidence judgments; confidence is often adjusted accordingly when noises arises from one’s internal cognitive capacity such as lack of attention whereas the impact of the physical property/environment of perceptual stimuli is not picked up by observers during metacognitive judgments.



ROC analysis of SDI for group data under visibility suppression.Adapted from Kanai and Walsh (2010)

Taken together, these studies have shown that making subjective confidence judgments for decisions about which category the target belongs to is simpler than making subjective confidence judgments for decisions about whether something was present or not (Meuwese et al., 2014). Furthermore, the poor metacognitive performances in detection tasks seemed to be driven by the absence conditions, in which participants do not have access to sensory evidence to the selected choice that could potentially impair their metacognitive judgement according to positive evidence theory (Zylberberg et al., 2012). However, the impairment of metacognition in absence trials was also found to be dependent on the source of noise, such that observers can recognize cognitive difficulties (e.g. attentional blink) and make confidence ratings accordingly while they did not take into account of sensory difficulties for confidence ratings.

### 2.3. Neural underpinnings of confidence

What brain regions support the formation of metacognitive judgments? Developments in neuroimaging method allow us to reveal the neural underpinnings of confidence formation in both animals and human. Studies have shown rostrolateral and dorsolateral prefrontal cortices are key brain areas involved in confidence judgments in human ( Fleming, Huijgen, and Dolan (2012) ; see S. M. Fleming and Dolan (2012) for a review). However, as metacognitive judgments often co-occur with decision-makings, some researchers suggest that confidence judgments share a neural system employed by decision making, while others presented evidence in favour of an independent neural network specific for metacognitive judgments (Qiu et al., 2018). Qiu and colleagues thus conducted a study to locate the neural systems supporting metacognition independent of decision-makings, in which they identified an independent network in the PFC supporting metacognition.

Given the behavioural studies described in previous section, showing different metacognitive performance in discrimination and detection, it is reasonable to expect distinctive neural representations supporting these judgments. Mazor, Friston, and Fleming (2020) thus investigated the neural contributions to confidence judgments in detection and discrimination tasks. Participants were asked to judge whether a grating was present (detection) or to discriminate between two grating orientations (discrimination), after which they reported their confidence ratings. Their results revealed distinct neural representations of confidence in detection and discrimination in several areas. Specifically, a more pronounced quadratic effect of confidence was observed in the frontopolar cortex (FPC) in the detection task compared with the discrimination task, which provides neural data to demonstrate that making subjective judgments differs between detection and discrimination. As discussed before, this is likely to be due to a unique subjective evaluation process related to absence trials, which was supported by the finding that quadratic confidence effects were stronger for absence decisions than presence decisions in the right temporoparietal junction (rTPJ). The rTPJ was known to be involved in various pivotal cognitive functions such as attentional control (Marois et al., 2004), awareness of actions (Blakemore, Wolpert, and Frith 2002) and social cognition (Decety and Lamm 2007) etc.. Mazor and colleagues proposed two different approaches to explain this pattern of activation in rTPJ, which will be discussed in section 2.4.

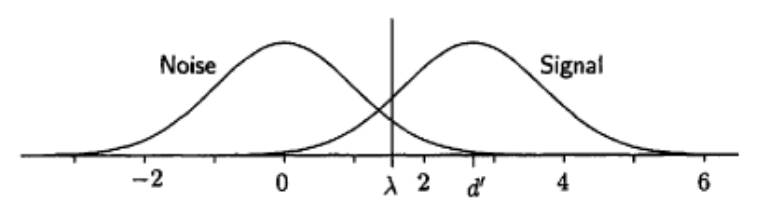
Previous research in metamemory also contributes to the discussion of making confidence judgments about absence. Miyamoto et al. (2018) found that Brodmann area 10, which is known to be involved in a series of high order executive functions, is also recruited for metacognitive evaluation of non-experienced events in macaque monkeys. The frontopolar cortex plays a causal role to confer subjective experiences of non-experienced events according to their results. This finding reflects the possibility that the frontopolar cortex is involved in making confidence judgments of absence since absence can be regarded as ‘non-experience of the stimulus’, which would reflect a qualitative distinctiveness of absence and hence detection from discrimination.

## 2.4. Models of confidence formation

After reviewing the literatures on detection and discrimination, this section will discuss some computational models that explain confidence formation generally, namely signal detection theory (SDT) and Bayesian models. We then introduce the two explanations we propose to account for differences in metacognition between detection and discrimination based on the first- and higher-level of interpretations.

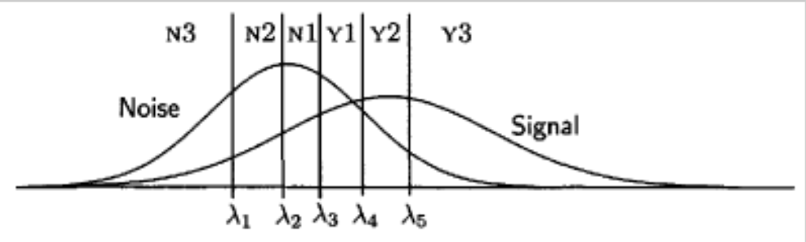
### 2.4.1.Confidence in SDT

The most common application of signal detection theory is to provide a measure for calculating the ability of a classifier/observer to distinguish signal from noises (Wickens 2002). The most basic representation of a binary perceptual task form SDT is shown in figure below, where the two distributions corresponds to the probability/likelihoods of the detected stimuli being a signal or noise given certain amount of evidence on the x axis. The distance between the peak of the noise and signal distributions corresponds to the level of sensitivity of the observer which known to be dprime (d’) and the position of criterion (λ) determines how an observer responds given certain amount of evidence.



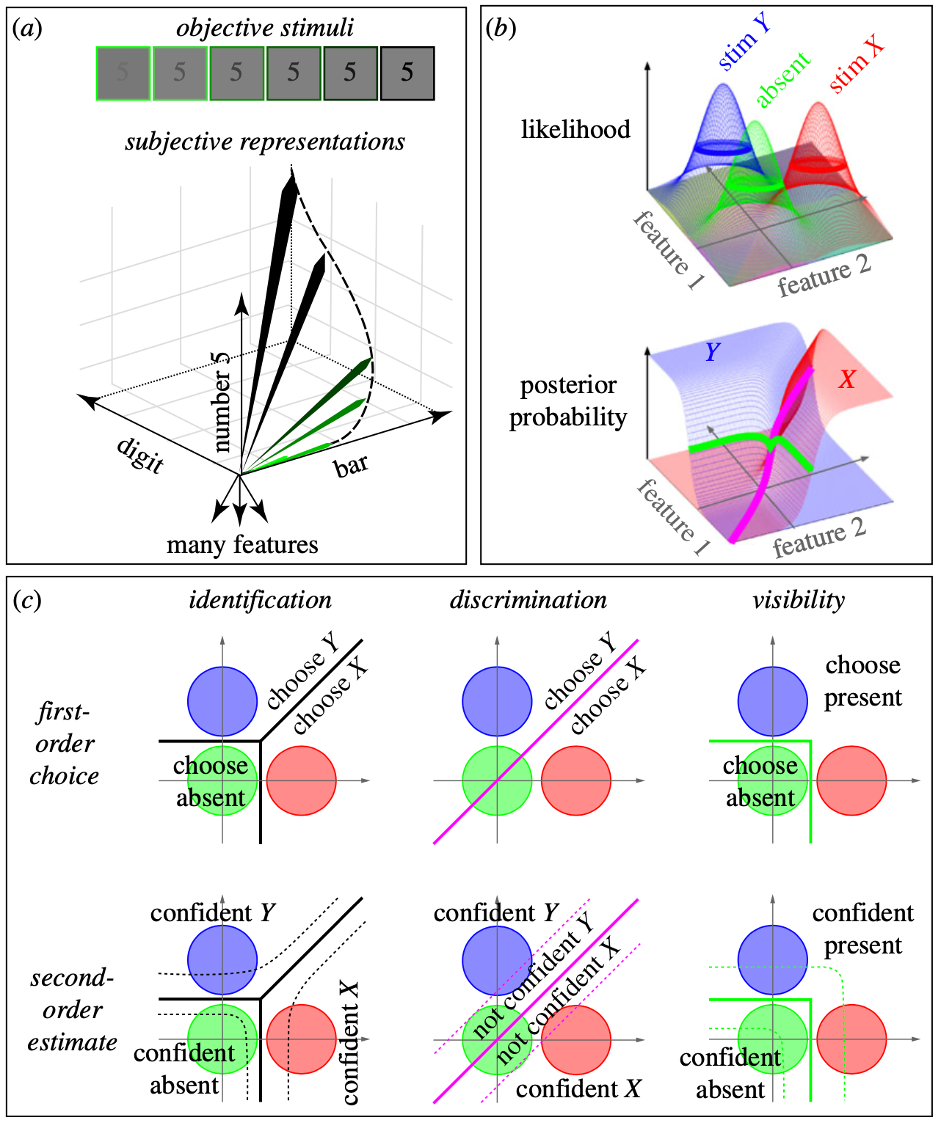
An equal-variance Gaussian distribution model of SDT, from Wickens, 2002.

SDT not only explains how observers distinguishes signal from noise, it can also be adapted to account for confidence rating process (Wickens, 2002). In an elementary detection framework, only one criterion is needed for signal and noise decisions. However, by placing various decision criterions that correspond to different confidence levels for each response, observers’ performance can be measured in a task requiring confidence ratings (see figure below). In this adapted version of SDT model, the amount of evidence supporting one choice is regarded as relational to the confidence towards that choice and this is visualised as the distance of the sensory sample from the criterion on the distribution plot.



Confidence rating in SDT. From Wickens, 2002

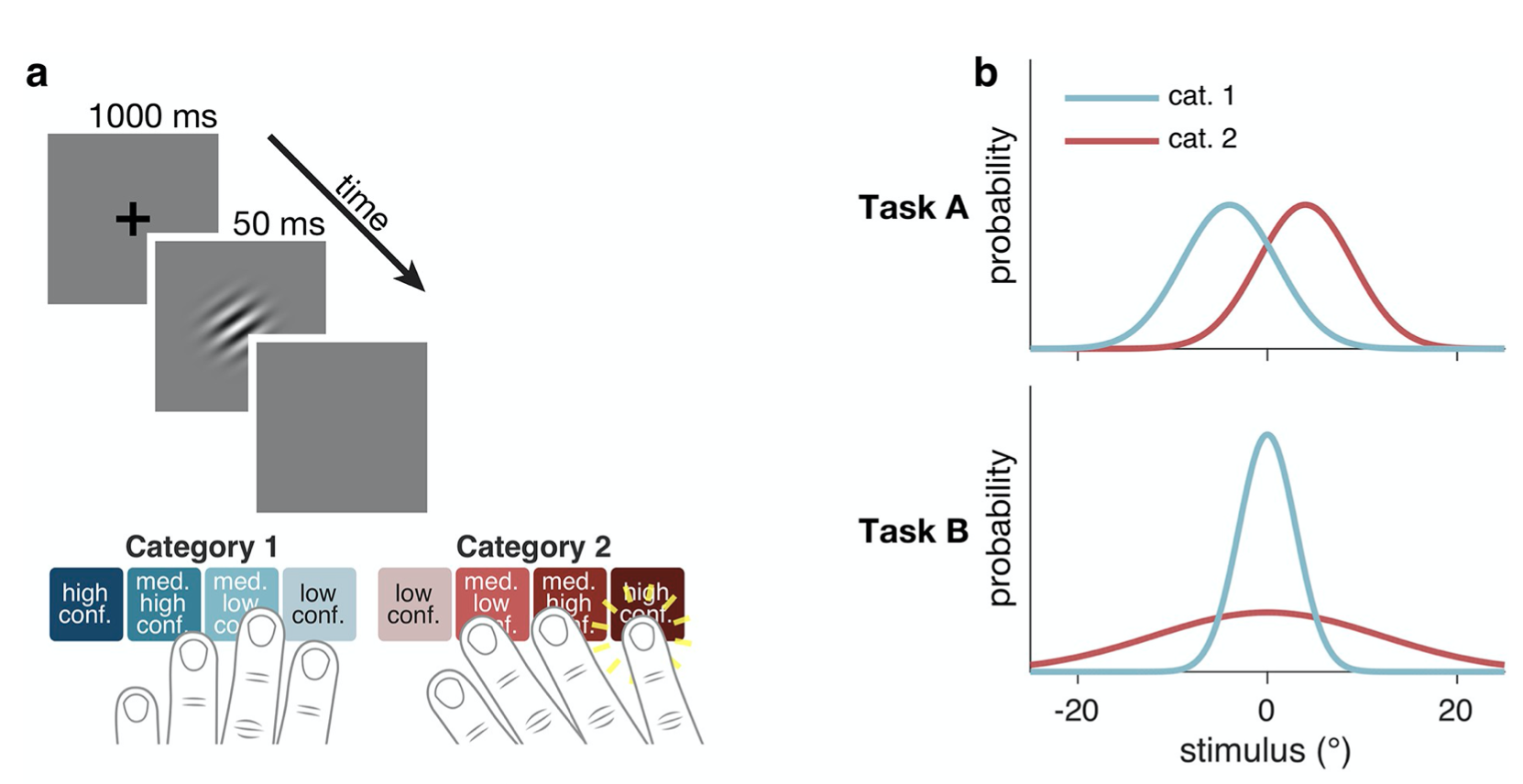
Based on SDT and Bayesian inference, a first-order model was put forward previously to explain the process of forming subjective reports (King and Dehaene 2014). In this model, subjective judgments were based on the same inferential process as objective perceptual decisions; both influenced by the strength of the sensory stimuli and prior beliefs. According to this model, the content of subjective report such as confidence is the outcome of an inferential decision process, influenced by the stimuli inputs, with a rich set of different output classes that have incorporated prior beliefs, rather than a simple present or absent decision (see fig.4 below). In addition, this model shows that the formation of second-order confidence judgments which reflects the probability of a choice being correct, is unique in each task (Fig.4, panel c). In such first-order models, confidence judgments are made through the same level of cognitive process as first order judgments without involving higher cognitive spaces.



### 2.4.2. Bayesian models of confidence formation

Apart from SDT type of account for confidence formation, another widely accepted view is that human confidence report reflects a Bayesian probabilistic estimation of a choice being correct. That is the judgments of confidence are made based on the posterior probability of the decision made being correct formed by the observer.

To test this assumption, Adler and Ma (2018) compared the Bayesian and Non-Bayesian alternative models for forming confidence using two categorisation tasks. The tasks essentially require the participants to report the direction of the stimulus shown which could be a visual grating or ellipse, as well as the level of confidence in their choices simultaneously in one press. The contrast of the grating and elongation of the ellipse were manipulated to give rise to the measure of stimuli reliability and corresponds to the level of sensory uncertainty (see Fig 7).



Experimental design from Adler and Ma (2018)

They found subjects did take into account the level of sensory uncertainty, the perquisite of Bayesian reasoning for confidence judgments without forming a posterior. Confidence judgments, according to their analysis, are affected by global beliefs about sensory uncertainty in perception, but in a simpler heuristic way than fully Bayesian. To be distinguished from SDT models where confidence judgments remains unaffected by the beliefs of uncertainty and based solely on sensory inputs.

Models described so far explain the general confidence formation process. However, they do not account for the differences between detection and discrimination discussed in previous sections. In the next two sections, a qualitative account (counterfactual reasoning model) and a quantitative account (unequal variance model) will be introduced to explain some of the previous findings.

### 2.4.3. Counterfactual reasoning models

It has been argued that detection and discrimination are qualitatively different thus leading to distinct metacognitive processes. Specifically, the unique absence condition in detection might motivate participants to adopt alternative mechanisms for metacognitive judgments that is qualitatively different from mechanisms used in conventional discrimination. One proposal is that due to the lack of sensory input for absence trials, participants might instead refer to their perceived likelihood of having detected a hypothetical target, i.e. the counterfactual situation, in order to make confidence rating for absence trials. Recall the study by Zylberberg et al. (2012), in which they found that metacognitive judgments are influenced by evidence in favor of the selected choice, this positive evidence effect and lack of evidence in absence together is thus likely to encourage participants to employ a counterfactual reasoning process. Such counterfactual reasoning process is expected to be influenced by their current state of attention: participants are likely to give low confidence ratings when they evaluate themselves as less attentive as their judgments about the possibility of the counterfactual situation is less certain and vice versa. Behavioural results have supported a role of attention monitoring in subjective judgments of performance (Kanai, Walsh, and Tseng 2010). People seem to be able to distinguish their correct rejections from misses when the task difficulty arises from competing attention resources in contrast to when it arises from lack of sensory input. In discrimination task, it is also found that uncertainty related to attention influences observers’ categorization performance and confidence rating in a Bayesian style (Denison et al. 2018). Neuroimaging data also suggested a role of attention in judging subjective experiences in a perceptual filling-in task where participants confidence rating was negatively correlated with attention (Davidson et al. 2020). Therefore, the process of a counterfactual reasoning, under the influence of monitoring internal states, is possible in the absence trials for detection.

According to this account, the differences in frontopolar activations in detection and discrimination might reflect a role of this area in referring to an internal hypothetical state. This hypothesis shares a ground with previous finding that changes in the connectivity of the frontopolar cortex tracks the switching to an alternative choice (Boorman et al. 2009) as such switching requires one to first hold a hypothetical representation of the alternative choice. Similarly, the more proNounced quadratic effect of confidence found in rTPJ in absence might also be a representation of counterfactual reasoning process in those trials or an interactive modulation of attention occurred in counterfactual reasoning.

However, it is important to note here that such counterfactual model incorporating attentional control could only account for distinct activations in ‘No’ responses whereas Mazor, Friston, and Fleming (2020) found similar pattern in ‘Yes’ trials as well. Further investigations are therefore needed to confirm the role of counterfactual reasoning in detection and discrimination.

### 2.4.4. Unequal variance model

SDT does not only provide an account for general confidence formation process, it can also explain the differences in the confidence formation found in detection and discrimination in an unequal variance model.

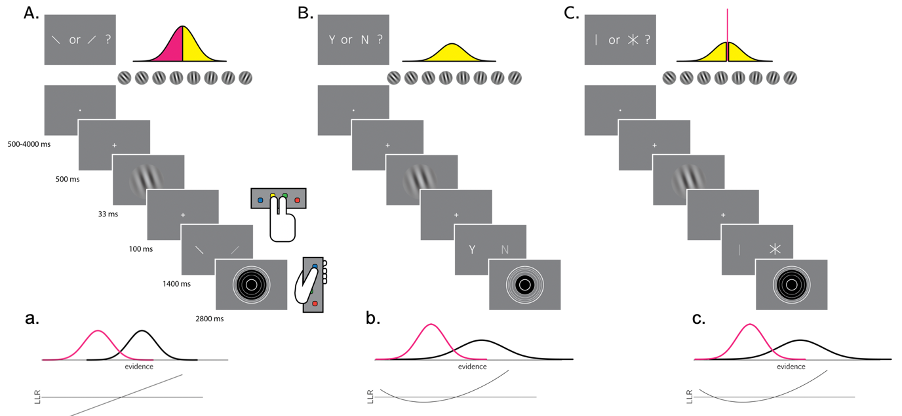
In a detection task, the sensory evidence for two competing choices is commonly found in model fits to be unequally distributed, with a greater variance in the presence trials than the absence trials (Fig.1b) as theoretically No sensory stimulus is available in absence trials. In contrast, in a typical discrimination task the two alternatives are equally distributed (Fig.1a). According to the likelihood-ratio calculation of decision making for a binary task, , equal variance implies a linear function of log likelihood-ratio with respect to sensory evidence, whereas unequal variance implies a quadratic function (Wickens 2002). Therefore, a more proNounced quadratic effect of confidence in FPC in the detection task might arise as a result of activation in this area representing the log likelihood ratio of committing to ‘Yes’ or ‘No’ responses drawing from unequally distributed evidence. In contrast, the confidence effect would be linear when neurons were representing the relative likelihood of ‘Left’ or ‘Right’ drawing from equally distributed evidence. Such unequal variance model attributes the differences in confidence judgments between perceptual tasks to the sensory properties of alternative choices, which reflects a quantitative aspect of consideration.

It is important to emphasize here that this model differs profoundly from the counterfactual reasoning model such that the former considers metacognitive judgments were produced through basic level sensory processing whereas the later treats it a distinct cognitive process that requires higher order capacities.

### 2.5. The current study

To untangle the first order and higher order models described above, the current study includes another discrimination task with unequal distributions (Fig.7C) in addition to a detection and a discrimination task that are typically used in psychophysics. This will allow us to separate the quantitative feature of the task, i.e. type of distribution, from the qualitative ones, i.e. type of task. We built on the design of Mazor, Friston and Fleming (2020) including three perceptual tasks while participants’ brain activations are being monitored in an MR scanner: A). discrimination task with equal variance (whether gratings are Tilted to the right or left; B). detection task with unequal variance (whether a grating is present or not); C). discrimination task with unequal variance (whether a stimulus is a Vertical grating or Tilted towards any direction) which we refer as tilt recognition task in this project. Participants are asked to rate their confidence about their decision following each decision.

In line with the unequal variance model, a discrimination task with the attributes of unequally distributed evidence for two alternatives would also lead to similar performance to a detection task. Therefore, if participants exhibit similar pattern of behaviour in the new tilt recognition task as that in detection, it would provide support to the first order accounts such as unequal variance model. On the other hand, if performance in the tilt recognition task more resemble that in discrimination, the higher order interpretations are more plausible. For neural data, if quadratic effects of confidence on brain activations are observed in both the discrimination task with unequal variance and the detection task, it is likely that neurons in this area represent the relative likelihood of two competing choices in perceptual tasks as predicted by the SDT model. On the other hand, if the discrimination task with unequal variance shares a similar linear activation profile with the equal-variance discrimination task, it would suggest that the difference in neural representations of detection and discrimination is driven by some qualitative features of detection, such as counterfactual reasoning proposed in current study.



Schematic representation of the experimental procedure

An important alternation we make to the task used in previous study by Mazor, Friston and Fleming (2020) is that instead of having the target and noise emerges from a blank background, in the current study, dynamic noise was used as background and signals emerged from noise instead. The appearance of noise from blank background could be argued as a form of presence, and thus this type of detection could be considered as a pseudo detection, which resembles a discrimination task, i.e. discrimination the appearance of the noise from the appearance of the target from a balk background. It was reported that this type of pseudo detection did not impair metacognitive sensitivity in absence trials as it would in true detection (Maniscalco and Lau 2011). Therefore, to ensure the use of a true detection task, the stimulus presentation method was changed in the current project.

## 2.6. Aims and Hypothesis

The current study aims to study how human form confidence in detection and discrimination tasks both from behavioural and neural perspectives. Specifically, this project aims to:

* Compare the objective performance of the new tilt recognition task with detection and discrimination
* Compare the metacognitive performance of the new tilt recognition task with detection and discrimination.
* Replicate the finding from (Mazor, Friston, & Fleming, 2020) that an interaction between task (discrimination/detection) and the quadratic effect of confidence, in medial and lateral frontopolar cortex.

We make the following predictions based on the two models of interests:

**1). Unequal variance model:**

According to this model, we expect Tilted responses to behave like Yes responses (as both correspond to the high-variance condition) and Vertical responses to behave like No responses (as both correspond to the low variance condition). We thus expect to see:

* Response time is faster for Tilted than for Vertical responses in the tilt recognition task.
* Participants are more confident in Tilted than in Vertical responses in tilt recognition task.
* Higher metacognitive sensitivity for Tilted response than for Vertical response in the tilt recognition task.
* More pronounced quadratic expansion of confidence in the frontopolar cortex in tilt recognition than in discrimination.

**2). Counterfactual reasoning model:**

According to this model, we expect the new tilt recognition task to resemble the discrimination task.

* Response time is not significantly different between Tilted and Vertical responses in tilt recognition task.
* Participants’ confidence ratings are not significantly different between Tilted than in Vertical responses.
* Metacognitive sensitivity is not significantly different between Tilted response and Vertical response in the tilt recognition task.
* The quadratic effect of confidence in the frontopolar cortex is stronger for detection than discrimination and tilt recognition and this quadratic expansion is not significantly different between discrimination and tilt recognition

# 3. Method

## 3.1. Overview

The current study involved two parts: a behavioural training session lasting around 60 mins and a scanning session lasting around 90 mins with intervals No longer than two weeks. Both parts of the experiments took place at the Wellcome Centre for Human Neuroimaging, University College London. This project followed a pre-registration (<https://github.com/efkccc/Data-analysis/blob/master/pre-registration%20document.pdf>) and open science practices. All data and analysis scripts are available on Github and this dissertation is fully reproducible via RMarkdown (<https://github.com/efkccc/Data-analysis>).

## 3.2. Participants

34 healthy participants were recruited, and according to the preregistration exclusion criteria (see section 3.3.1 for details), in total data from 25 participants were included in the final analysis. Participants received cash payments as compensation for their time, £10 for the behavioural session and £20 for the scanning session. To motivate participants to perform their best in our tasks, we also offered bonus payment for good performance and accurate confidence ratings (see procedure below for details on bonus calculation).

## 3.3. Experimental Procedure

### 3.3.1. Behavioural session

During behavioural training session, participants first received introductions of the study, including general procedure, ethic and data protection protocols. The structure of the three tasks (see fig.4 for the schematic representation) were explained to participants as following:

* Detection task: On half of the trials, a Noisy grating will appear after the fixation cross and on the other half there would be No grating shown and you need to decide whether there was a grating present.
* Discrimination task: A grating will appear on the screen every few seconds after the fixation cross, which will be Tilted Clockwise in half of the trials and AntiClockwise in other half. Participants were asked to decide which direction the grating was Tilted to.
* Tilt recognition task: A grating will appear on the screen every few seconds after the fixation cross, which will be Tilted (to any direction) in half of the trials and Vertical in other half. Participants were asked to decide whether the grating was Tilted or Vertical.
* Confidence rating: In all three tasks, immediately after making a choice, you need to indicate how confident you are in your decision by changing the size of the circle.

This session contained a practice block, a calibration block and several training blocks for all three tasks. No feedback was provided to participants about their performance. The response mapping was counterbalanced between blocks. As a result, participants alternated between two different keys for each response between blocks. For example, in the detection task, half of the blocks requires participants to press key “8” to for Yes response and in the other half they would need to press key “9” to give the Yes response. This is to avoid potential associations between the key and the response that might bias subjects’ choices. Confidence ratings were achieved by changing the size of the circle and participants were divided into two groups such that for half group of participants bigger circle corresponds to higher confidence levels and for the other half bigger circle corresponds to lower confidence levels. Such design of counterbalancing is to avoid unrelated visual activation inside the scanner, confounding the effects of confidence of our interests.

During this session, participants’ performance was controlled around 70 % accurate, by manipulating the task difficulty independently for the three tasks. This was achieved by using the common 1 up 2 down staircase procedure on stimulus visibility (discrimination and detection task) and on the standard deviation of the orientation distribution (tilt recognition). Participants were not invited back to continue the scanning session if: 1.) their accuracy were lower than 60% or higher than 80%; 2.) had strong response bias, i.e. used the same response in more than 80% of the trials; 3.) had strong confidence bias, i.e. the same confidence level was reported for more than 90% of the trials.

### 3.3.2. Scanning session

The structure of the three tasks were the same as behavioural session. To motivate participants perform we offered bonus in addition to the baseline payment for the scanning session. Bonus is calculated use following rule:  
bonus=£. Where is a vector of 1 and -1 for correct and incorrect responses, and is a vector of integers in the range of 1 to 6, representing confidence reports for all trials. The rule for bonus calculation was explained to participants in both sessions. The scanning session started with a calibration phase to further calibrate participants performance during which time the structural scan for each participant was also obtained. At scanning, 10 discrimination and detection blocks were presented in 5 scanner runs.

## 3.4. Stimulus

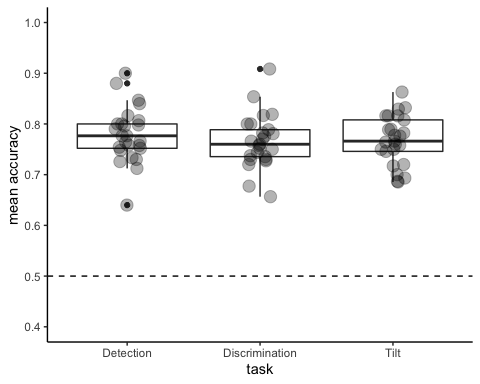
Each trial started with a fixation cross that was an indicator for participants as the beginning of tasks, which lasted for 500 milliseconds. We then presented the stimulus on the screen for 500 milliseconds, which consisted of 10 grayscale frames. Those frames were presented at 20 frames per second within a circle of diameter 3°. The details for generating stimuli are following (as stated in pre-registration <https://github.com/efkccc/Data-analysis/blob/master/pre-registration%20document.pdf>):

* Generate 10 grayscale frames ( *F* … *F* ), each an array of 142 by 142 random luminance values.
* Create a 142 by 142 sinusudial grating ( 24 pixels per period, random phase). The orientation of the grating is determined according to the trial type.
* The grating visibility for frame *i* is *pi* = *v* × *exp*(-/2) with *v* being the visibility level in this trial (0 for target-absent trials).
* For each pixel in the frame , replace the luminance value for this pixel with the luminance value of this pixel in the grating with a probability of.

## 3.5. Scanning parameters

Scanning took place at the Wellcome Centre for Human Neuroimaging, London. The structural images were obtained using an MPRAGE sequence (1x1x1 *mm* voxels, 176 slices, in plane FoV = 256x256 *mm* 2), followed by a double-echo FLASH (gradient echo) sequence with TE1=10ms and TE2=12.46ms (64 slices, slice thickness = 2 *mm*, gap = 1 *mm*, in plane FoV= 192×192 *mm* 2, resolution = 3×3 *mm* 2) that were later used for field inhomogeneity correction. Functional scans were acquired using a 2D EPI sequence, optimized for regions near the orbitofrontal cortex (3.0x3.0x3.0 *mm* voxels, TR=3.36 seconds, TE = 30 ms, 48 slices Tilted by -30 degrees with respect to the T¿C axis, matrix size = 64x72, Z-shim=-1.4).

# 4. Results

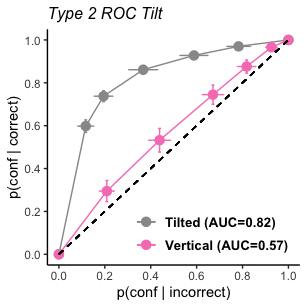
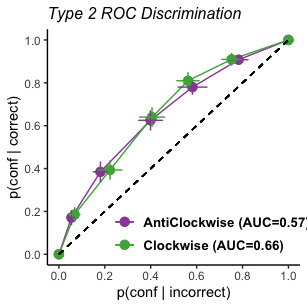
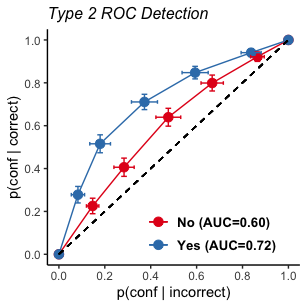


Mean accuracy across three tasks

## 4.1. Behavioural results

### 4.1.1. Performance across threee tasks

The performance across three tasks, detection (accuracy= 0.78, d’= 1.62), discrimination (accuracy= 0.77, d’= 1.60) and tilt recognition (accuracy = 0.77, d’= 1.60) was similar. Accuracy was computed as the proportion of correct trials out of total number of trials and d’ was calculated using the following formula: d’ = z(FA) – z(H). A one-way ANOVA failed to detect a significant difference between the accuracy of these three tasks , , , and d’ , , , ; see Figure )), which reflected a good control over the difficulty level and objective performance across the three tasks.



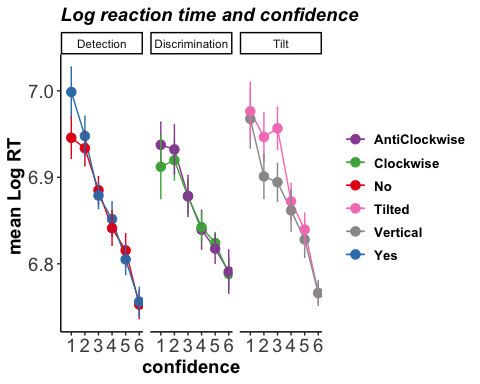
### 4.1.2. Metacognitive sensitivity

Metacognitive sensitivity, which is quantified as the area under Type 2 ROC curve (see Figure ), was significantly higher for Yes (0.72) than No (0.60) response (, 95% CI , , , ), which was consistent with previous findings in Meuwese et al. (2014) and Mazor, Friston and Fleming (2020). This suggests that participants’ confidence ratings are more diagNostic to accuracy in the judgments of a target stiZmulus being present than being absent. An even bigger discrepancy was observed in the tilt recognition task, where the AUC is significantly higher for Tilted (0.82) than Vertical (0.57) responses (, 95% CI , , , ), which suggested that the correct judgments of a stimulus being Tilted is better reflected by participants’ confidence ratings than the correct judgments of a stimulus being Vertical. In discrimination, participants’ metacognitive sensitivity was not significantly different between the two responses (, 95% CI , , , ) and this symmetry was again, consistent with previous literates (Mazor, Friston, and Fleming 2020; Meuwese et al. 2014).

### 4.1.3. Response Bias

To test whether participants were biased towards certain response, we computed the probability of giving each response in the three tasks. The probability of responding No in detection was 0.54 (± 0.07), and was significantly different from 0.5 (, ), showing that participants were more likely to give No responses than Yes response in detection task. The probability of responding Clockwise was 0.51 (± 0.11) and was not significantly different from 0.5 (, 95% CI , , , ). For the tilt recognition task, the probability of responding Tilted was 0.43 (± 0.07), which was significantly lower than responding Vertical 0.57 ± (0.07) (, 95% CI , , , ). This result reveals that participants were more biased towards giving Vertical response than tilted response.

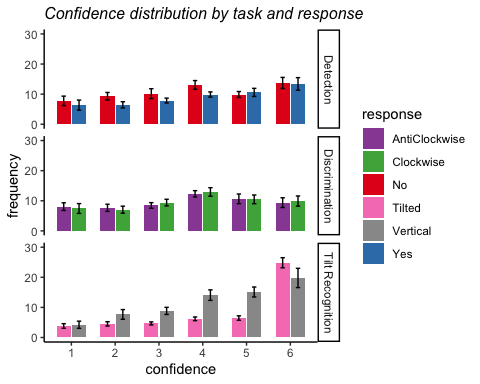
### 4.1.4. Response time



Log transformed reaction time across three tasks

Consistent with previous studies, response time was faster for correct responses (1st quantile= 908.91, median= 916.80, 3rd quantile= 916.85 milliseconds) than incorrect responses (1st quantile= 970.97, median= 975.06, 3rd quantile= 983.49 milliseconds) according to a paired t test on the log reaction time (, 95% CI , , , ). No significant overall effect of responses type on log reaction time in detection task (Yes vs. No, , 95% CI , , , ), discrimination task (Clockwise vs. AntiClockwise, , 95% CI , , , ) and tilt recognition task , 95% CI , , , ) were detected according to the results from paired sample t tests. Results of reaction time were not consistent with previous findings that participants were faster in responding Yes than No in detection task reported by Mazor, Friston and Fleming (2020). Figure shows participants log transformed reaction time as a function of confidence in the three tasks.

### 4.1.5. Confidence distributions



Confidence distributions across three tasks and responses. Error bars represent the standard error of the mean.

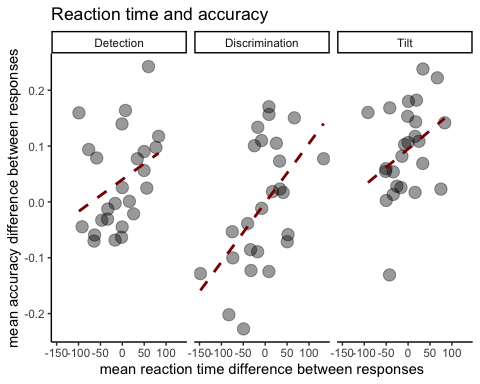
Within detection, No significant difference in mean confidence was observed between Yes (target present) and No (target absent) responses (see Figure ) (, 95% CI , , , ) and similarly, in the discrimination task participants’ mean confidence ratings did not differ significantly between the Clockwise and AntiClockwise responses (, 95% CI , , , ). This was in contrast to the previous finding that participants were more confidence in their Yes responses than No responses(Mazor, Friston, and Fleming 2020). However in the tilt recognition task, a paired sample t test revealed that the effect of response type had a significant impact on mean confidence ratings (, 95% CI , , , , such that participants are more confident in Tilted (4.75) than Vertical response (4.32).

### 4.1.6. Exploratory Analysis

Based on the unequal variance model, one of our hypotheses was that participants’ performance would be similar between detection and tilt recognition task. We thus conducted further exploratory analysis to assess such resemblance from several perspectives.

Given that participants displayed asymmetry of metacognition in both detection and tilt recognition, we conducted further exploratory analysis to assess the level of within-subject correlation in AUC, i.e. metacogntiive sensitivity, differences between two responses in detection and tilt recognition. In other words, we were interested whether participants whose metacognitive sensitivity differed greatly between Yes and No responses were also those differed significantly between Tilted and Vertical response. To test this, we first calculated the differences in AUC for each participants in detection and discrimination and fitted a Spearman’s correlation test to them, the results from which revealed no such association , 95% CI , , , . Therefore we did not establish a within-subject correlation in the differences of metacognitive sensitivity between detection and tilt recognition task.

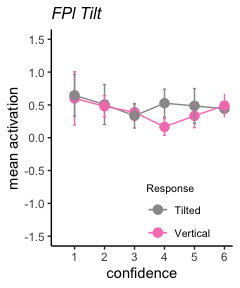
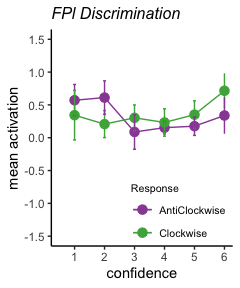
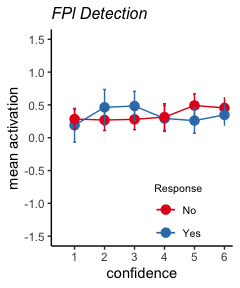
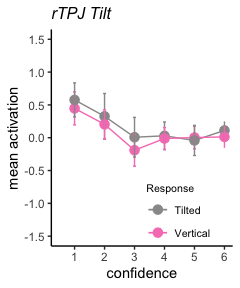
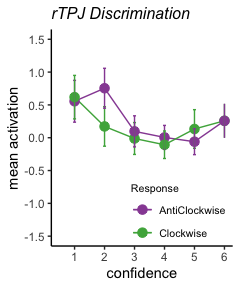
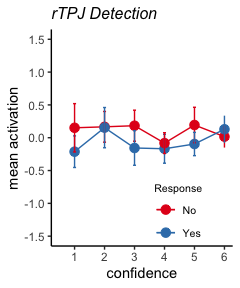
Furthermore,given the existence of significant response bias in both detection and tilt recognition task and the hypothesis for a resemblance between them two based on unequal variance model, statistical tests were conducted to assess whether individuals who are more likely to give No responses in detection are also more likely to give Vertical response in tilt recognition. A linear model was fitted to the probabilities of giving these two response within each individual and this model fit result suggest that this relationship was not significant (, 90% CI , , , ). Therefore, this within-individual correlation was not established.



Correlation between accuracy and log-transformed reaction time across three tasks.

To explore the relationship between reaction time and accuracy in each task, further linear regression analysis was carried out. Namely, we are interested in whether participants who took longer to give one response (Yes, Clockwise and Tilted) than the opposite (No, AntiClockwise and Vertical) were also more likely to be correct in those choices. We computed the accuracy difference and response time difference between each response pairs in three task and tested the level of correlation between them, the results from which reveals a significant correlation in discrimination (, 95% CI , , , ) but not in detection (, 95% CI , , , ) and tilt recognition (, 95% CI , , , ) (see Figure ).

## 4.2. Imaging results



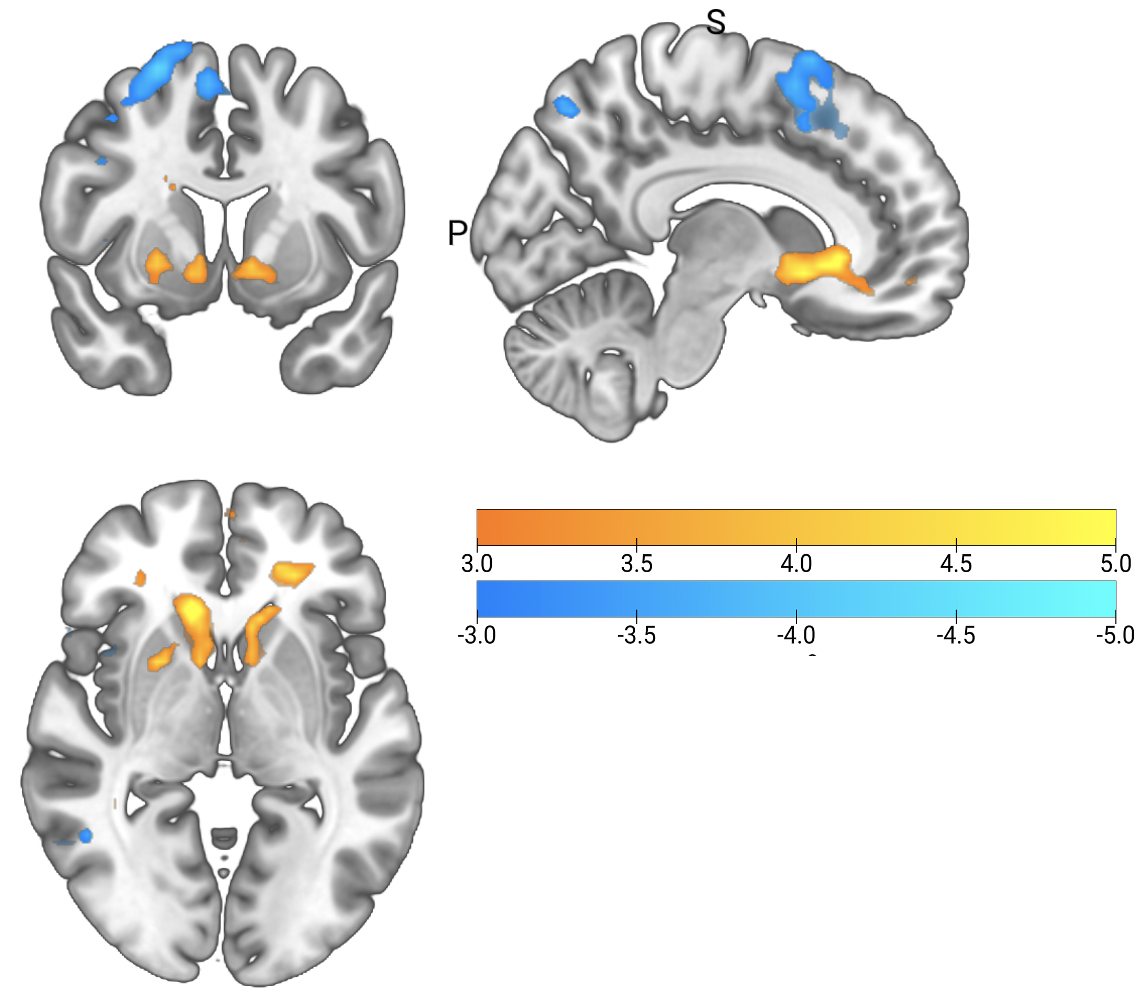
### 4.2.1. Effect of confidence on BOLD signal in ROIs

From our data, negative linear confidence-related effects were observed in right Temporoparietal Junction (rTPJ) (=-0.04, , ), Posterior Medial Frontal Cortex (pMFC)(=-0.11, , ), as well as positive linear correlation between confidence and BOLD signals in Ventromedial Prefrontal Cortex (vmPFC)(=0.12, , ), Medial Frontopolar Cortex (FPm)(=0.05, , ) and Lateral Frontopolar Cortex (FPl)(=0.01, , ).These linear relationships were consistent with long established findings on confidence research ( e.g. Morales, Lau, and Fleming (2018)).

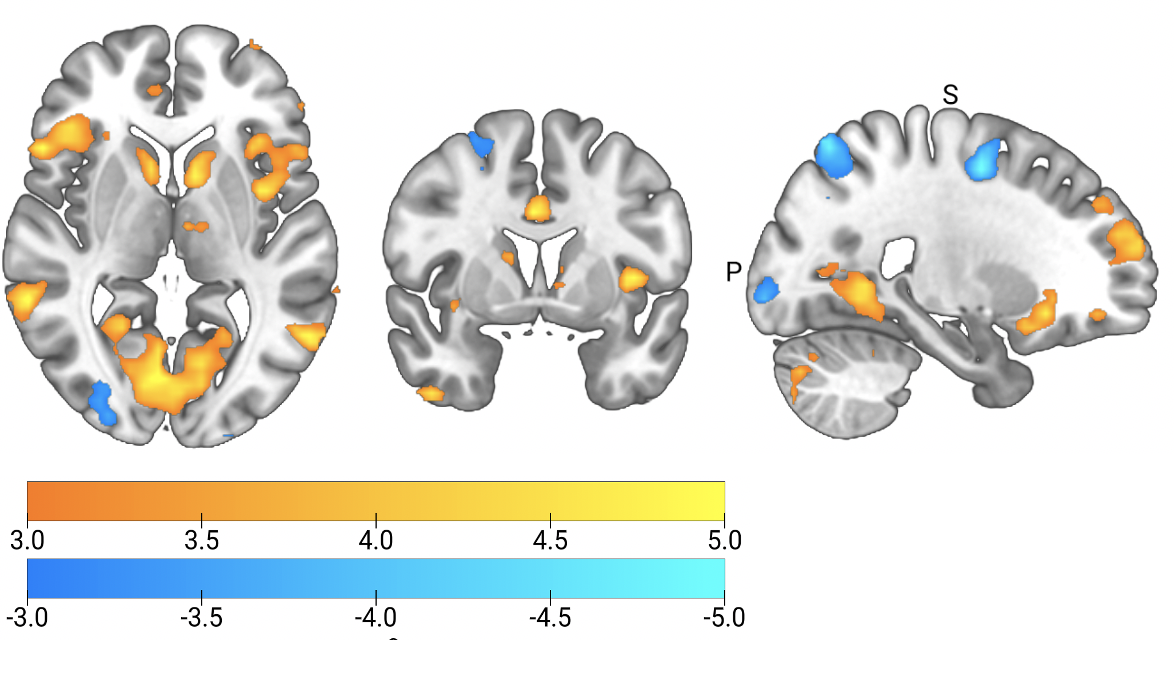
To investigate whether linear effects of confidence on BOLD signal were influenced by the type of task (Detection vs. Discrimination vs. Tilt recognition) and/or response (Yes vs. No; Clockwise vs. AntiClockwise; Vertical vs. Tilted), linear models with interactions were tested. The effects of confidence failed to show a significant difference between three tasks in all ROIs (all p values > 0.32), according to uncorrected statistical tests for multiple comparisons. However, when looking at the interaction of response within each task, giving Yes or No responses had significant impact on the linear relationship of confidence on BOLD signals in the FPl region r apa(FPl\_det, 2, T) and so did giving Clockwise or AntiClockwise (p= 0.001). In the rTPJ, the effect of response type was also significant (p= 0.041) in discrimination. No significant effect between different responses was observed for three tasks in other regions (p values > 0.13).

### 4.2.2. Whole brain analysis

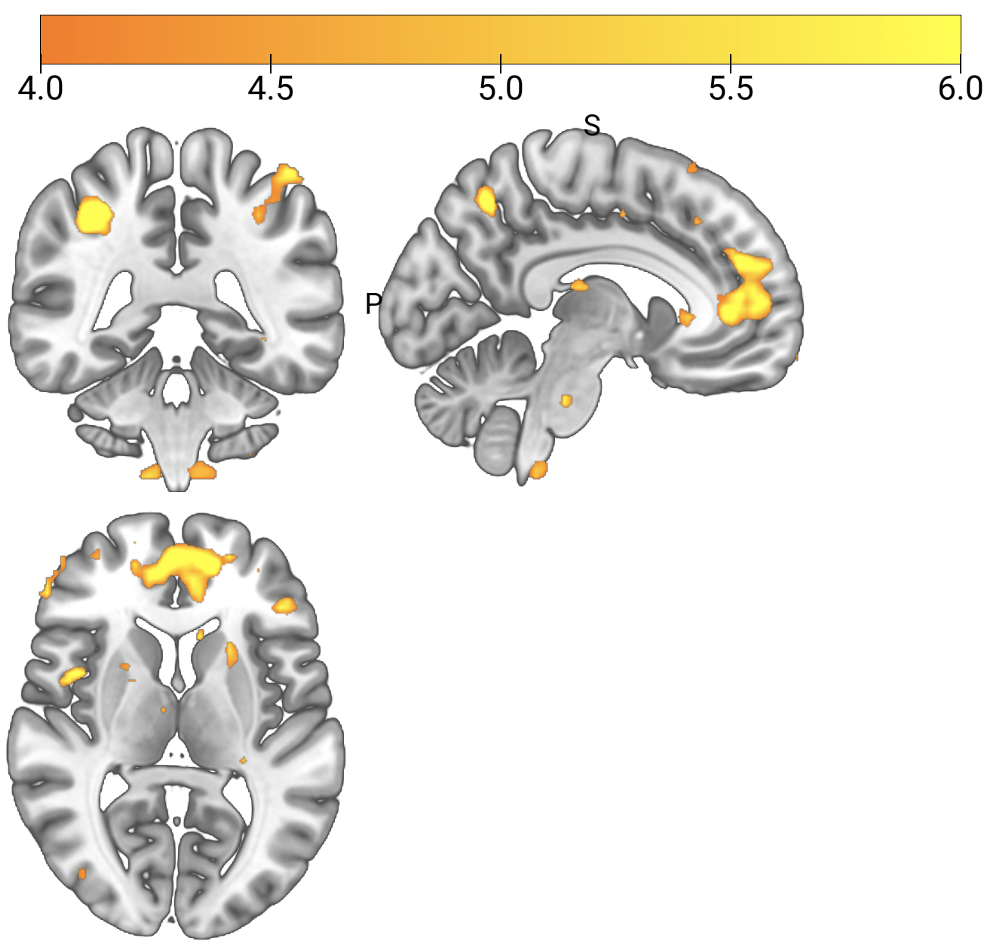
In the uncorrected whole brain analysis, we found linear relationships between confidence and activation in several brain regions, which is shown in the map below. These are the regions that were found to relate with confidence judgments repeatedly by previous studies (Morales, Lau, and Fleming 2018). However, No significant interaction of task type (detection vs. discrimination, tilt recognition vs. discrimination) was found on the quadratic effect of confidence at significant level of <.001.



Linear effects of confidence. MNI(73.68, 2.68, 2.68)



Quadratic effects of confidence. MNI(23.8, 2.81, 3.46)



Main effect of task. MNI(-6.4, -40.7, 7.15)

# 5. Discussion

Detection and discrimination are two distinctive perceptual processes that constitute some of the most fundamental tasks we perform on a daily basis. Detection involves identifying the presence of a target whereas discrimination requires one to categorise the target into certain group. The differences in the nature of those two tasks have been demonstrated in the alignment of one’s objective performance (Mack and Palmeri 2010), as well as confidence judgments with their performance (Meuwese et al. 2014), with research showing that such judgments are more accurate in discrimination than detection. In particular, the judgments of confidence for absence in detection appeared to be more difficult comparing with presence (Meuwese et al. 2014). Differences between detection and discrimination were also manifested in a unique quadratic effect of confidence that were found in rTPJ and frontopolar cortex in detection task in contrast to discrimination task (Mazor, Friston, & Fleming, 2020). The current study aims to investigate whether those differences reflect a qualitative difference, i.e. distinctive cognitive processes involved, or a quantitative difference, i.e. equal or unequal distribution of sensory evidence, in forming confidence between detection and discrimination. To test those different accounts, a novel discrimination task, which at the same time, also fulfils the characteristics of an unequally distributed detection task from the perspective of SDT, was introduced. The comparison of differences in the way participants behave between the new tilt recognition task and conventional detection/discrimination tasks would allow us to gain understandings of the nature of metacognitive judgments in detection and discrimination: a similarity between detection and tilt recognition would suggest that metacognitive judgments were made through first-order processes as sensory processing whereas a resemblance between discrimination and tilt recognition would suggest that metacognitive judgments involve higher-order cognitive capacities to be distinguished from basic sensory processing.

## 5.1. Metacognitive sensitivity in three perceptual tasks

The level of metacognitive sensitivity, which was quantified as the area under the Type-II ROC curves, was calculated for each individual that was then used for group-level analysis. Consistent with previous findings, AUC is higher for Yes than No responses in detection task. Participants’ metacognitive sensitivity did not differ significantly between the two responses in discrimination, which was also in line with previous findings. We initially hypothesized that the new tilt recognition task would resemble the detection task behaviourally as they share similar features from the perspective of SDT. This hypothesis was supported by our results showing that in the new tilt recognition task also share a discrepancy, that was even wider than detection, in the metacognitive sensitivity between the two responses.

The lower level of accuracy in metacognitive judgments in absence responses in comparison to presence response has been shown repeatedly by previous studies (e.g.Kanai et al., 2010; Mazor, Friston & Fleming 2020). This discrepancy has been explained mainly from two perspectives. The first attributes it to the stimuli property in absence trials at a basic sensory processing level. Based on SDT, the structure of the detection task entails two unequal distributions for absence and presence choices and furthermore, the widely distributed presence choices would translate into greater AUC in Type-II ROC curves (Maniscalco & Lau, 2011), which is line with our results.

However, we did not establish a within-subject correlation in the differences of metacognitive sensitivity between detection and tilt recognition task. This could be due to the fact that the noise in the two tasks arise from different sources. In detection, the task difficulty was manipulated by the contrast of the grating, where noisy conditions corresponds to low level of contrast and thus visibility. In the new tilt recognition task, the task was made difficult by reducing the angle to which the grating was tilted to so that in the more difficult conditions, the gratings look more like vertical. Therefore, orientation and visibility are two distinctive type of noise that might diver participants to adopt different decision making strategies in these two tasks.

## 5.2. Response bias

In the current study, we found that participants were more likely to give No response than Yes response while they were also more likely to give Vertical than Tilted response. The probability of giving two responses in discrimination task was well balanced. Results from the detection task was not consistent from that in Mazor et al., (2020) in which they reported no significant bias towards any response.

Since the new task was designed to resemble the SDT features of the detection task and the fact we found response bias in both, we then conducted exploratory analysis to assess whether participants who had biases toward No response in detection also share bias in tilt recognition. The response bias in these two tasks was not correlated at an inter-individual level, i.e. participants who are more likely to give No response did not necessarily tend to give Vertical responses. This again could be due to the fact that the two tasks introduce noise from different sources that had driven participants to use different decision-making strategies.

## 5.3. Response time

Results from our log transformed reaction time analysis showed that participants were generally faster at giving correct responses than incorrect responses, in line with previous results (Mazor, Fleming & Friston, 2020). This regular pattern of decision making is rather straightforward to explain from either direction: correct choices are those easier and thus faster to make or faster choices are more likely to be correct.

Effect of response type on reaction time was investigated in detection, discrimination and tilt recognition tasks respectively. No significant difference of reaction time and responses pairs (Yes/No, Clockwise/AntiClockwise and Vertical/Tilted) were identified in our analysis. This diverges from the results for detection reported in Mazor, Fleming and Friston (2020), in which they reported that Yes responses were significantly faster than No responses.

One possibility is that this differences in results arise from variations in task structures, specifically the extension of stimulus presentation window (from 33s to 500s) in the current study. It is possible that when duration of showing stimulus is long, the time taken to make a response is not as meaningful as when duration is short in informing participants decision-making process. This is because long stimulus window potentially allows one to decide during presentation window already, in which case response time captured during designed response window is merely the motor execution of that already-made decision rather than a complex cognitive process.

We also explored the relationship between any potential reaction time and accuracy trade-offs. To investigate whether participants who took longer to make certain response are also those who have higher accuracy level for these responses, we computed the reaction time differences and accuracy difference for each individual between two response types in all three tasks. In discrimination, participants who took longer to give Clockwise responses than AntiClockwise responses were also more likely to be correct in these judgments. Although this relationship was not statistically significant in the detection and tilt recognition task (which might be due to sample being underpowered, see section 5.4 below), there is some evidence to suggest a positive correlation between reaction time and accuracy in general. This pattern resembles the evidence accumulation model for decision making to certain degree, where the position of the accumulation starting point affects differences in reaction time and in accuracy in opposite directions. However, a more complete sample is required to study this further, and to establish task-specific differences in diffusion-model parameters.

## 5.4. Imaging results

Results from current project did not replicate the previous quadratic effect of confidence in the detection task in any ROIs. The stronger quadratic confidence effect was thus also not found in absence conditions. The global linear effect of confidence was found in expected brain regions, consistent with previous findings (e.g., Morales et al., 2018), which reassured the validity of the design of our study. The failure to replicate previous results could be caused by two reasons which is explained below. Due to COVID-19, the data collection of the current study was interrupted. As a result, instead of the target sample of 35 participants, we were only able to include data from 25 participants. According to the effect size reported in Mazor, Friston and Fleming (2020), the power of replicating a significant differences in quadratic confidence effect between detection and discrimination was 0.228 based on the current sample. Therefore, with this size of sample, it was theoretically not likely to detect a significant impact and a more complete sample could make differences to results. Since previous finding of a unique quadratic effect of confidence in detection arise from exploratory analysis that was not theory driven, and therefore, lied on a less solid ground, another possibility is that such effect was a noise in the data collected and was specific to that particular sample. In this case, the quadratic effect observed previously was not a generalizable pattern and thus could not contribute to evidence supporting differences between detection and discrimination. However, it is difficult to draw conclusions from incomplete sample, thus we leave this as an open question to be answered by future research.

# 6. Ethics of animal research

An important consideration we draw readers attention to is the aspect concerning ethics of the methodology used in cited animal studies (Katharina Merten and Andreas Nieder (2012), Miyamoto et al. (2018) Okuyama, Kuki, and Mushiake (2015), Ramirez-Cardenas, Moskaleva, and Nieder (2016) and Field, Biagi, and Inman (2020)). These studies often involve training animals (e.g. rhesus monkeys) to perform tasks to serve the purpose of research. In many cases, the training of these animals involves restraining animals from satisfying their physical needs (e.g. drinking water, connection with conspecifics), and introducing physical harm to their bodies. Although these existing studies do provide different perspectives from human research, here we call researchers to reconsider the ethics of such methods and develop alternative training systems that for instance, emphasis more on rewards without harming basic needs of animals.

# 7. Reference

Adler, William T., and Wei Ji Ma. 2018. “Comparing Bayesian and Non-Bayesian Accounts of Human Confidence Reports.” *PLoS Computational Biology* 14 (11): e1006572. <https://doi.org/10.1371/journal.pcbi.1006572>.

Ariel Ezylberberg, Ariel Ezylberberg, Ariel Ezylberberg, Pablo Ebarttfeld, and Mariano Esigman. 2012. “The Construction of Confidence in a Perceptual Decision.” *Frontiers in Integrative Neuroscience* 6: 79. <https://doi.org/10.3389/fnint.2012.00079>.

Bennur, Sharath, and Joshua I. Gold. 2011. “Distinct Representations of a Perceptual Decision and the Associated Oculomotor Plan in the Monkey Lateral Intraparietal Area.” *The Journal of Neuroscience* 31 (3): 913–21. <https://doi.org/10.1523/JNEUROSCI.4417-10.2011>.

Blakemore, Sarah-Jayne, Daniel M. Wolpert, and Christopher D. Frith. 2002. “Abnormalities in the Awareness of Action.” *Trends in Cognitive Sciences* 6 (6): 237–42. <https://doi.org/10.1016/s1364-6613(02)01907-1>.

Boorman, Erie D., Timothy E. J. Behrens, Mark W. Woolrich, and Matthew F. S. Rushworth. 2009. “How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action.” *Neuron* 62 (5): 733–43. <https://doi.org/10.1016/j.neuron.2009.05.014>.

Davidson, Matthew J., Will Mithen, Hinze Hogendoorn, Jeroen J. A. van Boxtel, and Naotsugu Tsuchiya. 2020. “A Neural Representation of Invisibility: When Stimulus-Specific Neural Activity Negatively Correlates with Conscious Experience.” *bioRxiv*, April, 2020.04.20.051334. <https://doi.org/10.1101/2020.04.20.051334>.

Decety, Jean, and Claus Lamm. 2007. “The Role of the Right Temporoparietal Junction in Social Interaction: How Low-Level Computational Processes Contribute to Meta-Cognition.” *The Neuroscientist (Baltimore, Md.)* 13 (6): 580–93. <https://doi.org/10.1177/1073858407304654>.

Denison, Rachel N., William T. Adler, Marisa Carrasco, and Wei Ji Ma. 2018. “Humans Incorporate Attention-Dependent Uncertainty into Perceptual Decisions and Confidence.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (43): 11090–5. <https://doi.org/10.1073/pnas.1717720115>.

Field, David T., Nicol‘o Biagi, and Laura A. Inman. 2020. “The Role of the Ventral Intraparietal Area (VIP/pVIP) in the Perception of Object-Motion and Self-Motion.” *NeuroImage* 213 (June): 116679. <https://doi.org/10.1016/j.neuroimage.2020.116679>.

Fleming, Stephen M., and Raymond J. Dolan. 2012. “The Neural Basis of Metacognitive Ability.” *Philosophical Transactions of the Royal Society B* 367 (1594): 1338–49. <https://doi.org/10.1098/rstb.2011.0417>.

Fleming, Stephen M, Josefien Huijgen, and Raymond J. Dolan. 2012. “Prefrontal Contributions to Metacognition in Perceptual Decision Making.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 32 (18): 6117–25. <https://doi.org/10.1523/JNEUROSCI.6489-11.2012>.

Fleming, Stephen M., and Hakwan C. Lau. 2014. “How to Measure Metacognition.” *Frontiers in Human Neuroscience* 8: 443. <https://doi.org/10.3389/fnhum.2014.00443>.

Grefkes, Christian, and Gereon R. Fink. 2005. “REVIEW: The Functional Organization of the Intraparietal Sulcus in Humans and Monkeys.” *Journal of Anatomy* 207 (1): 3–17. <https://doi.org/10.1111/j.1469-7580.2005.00426.x>.

Grill-Spector, Kalanit, and Nancy Kanwisher. 2005. “Visual Recognition: As Soon as You Know It Is There, You Know What It Is.” *Psychological Science* 16 (2): 152–60.

Grimaldi, Piercesare, Hakwan Lau, and Michele A. Basso. 2015. “There Are Things That We Know That We Know, and There Are Things That We Do Not Know We Do Not Know: Confidence in Decision-Making.” *Neuroscience and Biobehavioral Reviews* 55: 88–97. <https://doi.org/10.1016/j.neubiorev.2015.04.006>.

Kanai, Ryota, Vincent Walsh, and Chia-Huei Tseng. 2010. “Subjective Discriminability of Invisibility: A Framework for Distinguishing Perceptual and Attentional Failures of Awareness.” *Consciousness and Cognition* 19 (4): 1045–57. <https://doi.org/10.1016/j.concog.2010.06.003>.

Katharina Merten, and Andreas Nieder. 2012. “Active Encoding of Decisions About Stimulus Absence in Primate Prefrontal Cortex Neurons.” *Proceedings of the National Academy of Sciences* 109 (16): 6289–94. <https://doi.org/10.1073/pnas.1121084109>.

King, J.-R., and S. Dehaene. 2014. “A Model of Subjective Report and Objective Discrimination as Categorical Decisions in a Vast Representational Space.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 369 (1641): 20130204. <https://doi.org/10.1098/rstb.2013.0204>.

Mack, Michael L., and Thomas J. Palmeri. 2010. “Decoupling Object Detection and Categorization.” *Journal of Experimental Psychology: Human Perception and Performance* 36 (5): 1067–79. <https://doi.org/10.1037/a0020254>.

Maniscalco, Brian, and Hakwan Lau. 2011. “On a Distinction Between Detection and Discrimination: Metacognitive Advantage for Signal over Noise.” *Journal of Vision* 11 (11): 163–63. <https://doi.org/10.1167/11.11.163>.

Mazor, Matan, Karl J Friston, and Stephen M Fleming. 2020. “Distinct Neural Contributions to Metacognition for Detecting, but Not Discriminating Visual Stimuli.” Edited by Thorsten Kahnt, Joshua I Gold, and Michael Graziano. *eLife* 9 (April): e53900. <https://doi.org/10.7554/eLife.53900>.

Meuwese, Julia D. I., Anouk M. Van Loon, Victor A. F. Lamme, and Johannes J. Fahrenfort. 2014. “The Subjective Experience of Object Recognition: Comparing Metacognition for Object Detection and Object Categorization.” *Attention, Perception & Psychophysics* 76 (4): 1057–68. <https://doi.org/10.3758/s13414-014-0643-1>.

Miyamoto, Kentaro, Rieko Setsuie, Takahiro Osada, and Yasushi Miyashita. 2018. “Reversible Silencing of the Frontopolar Cortex Selectively Impairs Metacognitive Judgment on Non-Experience in Primates.” *Neuron* 97 (4): 980–989.e6. <https://doi.org/10.1016/j.neuron.2017.12.040>.

Morales, J., H. Lau, and S. M. Fleming. 2018. “Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex.” *Journal of Neuroscience , 38 (14) Pp. 3534-3546. (2018)*.

Okuyama, Sumito, Toshinobu Kuki, and Hajime Mushiake. 2015. “Representation of the Numerosity ‘Zero’ in the Parietal Cortex of the Monkey.” *Scientific Reports* 5 (1): 10059. <https://doi.org/10.1038/srep10059>.

Ramirez-Cardenas, Araceli, Maria Moskaleva, and Andreas Nieder. 2016. “Neuronal Representation of Numerosity Zero in the Primate Parieto-Frontal Number Network.” *Current Biology* 26 (10): 1285–94. <https://doi.org/10.1016/j.cub.2016.03.052>.

Rinaldi, Luca, and Luisa Girelli. 2016. “A Place for Zero in the Brain.” *Trends in Cognitive Sciences* 20 (8): 563–64. <https://doi.org/10.1016/j.tics.2016.06.006>.

Victor De Lafuente, and Ranulfo Romo. 2005. “Neuronal Correlates of Subjective Sensory Experience.” *Nature Neuroscience* 8 (12): 1698–1703. <https://doi.org/10.1038/nn1587>.

Wason, P. C. 1959. “The Processing of Positive and Negative Information:” *Quarterly Journal of Experimental Psychology*. <https://doi.org/10.1080/17470215908416296>.

Wickens, Thomas D. 2002. *Elementary Signal Detection Theory / Thomas D. Wickens.* Oxford Scholarship Online. Oxford: University Press.