# Confidence in Detection and Discrimination: an fMRI Study

Matan Mazor, Karl J. Friston and Stephen M. Fleming

## Objective

the current study aims to compare the brain processes that govern perceptual discrimination and detection, and the neural mechanisms that allow for metacognitive evaluations of these processes.

A fundamental property that distinguishes detection from discrimination tasks is the asymmetry in the availability of evidence for 'yes' and for 'no' responses. While discrimination requires a comparison between the relative evidence for different options, in a detection setting evidence can only be available for the presence of a stimulus and not for its absence. Conceptually, this means that confidence in the absence of a stimulus cannot rely on the magnitude of evidence for its absence and may rely instead on counterfactual reasoning regarding the likelihood of the stimulus to be detected had it been presented. Behaviorally, this difference is reflected in general lower confidence and in a weaker association between objective accuracy and subjective confidence for 'no' responses (Kanai, Walsh, & Tseng, 2010; Meuwese, van Loon, Lamme, & Fahrenfort, 2014) in detection but also in detection-like tasks (such as recognition memory; Higham, Perfect, & Bruno, 2009).

It is still unknown what are the brain mechanisms that give rise to these behavioral differences. While previous studies compared structural and functional correlates for metacognitive sensitivity ratings across domains (Mccurdy, Maniscalco, Metcalfe, & De Lange, 2013; Morales, Lau, & Fleming, 2018), great care was taken to equate task requirements and avoid the asymmetry inherent to true detection tasks. For example, instead of asking participants to perform 'old'/'new' recognition judgments, participants were asked to answer which of two presented stimuli is old. Similarly, instead of asking participants whether they detected a signal or not, a 2 Interval Forced Choice (2IFC) approach is often preferred, where participants are asked to report whether the signal was presented in the first or the second interval.

Here we wish to compare detection and discrimination within the same low-level perceptual task, while controlling for task performance. The objectives of this study are:

1. Replicate previous findings of inter-subject correlations of structural and functional properties of the lateral prefrontal cortex (lPFC) with metacognitive sensitivity in discrimination (Fleming, Weil, Nagy, Dolan, & Rees, 2009; Mccurdy et al., n.d.).
2. Find inter-subject functional and structural correlates of metacognitive sensitivity in detection. Specifically, we will be interested to see if any dissociations can be found between brain structures that predict metacognitive sensitivity in detection and in discrimination.
3. Replicate previous findings of general confidence signal in ventromedial prefrontal cortex (De Martino, Fleming, Garrett, & Dolan, 2013; Morales et al., 2018).
4. Test for an interaction between confidence level and task (detection/discrimination) in BOLD response, specifically in the prefrontal cortex.
5. Within detection, test for an interaction between confidence level and response (yes/no) in BOLD response, specifically in the prefrontal cortex and in regions that have previously been associated with counterfactual reasoning (Boorman, Behrens, & Rushworth, 2011; Neubert, Mars, Thomas, Sallet, & Rushworth, 2014).
6. Test for an interaction between task and within-subject fluctuations in metacognitive sensitivity. Specifically, test the hypothesis that the frontopolar cortex is more associated with fluctuations in metacognitive sensitivity in detection trials when the subject reported the target to be missing (Miyamoto, Setsuie, Osada, & Miyashita, 2018).

## Design

We will test 30 healthy subjects in a 3 Tesla MRI scanner in the Welcome Centre for Human Neuroimaging, Institute of Neurology, University College London.

Participants will undergo 6 interleaved detection and discrimination scanner runs, each consisting of 72 trials. After a temporally jittered rest period of 500-4000 milliseconds, the trial will begin with a cue fixation cross (500 milliseconds), followed by a random dot kinematogram (RDK; 700 milliseconds). In discrimination blocks, participants will be asked to report on a continuous scale the direction of movement and their confidence in their decision. On each discrimination trial, a subset of the dots will be moving coherently to one of two opposite directions with equal probability (up/down or right/left, counterbalanced between participants). In detection blocks, on half of the trials all dots will be moving randomly, and on the other half a subset of the dots will be moving to one of two opposite direction (the same directions as in the discrimination task) with equal probabilities. Participants will be asked to report on a perpendicular continuous scale the degree to which they believe a subset of the dots moved coherently or not. The number of coherently moving dots will be fitted independently for the detection and discrimination tasks, to match performance on the two tasks as much as possible. No feedback will be delivered to subjects about their performance.

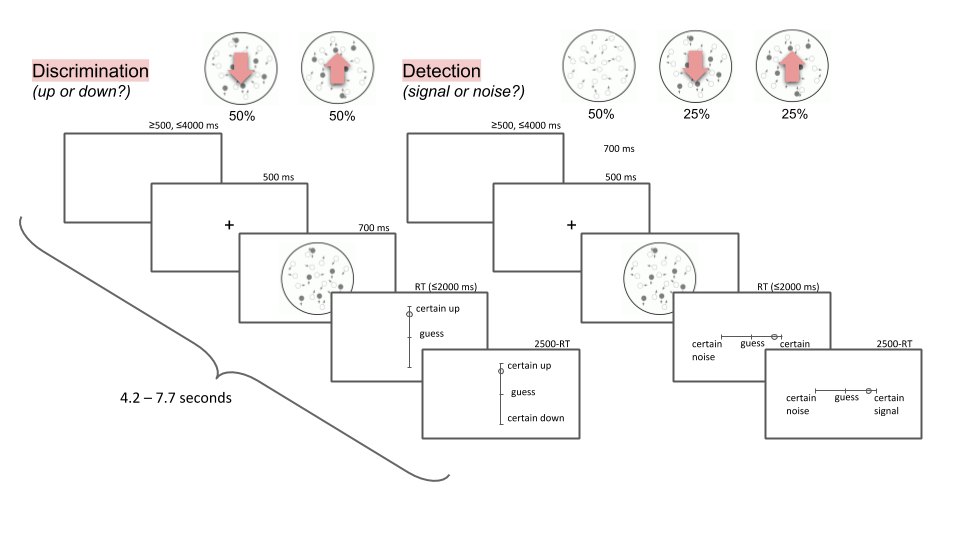


Figure 1: Experimental design for discrimination and for detection trials.

## Analysis

### fMRI data preprocessing

Data preprocessing will follow the procedure described in Morales and colleagues (2018):

*Imaging analysis was performed using SPM12 (Statistical Parametric Mapping; www.fil.ion.ucl.ac.uk/spm). The first five volumes of each run were discarded to allow for T1 stabilization. Functional images were realigned and unwarped using local field maps (Andersson et al., 2001) and then slice-time corrected (Sladky et al., 2011). Each participant’s structural image was segmented into gray matter, white matter, CSF, bone, soft tissue, and air/background images using a nonlinear deformation field to map it onto template tissue probability maps (Ashburner and Friston, 2005). This mapping was applied to both structural and functional images to create normalized images to Montreal Neurological Institute (MNI) space. Normalized images were spatially smoothed using a Gaussian kernel (8 mm FWHM). We set a within-run 1 mm rotation and 4 mm affine motion cutoff criterion, which led to the exclusion of 4 subjects, leaving a total of 24 subjects whose functional and behavioral data were fully analyzed.*

### Univariate Analysis

The design matrix for the univariate GLM analysis will consist of the following regressors of interest:

1. *Signal*: Detection trials in which the subject reported detecting a signal. A boxcar regressor with nonzero entries at the 2500 millisecond response interval will be convolved the canonical hemodynamic response function (HRF).
2. *Noise*: Detection trials in which the subject reported not detecting a signal.
3. *S1*: Discrimination trials in which the subject reported a rightward/upward moving signal.
4. *S2*: Discrimination trials in which the subject reported a leftward/downward moving signal.
5. *Conf*: Confidence ratings will be z-scored within subject and across tasks and responses. The normalized ratings will be fed into the design matrix as a parametric modulator.
6. *MC:* A dynamic estimate of the metacognitive efficiency, tracked separately for detection signal, detection noise and discrimination responses. The M-ratio estimates will be fed into the design matrix as a parametric modulator.

In addition, interaction regressors will be included to test for interactions between condition and confidence and condition and metacognitive sensitivity, and nuisance constant and motion regressors will be included to account for movement-related activations.

We will apply the following contrasts to the voxel-wise beta estimates:

|  |  |  |
| --- | --- | --- |
| Contrast | Interpretation | Pre-specified regions of interest |
| 1. Conf | Brain regions showing linear modulation of reported confidence on BOLD signal during the response stage. | vmPFC (MNI [12, 47, −11]; De Martino et al., 2012)  rlPFC (MNI [39, 41, 16]; Negative effect; ibid.)  Ventral striatum (MNI [3, 11, 7], [9, 9, -3], [-9, 9,-3] Morales et al., 2018; Hebart et al., 2016);  Right Middle frontal gyrus (negative effect; MNI [45, 26, 20], ibid.)  Pre-SMA, BA8 (negative effect; MNI [0, 14, 50], ibid.) |
| 1. (SignalXConf + NoiseXconf)- (S1XConf + S2XConf) | Brain regions showing differential modulation for confidence as a function of task. | We will use the whole-brain map of contrast number 1, thresholded at alpha=0.01 as a functional localizer for this contrast, and a small-volume correction will be applied to positive results within this mask. In addition, BA10 and BA46 will be defined anatomically as *a priori* regions of interests. |
| 1. (Signal+Noise)- (S1+S2) | Brain regions showing a main effect for task. | BA10 and BA46 will be defined anatomically as *a priori* regions of interests. |
| 1. Noise-Signal | Brain regions showing a main response effect for the detection task. | We will use the whole brain map of contrast number 3 thresholded at alpha=0.01 as a functional localizer for this contrast, and a small-volume correction will be applied to positive results within this mask. In addition, BA10 and BA46 will be defined anatomically as *a priori* regions of interests. |
| 1. NoiseXConf – SignalXConf | Brain regions showing a differential modulation of confidence as a function of response, within the detection task. | We will use the whole brain map of contrast number 2 thresholded at alpha=0.01 as a functional localizer for this contrast, and a small-volume correction will be applied to positive results within this mask. In addition, BA10 and BA46 will be defined anatomically as *a priori* regions of interests. |
| 1. MC | Brain regions showing a modulation of metacognitive efficiency on the BOLD signal during the response stage. | BA10, BA46 and the precuneus will be defined anatomically as *a priori* regions of interests for this contrast (Fleming et al., 2009). |
| 1. (SignalXMC + NoiseXMC)- (S1XMC + S2XMC) | Brain regions showing differential modulation for metacognitive sensitivity as a function of task. | We will use the whole-brain map of contrast number 6, thresholded at alpha=0.01 as a functional localizer for this contrast, and a small-volume correction will be applied to positive results within this mask. In addition, BA10 and BA46 will be defined anatomically as *a priori* regions of interests. |
| 1. NoiseXMC-SignalXMC | Brain regions showing differential modulation for metacognitive sensitivity as a function of response type within detection. | We will use the whole-brain map of contrast number 7, thresholded at alpha=0.01 as a functional localizer for this contrast, and a small-volume correction will be applied to positive results within this mask. In addition, BA10 will be defined anatomically as an *a priori* region of interests (Miyamoto et al., 2018). |

### Between-subject correlations

We will use voxel-based morphometry (VBM) to find brain structures that are associated with metacognitive efficiency for detection and for discrimination separately. Metacognitive efficiency will be defined as meta-d'/d' (Maniscalco & Lau, 2012), and will be correlated against gray-matter volume as measured with T1-weighted anatomical images (Fleming et al., 2009). For discrimination metacognition, BA10 ([24,65, 18], [-20,53,12], [33,50,9], [-12,54,16]), precuneus ([6,-57,18]) and BA46 ([36,39,21]) will be defined as *a priori* regions of interest based on Fleming and colleagues (2009) and McCurdy and colleagues (2013). BA10 will be defined anatomically as an *a priori* region of interest for metacognition in detection blocks.

### Multivariate analysis

Multi-voxel pattern analysis (Norman, Polyn, Detre, & Haxby, 2006) will be used to test for consistent spatial patterns in the fMRI data. We will follow the procedure described in Morales and colleagues (2018).

|  |  |  |  |
| --- | --- | --- | --- |
| Train | Test | Interpretation | Pre-specified regions of interest |
| High metacognitive accuracy vs. Low metacognitive accuracy | High metacognitive accuracy vs. Low metacognitive accuracy | Spatially multivariate signal predicting metacognitive performance. | Fleming, McCurdy, Baird |
| Within discrimination: high metacognitive accuracy vs. Low metacognitive accuracy | Within discrimination: high metacognitive accuracy vs. low metacognitive accuracy | Spatially multivariate signal predicting metacognitive performance in discrimination. | Fleming, McCurdy,Baird |
| Within detection: high metacognitive accuracy vs. Low metacognitive accuracy | Within detection: high metacognitive accuracy vs. Low metacognitive accuracy | Spatially multivariate signal predicting metacognitive performance in detection. | Fleming, McCurdy,Baird, Rushworth |
| Within discrimination: high metacognitive accuracy vs. Low metacognitive accuracy | Within detection: high metacognitive accuracy vs. Low metacognitive accuracy | Spatially multivariate signal predicting metacognitive performance in detection and discrimination. |  |
| Within detection Noise responses: high metacognitive accuracy vs. Low metacognitive accuracy | Within detection Noise responses: high metacognitive accuracy vs. Low metacognitive accuracy | Spatially multivariate signal predicting metacognitive performance in the representation of absence. |  |

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual Choice and Learning in a Neural Network Centered on Human Lateral Frontopolar Cortex. *PLoS Biology*, *9*(6), e1001093. https://doi.org/10.1371/journal.pbio.1001093

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110. https://doi.org/10.1038/nn.3279

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2009). Relating Introspective Accuracy to Individual Differences in Brain Structure. *Science*, *324*(5928), 759–764. https://doi.org/10.1126/science.1169405

Hebart, M. N., Schriever, Y., Donner, T. H., Haynes, J.-D., & Hebart, M. (n.d.). The Relationship between Perceptual Decision Variables and Confidence in the Human Brain. https://doi.org/10.1093/cercor/bhu181

Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 57–80. https://doi.org/10.1037/a0013865

Kanai, R., Walsh, V., & Tseng, C. (2010). Subjective discriminability of invisibility: A framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition*, *19*(4), 1045–1057. https://doi.org/10.1016/J.CONCOG.2010.06.003

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*, 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Mccurdy, L. Y., Maniscalco, B., Metcalfe, J., & De Lange, F. P. (2013). Anatomical Coupling between Distinct Metacognitive Systems for Memory and Visual Perception Decoded Neurofeedback Project for development of diagnostic and therapeutic system for mental disorders View project Temporal dynamics of visual imagery View proje. https://doi.org/10.1523/JNEUROSCI.1890-12

Mccurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., De Lange, F. P., & Lau, H. (n.d.). Behavioral/Cognitive Anatomical Coupling between Distinct Metacognitive Systems for Memory and Visual Perception. https://doi.org/10.1523/JNEUROSCI.1890-12.2013

Meuwese, J. D. I., van Loon, A. M., Lamme, V. A. F., & Fahrenfort, J. J. (2014). The subjective experience of object recognition: comparing metacognition for object detection and object categorization. *Attention, Perception, & Psychophysics*, *76*(4), 1057–1068. https://doi.org/10.3758/s13414-014-0643-1

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*(4), 980–989.e6. https://doi.org/10.1016/j.neuron.2017.12.040

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *38*(14), 3534–3546. https://doi.org/10.1523/JNEUROSCI.2360-17.2018

Neubert, F.-X., Mars, R. B., Thomas, A. G., Sallet, J., & Rushworth, M. F. S. (2014). Comparison of Human Ventral Frontal Cortex Areas for Cognitive Control and Language with Areas in Monkey Frontal Cortex. *Neuron*, *81*(3), 700–713. https://doi.org/10.1016/J.NEURON.2013.11.012

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005