

## RESEARCH ARTICLE

WILEY

# Detecting spontaneous deception in the brain

Yen-Ju Feng<sup>1</sup>  | Shao-Min Hung<sup>2</sup> | Po-Jang Hsieh<sup>1</sup>

<sup>1</sup>Department of Psychology, National Taiwan University, Taipei, Taiwan

<sup>2</sup>Biology and Biological Engineering, California Institute of Technology, Pasadena, California, USA

## Correspondence

Yen-Ju Feng and Po-Jang Hsieh, Department of Psychology, National Taiwan University, Taipei 10617, Taiwan.

Email: [yenju0115@gmail.com](mailto:yenju0115@gmail.com) (Y.-J. F.);

[hsiehpj@ntu.edu.tw](mailto:hsiehpj@ntu.edu.tw) (P.-J. H.)

## Funding information

Ministry of Education (MOE), Yushan Young Scholar Program, Grant/Award Number: NTU-109V0202; Ministry of Science and Technology, Taiwan, Grant/Award Number: 109-2410-H-002-004-MY3; National Taiwan University, Grant/Award Number: 110L9A00701

## Abstract

Deception detection can be of great value during the juristic investigation. Although the neural signatures of deception have been widely documented, most prior studies were biased by difficulty levels. That is, deceptive behavior typically required more effort, making deception detection possibly effort detection. Furthermore, no study has examined the generalizability across instructed and spontaneous responses and across participants. To explore these issues, we used a dual-task paradigm, where the difficulty level was balanced between truth-telling and lying, and the instructed and spontaneous truth-telling and lying were collected independently. Using Multivoxel pattern analysis, we were able to decode truth-telling versus lying with a balanced difficulty level. Results showed that the angular gyrus (AG), inferior frontal gyrus (IFG), and postcentral gyrus could differentiate lying from truth-telling. Critically, linear classifiers trained to distinguish instructed truthful and deceptive responses could correctly differentiate spontaneous truthful and deceptive responses in AG and IFG with above-chance accuracy. In addition, with a leave-one-participant-out analysis, multivoxel neural patterns from AG could classify if the left-out participant was lying or not in a trial. These results indicate the commonality of neural responses subserved instructed and spontaneous deceptive behavior as well as the feasibility of cross-participant deception validation.

## KEYWORDS

angular gyrus, deception detection, inferior frontal gyrus, lying, multivoxel pattern analysis

## 1 | INTRODUCTION

Deception is a complex cognitive activity that usually occurs when people attempt to convince others to accept incorrect beliefs. Deception could come in many different forms, such as outright lies, exaggerations, omissions, and subtle lies (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996; Vrij, Edward, Roberts, & Bull, 2000), and it affects various aspects of life, including politics, marketing, and personal relationships. Given that human society is trust-based, deception may hamper communication and destroy the relationship, leading

to negative consequences such as loss of property and sometimes even life. Therefore, deception detection has attracted researchers' attention for decades and human society has long sought scientific methods for detecting deceptive behaviors.

Polygraph is one of the pioneering techniques in measuring peripheral responses such as blood pressure, pulse rate, respiration, and electrodermal responses (Lykken, 1981; Saxe, Dougherty, & Cross, 1985). However, the use of polygraphs could be problematic, since measured physiological responses tend to correlate closely with emotional reactions like anger, fear, and anxiety. These unspecified

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

responses may not necessarily be associated with lying behavior under real-world interrogations, which can result in false positives in deception detection (Steinbrook, 1992).

Over time, the traditional measures have been supplemented by electroencephalographic (EEG) and functional magnetic resonance imaging (fMRI). Prior fMRI studies on deception fall into two categories. On the one hand, researchers have been focusing on detecting deception in each and every individual. For instance, indexed by whole-brain resting-state functional connectivities (RSFC), a prior study demonstrated that brain networks involving executive controlling (dorsolateral prefrontal cortex, middle frontal cortex, and orbitofrontal cortex), social and mentalizing (the temporal lobe, temporo-parietal junction, and inferior parietal lobule), and reward (putamen and thalamus) could predict participants' deceptive behaviors in an independent task (Tang et al., 2018). Also, it has been demonstrated that within a single individual, lying can be differentiated from truth-telling with an accuracy of 78% (Langleben et al., 2005). All aforementioned works demonstrated the values of the utilization of fMRI in distinguishing deceptive from truthful responses.

On the other hand, scientists attempted to detect deceptive behaviors by exploring neural mechanisms underlying deception at the group level. Relevant results showed that in comparison with truth-telling, lying elicited higher activations in the ventral medial prefrontal cortex (VMPFC), the prefrontal cortex (PFC), the prefrontal motor cortex (PMC), and the orbital frontal cortex (OFC) (Giorgio Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003; Kozel et al., 2005; Kozel, Padgett, & George, 2004; Phan et al., 2005; Spence, 2004; Spence et al., 2001). Based on these findings, it appears that lying may necessitate additional executive functions such as working memory, inhibition, and task switching (Christ, Van Essen, Watson, Brubaker, & McDermott, 2009). However, with identical contrast (truthful vs. deceptive responses), apart from regions related to high-level cognitive functions, other studies also showed that regions related to perception such as cuneus, precuneus, and cerebellum had greater activations during lying (Lee et al., 2002, 2005; Nuñez, Casey, Egner, Hare, & Hirsch, 2005). These discrepancies may result from the heterogeneity in experimental tasks and protocols (e.g., playing cards, autobiographical knowledge, rehearsal scenario, or mock-crime scenario).

Even though tasks and protocols differed from one another, they could be roughly classified into two groups: the Control Question Task (CQT) and Conceal Information Task (CIT), also known as Guilty Knowledge Test (GKT). These two techniques are different in terms of their basic assumptions. CQT assumes that deceptive responses/activities can be identified by directly comparing truth-telling against lying. Thus, CQT usually employs forced-choice methods, during which participants were required to give "Yes" or "No" answers to questions related to personal information or general knowledge involving truths or deceptive responses (e.g., The earth is round). As for CIT/GKY, the underlying assumption is that lying behaviors/responses are only observable in concealers who hide the truth. Thus, during CIT/GKT, both concealers and innocents will be asked to

answer relevant, irrelevant, and neutral questions (usually served as distractors) and deceptive responses can be revealed by comparing the reaction differences between these three types of questions between two groups of participants (concealers and innocents).

Based on CQT or CIT the neural substrates of the deceptive responses have been investigated to some extent, yet several issues remained unresolved. First, when searching neural activities for deceptive responses, most prior studies directly compared the differences between truth-telling versus lying conditions. However, the cognitive burden for lying is heavier than that for truth-telling: lying may require more engagement of working memory, exogenous, endogenous attention, and cognitive control (Sánchez, Masip, & Gómez-Ariza, 2020; Vartanian et al., 2013). Usually, when people attempt to lie to a question, they need to intentionally suppress the default truthful response and then make a reverse reaction, which may therefore increase the difficulty of lying. Although prior studies attempted to control for the cognitive resources between truth-telling and lying (Spence, Kaylor-Hughes, Farrow, & Wilkinson, 2008), the effect of different difficulty levels remained to be tested.

Second, few studies ever investigated the generalizability of lie detection across instructed deception and spontaneous deception albeit the neural mechanism underlying spontaneous lying have been examined (Yin, Reuter, & Weber, 2016; Zhang, Liu, Pelowski, & Yu, 2017). In most of the prior fMRI studies, participants were instructed to tell the truth or lie in the laboratory settings which lacked a key ingredient of real-life deception, namely spontaneous behaviors. Therefore, whether the results of prior studies can be generalized to deception in real life remain unclear.

Lastly, one critical hurdle for a practical deception detector has been to overcome individual differences. Although prior brain imaging studies with cross-validation analysis have successfully utilized brain activity patterns gathered from a group to predict the left-out participants (Mason, Just, Keller, & Carpenter, 2003; Wang, Hutchinson, & Mitchell, 2004) the feasibility of this approach remains to be tested in deception detection.

In sum, to address the aforementioned questions, there are three specific aims for the current study. As a first step, a dual-task paradigm, consisting of Task1 and Task2, was created based on CQT, which is simpler and easier to control in comparison with CIT. To resolve the concern with regard to difficulty when attesting to deceptive behaviors, truth conditions in Task1 will be further divided into two groups: truth-telling-easy (T1Te) and truth-telling-difficult (T1Td) (details provided in the Methods). With this control, we aimed to explore neural signatures exclusive for deceptive responses by comparing lying versus truth-telling-difficult conditions and establish a physiological database for the followed-up analysis.

Second, to resolve the concern with regard to the generalizability of truthful and deceptive acts, we aimed to test how well the instructed responses predict participants' spontaneous behaviors in another task. More specifically, within-participant cross-task classifier accuracies will be attested during which neural patterns from Task1 will be recorded and served as a baseline to predict neural patterns from Task2.

Third, we aimed to explore the feasibility of training classifiers to distinguish truthful and deceptive responses *across* participants. Cross-participant cross-task classifier accuracies will be examined by first establishing a database of instructed truth and lie responses from a group of participants, and then used to predict left-out individuals' spontaneous truth and lying acts.

## 2 | METHODS

### 2.1 | Participants

Twenty-five healthy volunteers (male = 12) were recruited from a local community to participate in the study. Their ages ranged from 20 to 32 years (Mean age = 25.33; SD = 3.38). Before the main experiment, the process and instructions were clearly explained to the participants who were then asked to sign an informed consent form. All recruited participants had a normal or corrected vision and they were screened for any history of neurological or mental disorders. They were also assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) as being right-handed. After the study, all participants were reimbursed with USD \$25 for participating in a 60-min scanning session. This study was non-invasive, performed in accordance with the ethical standards of the Declaration of Helsinki, and approved by the Institutional Review Board of the National Taiwan University.

To ensure the quality of the following image analyses, there are two exclusive criteria. First, at the run level, if the success rate of following the task instructions (to tell the truth/lie) in Task1 is lower than 75% the whole run will be excluded. Second, at the individual level, if the included number of runs is less than two then images from that particular participant will be fully excluded during the analysis. Following these criteria, images from four out of 25 participants were fully

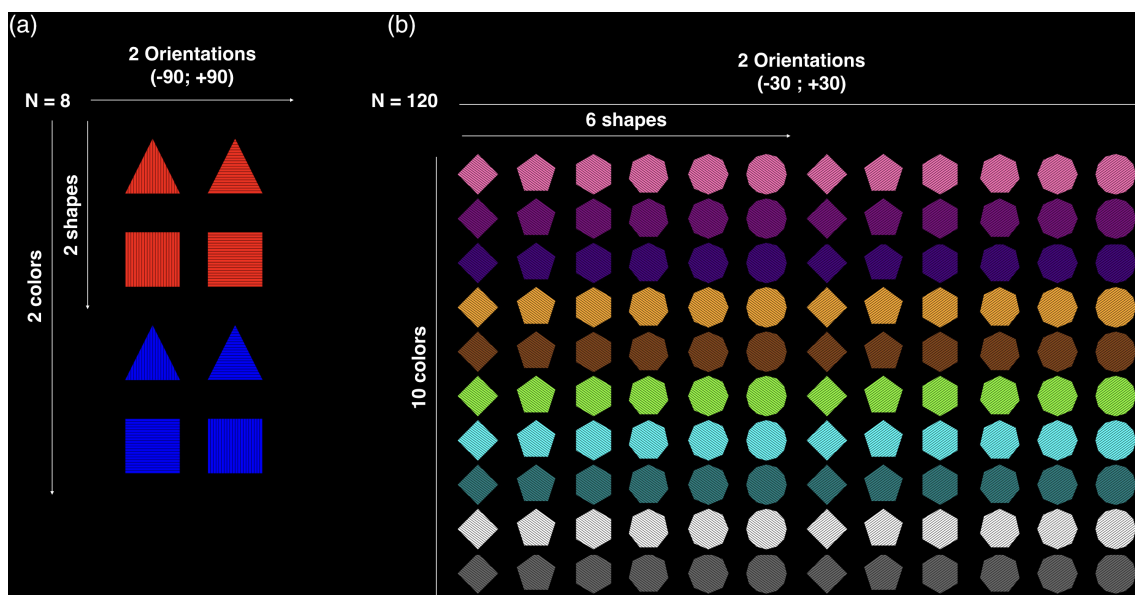
excluded. All the included participants had more than two runs and the average success rate in following the given task instructions (to tell the truth/lie) was higher than 75% (Figure S2). In this study, the number of included participants is comparable with prior fMRI studies testing on deceptions (Bhatt et al., 2009; Ganis, Rosenfeld, Meixner, Kievit, & Schendan, 2011; Nuñez et al., 2005; Yin & Weber, 2019; Zheltyakova, Kireev, Korotkov, & Medvedev, 2020).

### 2.2 | Stimuli

All experimental stimuli were created with Matlab. In Task1, eight types ( $2 \times 2 \times 2$ ) of visual stimuli were included: two shapes (triangle, square) overlaid with different orientations of lines (horizontal, vertical) and in different colors (blue, red) (Figure 1a). As for Task2, a material database, comprising 120 visual stimuli, was created from overlaying six shapes (diamond, pentagon, hexagon, heptagon, octagon, dodecagon) with either left- or right-tilted lines (vertical lines tilted by  $30^\circ$  in a clockwise or counterclockwise direction) and each of them had 10 different colors (Figure 1b). Notably, Task1 stimuli and Task2 stimuli were different in terms of shape, orientation of imbedded lines and color. Statements also varied depending on conditions. During the scanning, all the stimuli, as well as statements, were rear-projected to the center of the visual field using a video projector viewed through a head coil-mounted mirror (roughly 10–15 cm away from the participants).

### 2.3 | Task design

A dual-task paradigm, including two independent tasks namely Task1 and Task2, was employed in the current study. Task1 consisted of



**FIGURE 1** (a) An illustration of used visual stimuli in Task1. (b) An illustration of the material database created for Task2

three conditions: truth-telling-easy (T1Te), truth-telling-difficult (T1Td), and lying (T1L). Task2 consisted of two conditions: truth-telling (T2T) and lying (T2L). Notably, instructed lying was involved in Task1 while spontaneous lying was in Task2. In each run, the stimuli were presented in 4 blocks (3 blocks of Task1 and 1 block of Task2), each containing 8 trials. Each block began with a probe specifying the upcoming block as Task1 or Task2 and the duration for this period (probe duration) varied between 8 to 12 s.

Both Task1 and Task2 followed a similar trial procedure. Each trial consisted of four phases: the presentation of (1) instruction, (2) visual stimuli, (3) statements, and (4) feedbacks. Each phase lasted for 2, 3, 2, and 1 s, respectively, leading to 8 s per trial (Figures 2b,c). However, there were four main differences between Task1 and Task2, one for each phase. First, in Task1 participants were explicitly instructed to tell the truth or to lie with regard to the presented visual stimuli whereas in Task2, participants were allowed to make a response (either tell the truth or lie) of their free will. Second, in Task1, each type of stimuli created for Task1 was presented once. As for Task2, 4 out of 8 trials contained items randomly selected from 8 types of stimuli used in Task1, and the other 4 trials contained items randomly selected from the material database created only for Task2. Third, to control for difficulties between truth-telling and lying, simple and complex statements were created: simple statements were applied in

both truth-telling-easy (T1Te) and lying (T1L), whereas complex statements were used in truth-telling-difficult (T1Td). All the simple statements directly described one of the visual features characterized by presented visual stimuli whereas all the complex statements described the presented visual stimuli in an indirect manner (Table 1). In contrast, the statement in Task2 was always identical. Lastly, in Task1, feedback with regard to whether participants successfully sticking to a prior instruction was presented in each trial (“Correct”/“Incorrect”). Whereas in Task2, either caught information (“I Got You”/“Sorry I was wrong”) or blank was presented during the feedback phase. Notably, during Task2, participants were randomly caught two times.

Additionally, due to the task statement (i.e., “This figure has been presented in Task1”) used in Task2, the first block in the first run was always Task1, leading to the pseudo-randomized order of the task (Figure 2a). It is, therefore, important to note that for each run, Task1 always appeared before Task2 despite the fact that it could have been any condition from Task1.

## 2.4 | Task instruction

As per each of the phrases mentioned above, participants were instructed as below.



**FIGURE 2** An illustration of the run and trial procedure. (a) Each run consisted of four blocks, three designed for Task1 and one for Task2 (in a pseudo-randomized order). (b) In each block, a probe specifying upcoming task was presented from 8 to 12 s. Next, in each trial, four phases were presented successively (i.e., the presentation of instruction, visual stimuli, statement, and feedback). In both T1Te and T1L conditions, simple statements were utilized. Whereas in the T1Td condition, simple statements were replaced with complex statements. (c) During Task2, participants were free to tell the truth or lie on specific trials. The used statement was modified according to task requirements. The timeline for each phase was identical to other conditions from Task1. Abbreviations: T1L, lying condition from Task1; T1Te, truth-telling-easy condition from Task1; T1Td, truth-telling-difficult condition from Task1; T2L, lying condition from Task2; T2T, truth-telling condition from Task2

**TABLE 1** Statements and feedbacks used in each condition

| Task  | Condition | Statements  | Feedbacks   |
|-------|-----------|---|---|
| Task1 | T1L       | “Red”, “blue”, “triangle”, or “square”, “horizontal”, or “vertical”   | “Correct” or “incorrect”  |
|       | T1Te      |   |   |
|       | T1Td      | “The complimentary color is green”<br>“The complimentary color is yellow”<br>“Rotate 90° = horizontal”<br>“Rotate 90° = vertical”<br>“+1 edge = square”<br>“-1 edge = triangle” |   |
| Task2 | T2T/T2L   | “This figure has been presented in Task1”   | “You are lying. I got you”<br>“You are lying, sorry I am wrong” |

“In each run, you will experience two different tasks, namely Task1 and Task2. The presentation order of Task1 and Task2 was randomized. During Task1, you will be explicitly instructed to tell the truth or lie and the, and your response should be in line with the presented instruction. The aim of Task1 is to follow the presented instructions as closely as possible. As for Task2, you are free to either tell the truth or lie during which the computer algorithm will randomly detect whether you are lying or not. The goal is to deceive but avoid being detected as much as possible.”

## 2.5 | Image acquisition

MRI scanning was performed on a 3-Tesla Siemens Prisma scanner at the Imaging Center for Integrative Body, Mind, and Culture Research.

Functional data was collected with a Blood Oxygenation Level Dependent (BOLD) sequence (TR/TE = 2000/32 ms, FOV = 256 mm, matrix = 74 × 74, slice-thickness = 3.4 mm). For anatomical reference, registration and normalization of functional data to a standard T1 template (Montreal Neurological Institute, MNI) a T1 magnetization prepared, rapid-acquisition gradient echo (MPRAGE, TR/TE = 2000/2.28 ms, FOV = 256 mm, matrix = 256 × 256, slice-thickness = 1 mm) sequence was used to collect a high-resolution image of each participant’s brain. Task stimuli were presented via a projector and refracted to the subject’s visual field with a head-coil mounted mirror. X slices were collected with a 20-channel head coil. Slices were oriented roughly parallel to the AC-PC with whole-brain covered.

## 2.6 | Image analysis

To fit more precisely with the experimental purpose, the analytical scope was narrowed as the analysis progressed. Therefore, we first adopted whole-brain analysis to investigate brain areas associated with (1) difficulty processing and (2) deceptive behaviors, respectively. Then, to control for potential confound as well as to increase statistical power, we turned to ROI analysis. Our primary concerns in the ROIs analyses were: (1) whether an SVM trained with Task1 would be

able to differentiate between truth-telling and lying in Task2 (cross-task validation) and (2) whether such cross-task validation occurs across all participants. The following sections provide details for each step of analyses, including preprocessing, within-task validation (whole-brain analysis), ROIs determination, and cross-task validation (ROI analysis).

## 2.7 | Preprocessing

The preprocessing of fMRI data was conducted with SPM12. After converting all data from DICOM to NIFTI format, the following steps were performed on each experimental run. First, for each participant, the first volume in each run was aligned to the first volume of the first run (realigned) and then registered each image in each run to the first volume of that run (registered), using SPM 12. Second, realigned and registered images were being normalized to MNI space (ICBM 152 Nonlinear Asymmetrical template version 2009c) (Fonov et al., 2008).

After being processed with realigned, registered, and normalization, the general linear model (GLM) analysis was performed. Regressors of interest corresponded to the experimental condition: T1Te, T1Td, T1L. Each regressor was convolved with hemodynamic responses function (HRF). Notably, T1Te, T1Td, and T1L were modeled with the onset at the beginning of a new trial and the duration is 8 s. That is, the model was based on time series data collected during the entire trial rather than part of data derived from the statement. In addition, six motion parameters obtained from the realignment state were also included and served as nuisance regressors. With a high-pass filter of 128 s, we applied the default SPM12 options of grand mean scaling and auto-correlation modeling. On a voxel-by-voxel basis, a two-tailed *t*-test was then performed to test the null hypothesis that the BOLD signal was not explained by the experiment design. For each participant, beta values of regression coefficients were estimated, and *t*-contrasts between these values for regressors of interest and baseline BOLD signals were calculated (T1Te > baseline, T1Td > baseline, T1L > baseline). During the MVPA, the T-maps were then served as inputs for the training of SVM.

## 2.8 | Multivariate pattern analyses

The T-maps were then used as inputs for the multivariate analysis, which was conducted independently for each participant. During the MVPA, supervised Support Vector Machine (SVM), with a linear kernel, was employed in two separate stages, namely within-task classification and 2 sets of cross-task validations (within-participant and cross-participant), using the CoSMoMVPA package (<http://www.cosmomvpa.org/>) (Oosterhof, Connolly, & Haxby, 2016).

## 2.9 | Within-tasks validation

Two sets of whole-brain binary decoding were conducted to identify brain regions that differentiated experimental conditions within Task1: T1Td versus T1Te, and T1L versus T1Td. In both sets of decoding, the following procedures were followed.

First, using a searchlight method, a sphere with a radius of 3 mm was defined and centered over each voxel. Across the whole brain, a pattern of responses in each sphere is strung out in a vector for each condition. Next, for each participant, we split all runs into a training set ( $R-1$  runs,  $R$  denotes the number of all runs) and a testing set (the remaining run). Using the aforementioned feature vectors, two feature matrices representing the spatial patterns of the two sets of data were derived (one for training and one for testing). Following normalization of the training set data, an SVM model was constructed to solve two two-class problems. After repeating this process for all gray matter voxels (i.e., searchlight analysis) (Kriegeskorte, Goebel, & Bandettini, 2006) using the  $n$ -fold principle (leave-one-run out cross-validation), a three-dimensional accuracy map intended to represent the discriminating ability of classifying designated conditions was generated. (i.e., T1Te vs. T1Td & T1Td vs. T1L). To convert the accuracy map into a  $p$ -value map, a binomial distribution was tested, with the null hypothesis that no difference existed between the two groups. With a threshold of FWE-corrected  $p < .001$  and cluster size  $>5$  voxels, significant clusters to classify different groups were finally identified.

## 2.10 | ROIs determination

Even though the difficulty level between telling true (T1Td) and lying (T1L) were controlled behaviorally, it is still uncertain that Td and L are completely identical. Accordingly, during image analysis, we used a subtraction method and searched for brain areas purely related to deception. To achieve this, for each participant, the accuracy matrix obtained from the comparison of T1Te and T1Td was subtracted from the accuracy matrix obtained from the comparison of T1Td and T1L. Finally, we tested the differences in accuracy rate across all participants using one-sample  $t$ -test. Activations were considered significant at uncorrected  $p < .001$  and only clusters of 125 or more continuous voxels are reported ( $k \geq 125$ ; Langleben et al., 2005).

## 2.11 | Cross-task validation

Two sets of cross-task ROI analyses were conducted: (1) within-participant cross-task and (2) cross-participant cross-task. First, in the within-participant cross-task analysis, we examined how well the instructed responses can predict spontaneous responses. During this analysis, the patterns of responses in each functional ROI are strung out in a vector for each condition. Next, for each participant, feature vectors from Task1 constituted a training set, whereas feature vectors from Task2 constituted a testing set. Similarly, using the normalized training set data, an SVM model was constructed to solve two two-class problems (i.e., T2T vs. T2L). The same process was repeated for all gray matter voxels using the four-fold cross-validation scheme. This provided a three-dimensional accuracy map for representing the discriminability of designated conditions. The performance of the within-participant cross-task classifiers was evaluated based on the average accuracy across 21 participants (Figure 3a).

Second, to investigate whether patterns from one participant can predict that from another participant, the cross-participant cross-task analysis was conducted. To minimize the biased selection of voxels, the functionally defined ROIs were replaced with spherical ROIs (radius 8 mm) centered on the centers of mass of the prior functional ROIs, using the MarsBaR (<http://marsbar.sourceforge.net>). Accordingly, during this analysis, a pattern of responses in each anatomical ROI is strung out in a vector for each condition. In the next step, feature vectors of Task1 from  $N-1$  participants ( $N$  denotes the number of participants) were stacked and used as a training data set, while feature vectors of Task2 from the remaining participants were used as a testing set. Similarly, a SVM model was constructed to solve two two-class problems (i.e., T2T vs. T2L) following normalization of the training set data. After repeating this process for all gray matter voxels using the four-fold cross-validation scheme, a three-dimensional accuracy map intended to represent the discriminating ability of classifying designated conditions was generated. This procedure was repeated based on the leave-one-participant-out (LOPO) scheme, resulting in 21 folds. The performance of the cross-participant cross-task classifiers was attested on the average accuracy across 21 folds (Figure 3b). The raw data and code used for data analysis are available from the corresponding author upon reasonable request.

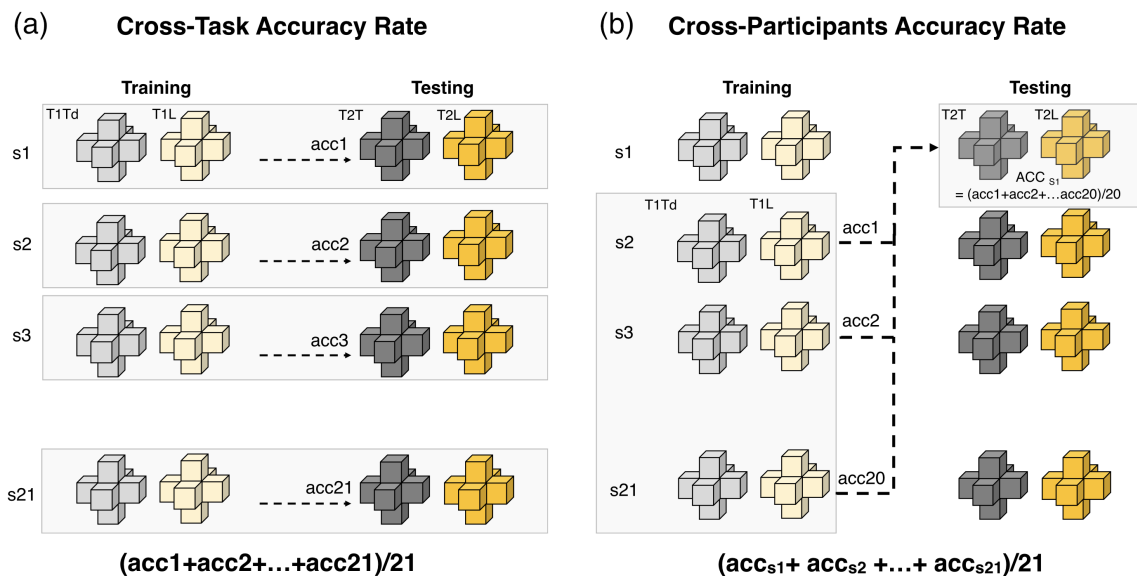
## 3 | RESULTS

### 3.1 | Behavior results

In order to obtain a clearer physiological database, in Task1, the success rate in following the task instruction was calculated for each run. For each of the included participants, the averaged success rate was higher than 75% (mean = 0.89, SD = 0.05), and the number of included runs ranged from 3 to 8.

To validate the manipulated difficulty level, a one-way ANOVA (within-subject, T1Te, T1Td, and T1L) on accuracy rate was conducted.





**FIGURE 3** (a) In the within-participant cross-task analysis, the classifiers were trained with T1Td and T1L and then used to distinguish T2T from T2L for each participant. The performances of the within-participant crosstask classifiers were examined on the average accuracy across 21 participants. (b) Cross-participant crosstask analysis was conducted based on the leave-one-participant-out (LOPO) scheme, resulting in 21 folds. For each fold, classifiers were trained with T1Td and T1L collected from 20 participants and then used to distinguish T2T from T2L in another participant who was left out for testing. The performance of the crossparticipant cross-task classifiers was attested to the average accuracy across 21 folds

Results yielded a significant effect of condition in accuracy (T1Te:  $0.95 \pm 0.06$ ; T1Td:  $0.88 \pm 0.08$ ; T1L:  $0.85 \pm 0.09$ ;  $F_{2,40} = 13.59$ ,  $p < .001$ , eta-square = 0.41). Post hoc comparisons indicated that the average accuracy was significantly higher for T1Te than T1Td ( $p_{adjusted} = .006$ ) and T1L ( $p_{adjusted} < .001$ ). Furthermore, there was no difference between T1Td and T1L in terms of accuracy rate ( $p = .219$ ). Statistical values ( $p$ -value) were corrected with the Bonferroni method. These results suggest that T1Te was easier than both T1Td and T1L while T1Td and T1L were comparable in terms of difficulty level (Figure 4).

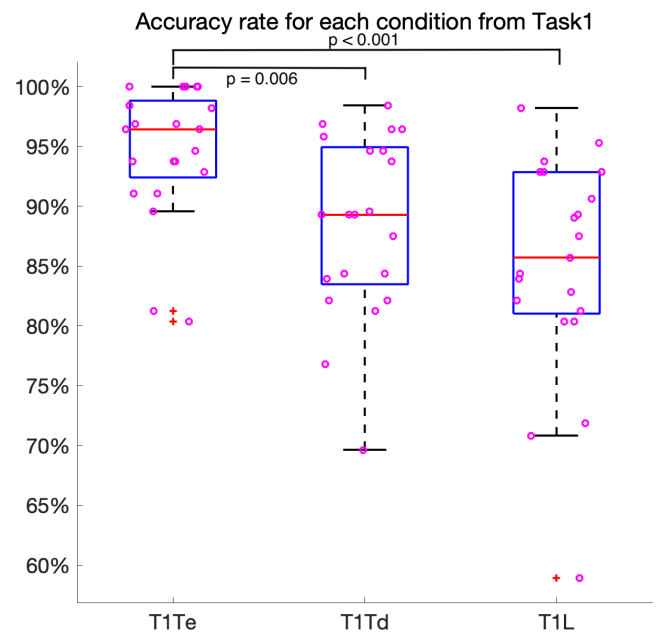
## 3.2 | Imaging results

### 3.2.1 | Whole-brain analysis

#### Within-Task decoding

With the utilization of supervised SVM, coupled with searchlight methods, two sets of binary decoding were conducted based on Task1. First, we decoded T1Td versus T1Te to examine the effect of difficulty. Results showed that brain regions including the left middle occipital gyrus (MOG), left precuneus (PrCUN), left inferior frontal gyrus (IFG), right lingual (LING), right middle occipital gyrus (MOG), right parietal lobe, right supplementary motor area (SMA), right middle temporal (MTG), right middle frontal gyrus (MFG), right frontal lobe, and bilateral superior parietal gyrus (SPG) can distinguish T1Td from T1Te (T1Td vs. T1Te, Figure S6a).

Second, we decoded T1L versus T1Td to explore the neural signatures exclusive for deception. Results showed that the left precuneus



**FIGURE 4** The accuracy of each condition in Task1. The accuracy rate from T1Te was significantly higher than both T1Td and T1L. There was no significant difference between T1Td and T1Te

(PrCUN), left postcentral (PoCG), left rolandic operculum (RO), left precentral (PrCG), right fusiform (FG), right middle temporal gyrus (MTG), right supplementary motor area (SMA), right middle frontal gyrus (MFG), right superior frontal gyrus (SFG), and right supramarginal gyrus (SMG) can differentiate T1L from T1Td (T1L

vs. T1Td, Supplementary Figure S6b). To further minimize the confound of difficulty, the second level of contrast was made by subtracting T1Td versus T1Te from T1L versus T1Td. The resulted ROIs were then employed in the follow-up validation analyses.

### 3.3 | ROI analysis

#### 3.3.1 | Within-participant cross-task validation

The generalizability of deception detection was verified by examining how well the classifiers trained with instructed behaviors can differentiate spontaneous responses. Across all functional ROIs, our results showed that after training with T1Td and T1L, classifiers from the right angular gyrus (AG), right inferior frontal gyrus (IFG), but not left postcentral gyrus (PoCG) can distinguish T2T and T2L (Figure 5a, Table S1). The average accuracy was  $0.54 \pm 0.098$  in the angular gyrus,  $0.56 \pm 0.127$  in the inferior frontal gyrus, and  $0.52 \pm 0.096$  in the left postcentral gyrus (PoCG).

#### 3.4 | Cross-participant cross-task validation

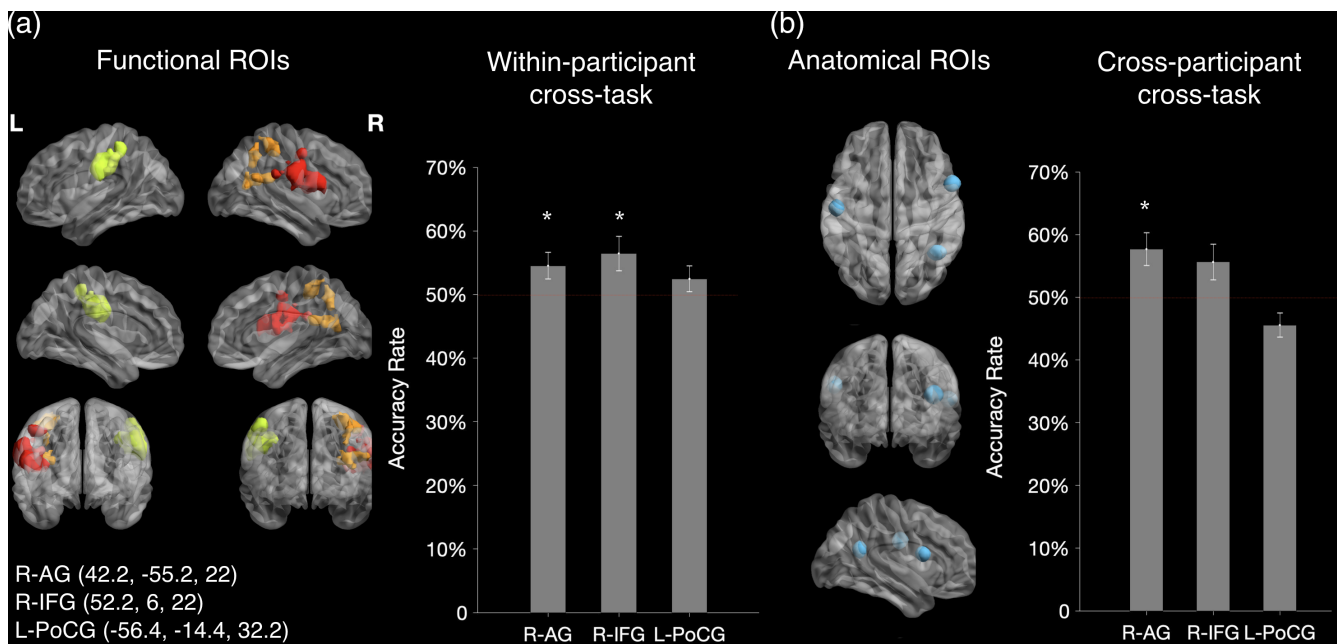
In the cross-participant cross-task analysis, we trained classifiers to distinguish T2T from T2L across participants, using a leave-one-participant-out (LOPO) scheme. Here, each of the 21 participants was used in turn as the test participant, while training on the remaining

20 participants. Then, using a *t*-test (one-tailed), the mean accuracy for each anatomical ROI was compared with the chance level (expected 0.50 accuracy).

In general, the average accuracy of validations for the left-out participant was  $0.58 \pm 0.027$  in the right angular gyrus (AG),  $0.56 \pm 0.029$  in the right inferior frontal gyrus (IFG), and  $0.46 \pm 0.020$  in the left postcentral gyrus (PoCG). Across three anatomical ROIs, accuracy in the right angular gyrus (AG) is significant compared to the null hypothesis that classification was by chance ( $p = .015$ , Cohen's  $d = 0.62$ ). This result suggests that the classifier trained to differentiate patterns between T1Td and T1L across participants allows for the classification of T2T and T2L in new participants (Figure 5b and Table S1).

## 4 | DISCUSSION

One of the long-lasting critiques for prior fMRI studies on deception detection was that relevant results might not be easily applied to forensic practices. To resolve this issue, researchers have focused on how well true responses could be differentiated from deceptive responses. For instance, using conventional univariate methods, it has been demonstrated that lying response can be distinguished from truth response within the same individual, with an accuracy of 78% (Langleben et al., 2005). Furthermore, using MRI scanning and multivoxel pattern analysis, the distinguishability between true and false responses has been increased to 100%. In addition to identifying



**FIGURE 5** (a) Within-participant cross-task validation was conducted in three functional ROIs, namely the right angular gyrus (AG), right inferior frontal gyrus (IFG), and left postcentral gyrus (PoCG). Results showed that accuracy rates in classifying T2T and T2L from the right angular (AG), right inferior frontal gyrus (IFG) but not left postcentral gyrus (PoCG) were significantly higher than chance level (50%). (b) Cross-participant cross-task validation was conducted in three 8-mm spherical regions, centered on the coordinates of the prior functional ROIs. Results showed that accuracy rates in classifying T2T and T2L from the right angular gyrus (AG) but not right inferior frontal gyrus (IFG) and left postcentral gyrus (PoCG) was significantly higher than the chance level (50%)



true and false responses within a single individual, results from cross-validation using leave-one-subject-out also yielded an accuracy rate of 88%, demonstrating the viability of cross-participant validation (Davatzikos et al., 2005).

Based on prior research, this study examined whether the ability to discriminate between true and false responses could be generalized across different tasks and individuals, which has not been explored previously. To our knowledge, the present study was the first attempt to examine the validity of deception detection research with the utilization of a dual-task paradigm, convolving two independent tasks. More specifically, with the control of difficulty level, several major

findings were obtained. First, our decoding results showed that the neural networks involved in difficulty and deception were highly overlapped. Thus, a secondary contrast was made ([T1L vs. T1Td] minus [T1Td vs. T1Te]), and our results showed that the right angular gyrus (AG), right inferior frontal gyrus (IFG), and left postcentral gyrus (PoCG) were more activated for lying than for truth-telling responses. Second, classifiers built with instructed lying and truth-telling can be used to predict spontaneous lying and truth-telling in a separate, independent task. Third, in the cross-participant analysis, classifiers trained on a separate group of participants successfully distinguished spontaneous lying and truth-telling in the left-out participants (Table 2).

**TABLE 2** Local maxima of BOLD changes while engaged in difficulty (T1Td vs. T1Te), deception (T1L vs. T1Td), and unbiased deception after subtraction ([T1Td vs. T1Te] – [T1L vs. T1Td]). Pooled group results for all subjects ( $N = 21$ ). Except for specifying, all clusters are significant at FWE corrected  $p < .001$ , and the only cluster of 5 or more continuous voxels are reported

| Brain regions(AAL)                                      | Side | BA    | MNI coordinate |       |      | k     | Z    |
|---|------|-------|----------------|-------|------|-------|------|
|   |      |       | x              | y     | z    |       |      |
| <b>T1Td vs. T1Te</b>                                    |      |       |                |       |      |       |      |
| Middle occipital gyrus                                  | R    | 18/19 | 25             | -86   | 2    | 131   | >8   |
| Middle frontal gyrus                                    | R    | 10    | 42             | 9     | 39   | 77    | 7    |
| Middle occipital gyrus                                  | L    | 18/19 | -33            | -82   | 8    | 112   | 6.65 |
| Inferior frontal gyrus (orbital part)                   | L    | 11    | -43            | 16    | -12  | 9     | 6.45 |
| Frontal lobe  | R    |       | 29             | 33    | 19   | 5     | 6.44 |
| Superior parietal gyrus                                 | L    | 5     | -22            | -69   | 53   | 24    | 6.39 |
| Precuneus   | L    | 7     | -12            | -55   | 42   | 8     | 6.14 |
| Superior parietal gyrus                                 | R    | 5     | 32             | -72   | 49   | 21    | 6.13 |
| Middle temporal gyrus                                   | R    | 21    | 49             | -55   | -2   | 7     | 6.1  |
| Supplementary motor area                                | R    | 6     | 8              | -14   | 66   | 12    | 5.97 |
| Parietal lobe   | R    |       | 32             | -45   | 46   | 5     | 5.83 |
| Lingual   | R    | 17    | 15             | -55   | 8    | 7     | 5.73 |
| <b>T1L vs. T1Td</b>                                     |      |       |                |       |      |       |      |
| Fusiform  | R    | 37    | 25             | -86   | -2   | 2,383 | >8   |
| Precentral  | L    | 4     | -50            | 3     | 36   | 99    | 7.14 |
| Supplementary motor area                                | R    | 6     | 5              | 6     | 49   | 167   | 7.1  |
| Rolandic operculum                                      | L    | 43    | -46            | -8    | 5    | 29    | 6.91 |
| Middle frontal gyrus                                    | R    | 10    | 49             | 23    | 32   | 162   | 6.88 |
| Supramarginal gyrus                                     | R    | 40    | 63             | -25   | 29   | 12    | 6.86 |
| Postcentral   | L    | 1/2/3 | -53            | -21   | 25   | 15    | 6.41 |
| Middle temporal gyrus                                   | R    | 21    | 52             | -45   | 8    | 14    | 6.29 |
| Supramarginal gyrus                                     | R    | 40    | 52             | -25   | 36   | 22    | 6.23 |
| Precuneus   | L    | 7     | -5             | -55   | 66   | 12    | 6.18 |
| Superior frontal gyrus                                  | R    | 4/6/8 | 8              | 43    | 42   | 12    | 6.18 |
| <b>[T1L vs. T1Td] minus [T1Td vs. T1Te]<sup>a</sup></b> |      |       |                |       |      |       |      |
| Postcentral gyrus                                       | L    | 1/2/3 | -56.4          | -14.4 | 32.2 | 158   | 8.43 |
| Inferior frontal gyrus                                  | R    | 11    | 52.2           | 6     | 22   | 325   | 7.46 |
| Angular gyrus   | R    | 39    | 42.2           | -55.2 | 22   | 191   | 6.46 |

<sup>a</sup>All clusters are significant at uncorrected  $p < .001$ , and the only cluster of 125 or more continuous voxels are reported.

Abbreviations: T1Td, truth-telling-difficult condition from Task1; T1Te, truth-telling-easy condition from Task1; T1L, lying condition from Task1; R, right; L, left.

## 4.1 | The neural signatures for deception

In line with most of the previous fMRI deception studies, we found that frontal (right superior frontal gyrus, right middle frontal gyrus, right supplementary motor area, and left precentral gyrus), temporal (right middle temporal gyrus, right supramarginal gyrus, and right fusiform), and parietal (left rolandic operculum, left postcentral gyrus, and left precuneus) areas can distinguish deceptive from truthful responses. These findings indicate that the neural processes involved in lying and telling the truth may differ, at least in the Dual-Task Paradigm. These large-scale activations suggest that deception required several cognitive functions, including visual perception, linguistic processing, working memory, inhibition, and intention.

To begin with, deception may require delicate visual and linguistic processing, as evidenced by higher decoding accuracy rates in the right middle temporal gyrus (MTG), right supramarginal gyrus (SMG), and right fusiform gyrus (FG). The MTG and SMG have been associated with reading (Cummine et al., 2017; Freeman, 2006) while FG has been associated with multisensory integration and perception (Gerlach et al., 2002). Both visual and linguistic processing are important components in working memory, which has been shown necessary to deception (Ito et al., 2012; Jiang et al., 2013). In light of these findings, we speculate that to deceive, one may need to retain relevant information in the working memory through strategies such as visual imagery or rehearsal.

Additionally, working memory could be modulated by executive functions, such as inhibition and flexibility (i.e., the ability to switch between tasks and demands) and even intention.

The inhibitory role of frontal regions in deception has been widely documented (Ganis et al., 2003; Ito et al., 2012; Kozel et al., 2004). For instance, in a study by Frank Andrew Kozel et al. (2004) participants were instructed to tell the truth as well as lie about whether the money was hidden below predesignated objects. More specifically, in the lying condition, participants were explicitly asked to indicate that money was not hidden in the predetermined object, but the other object. Therefore, lying participants were required to memorize the relevant task information, inhibit the truthful response as well as switch between tasks and demands. Together with frontal regions, SMA also exhibits inhibition from the perspective of complex motor control during deception (Ito et al., 2012; Frank Andrew Kozel et al., 2004). Lastly, the intention is also an integral part of deception (Lee, Leung, Lee, Raine, & Chan, 2013; Lissek et al., 2008). For instance, a study by Lissek et al. (2008) demonstrated that precuneus was activated in both cooperation and deception conditions, suggesting that from the perspective of social interaction, precuneus may be involved in emotion and intentions, requiring for belief reasoning and comprehension of cooperation and deception.

Similarly, in the current study, participants were instructed to tell the truth or lie in Task 1. To lie successfully, participants were required to memorize and then access their working memory to recall the details of what was presented. Additionally, they had to inhibit their truthful responses and make reverse reactions. All these activities were accomplished through the cooperation of the frontal,

temporal, and occipital regions. Importantly, we believe that these activities reflect lying-specific activities rather than the general differences between lying and truth-telling, such as different difficulty levels, which have been carefully controlled in our study.

## 4.2 | Predictability of deception

To examine the generalizability of deceptive behaviors, we first performed within-participant cross-task analyses to examine whether classifiers trained with neural patterns from instructed deception could predict spontaneous deception in a separate task. Among three functional ROIs, activation patterns from the right angular gyrus (AG) and right inferior frontal gyrus (IFG) successfully classified spontaneous truthful and deceptive behaviors from another task (Task2). These results suggest that instructed deception bears some similarity with spontaneous deception in the angular gyrus (AG) and inferior frontal gyrus (IFG).

Furthermore, we demonstrated that it is possible to train classifiers of lie-truth discrimination across participants. In the cross-participant cross-task analyses, we discovered that the validation accuracy of the classifier trained with the voxels in AG was significantly greater than chance. Across two sets of cross-task analysis, activities in both AG and IFG can predict deceptive behavior.

Related findings have been documented in prior literature. First, AG is implicated in a number of processes in deception. For instance, in a study by Chen et al. (2015), participants were asked to follow the instructions and answer questions either honestly (i.e., “answering correctly” condition), randomly (i.e., “answering randomly” condition) or dishonestly (i.e., “feigned memory impairment” condition). During the dishonest condition, participants were required to deliberately and tactfully falsify memory impairment. Their results showed that in comparison with the baseline, dishonest condition led to higher activation in AG. In addition, AG is also related to the intentional aspects of deception in social settings (Volz, Vogeley, Tittgemeyer, Von Cramon, & Sutter, 2015) and theory of mind (Aichhorn, Perner, Kronbichler, Staffen, & Ladurner, 2006; Seghier, 2013). In fact, apart from deception, AG is one of the centers for integrative semantic processing and knowledge retrieval (Binder, Desai, Graves, & Conant, 2009). It is proposed to be critical for the transfer and organization of multisensory information as well as for higher-level conceptualization (Wang, Baucom, & Shinkareva, 2013). Considering this, it is not surprising that AG could effectively distinguish between instructed truth and instructed lying, since these two activities may represent completely different concepts to individuals. Similarly, the discriminability of AG across two different tasks can also be explained by the conceptual difference between instructed and spontaneous deception. Together, these findings suggest that AG may be a good practical candidate in distinguishing lie-truth behaviors.

Inferior frontal gyrus (IFG) also played a key role in deception in prior reports (Hakun et al., 2008; Kireev, Korotkov, Medvedeva, Mas-haripov, & Medvedev, 2017; Sánchez et al., 2020; Vartanian et al., 2013). For example, a study conducted by Hakun et al. (2008)

revealed that IFG activity was predictive of a concealed target during a concealed information task (CIT). In addition, it has been demonstrated that in comparison with low working memory conditions, successful lying under high working memory conditions led to greater activation in the IFG (Vartanian et al., 2013). Furthermore, another study by the same team (Vartanian, Kwantes, & Mandel, 2012) also demonstrated that in comparison with less-skilled liars, successful ones had greater activation in IFG. More recently, it has also been shown that the activation of IFG was positively correlated with lying frequency across individuals (Yin & Weber, 2019) as well as memorized stimuli (Gamer, Klimecki, & Bauermann, 2009). All these findings suggest that individual differences in the ability to suppress the truth, indexed by activations from the right inferior frontal gyrus, are important in predicting lying.

In fact, using a single task, prior studies have demonstrated that it is possible to classify true and false responses with excellent accuracy (Langleben et al., 2005). This study, however, aimed to investigate whether instructed lying can predict spontaneous lying, using two tasks. Since responses from different tasks are independent, the accuracy rate is expected to be lower than prior reports, despite similar brain areas having been identified. Moreover, in our study the cross-participant analysis was conducted based on cross-task validation, therefore, it is expected to have a lower accuracy rate than prior studies (Davatzikos et al. 2005). Typically, events that occur in a real-life environment are far more complex than those occurring in a laboratory setting. Thus, to meet ecological validity, it is worthwhile to examine the degree to which lying responses are differentiated from true responses across multiple tasks. Although our approach only achieves moderate accuracy rates, it can still be used by future studies to gather large amounts of data from instructed lying to predict spontaneous lying.

## 5 | LIMITATIONS

The primary purposes of this study were to explore the neural signatures exclusive for deceptive behaviors, as well as to bridge the gap between instructed and spontaneous deception. Results from this study may contribute to future studies in this research line. However, we acknowledge that there are still several factors to be considered in future studies, particularly in the direction of strengthening the ecological validity of deception detection research.

First, it has been suggested that different types of deceptions may be associated with different patterns of brain activation (Giorgio Ganis et al., 2003). As a consequence, although our dual-task paradigm provides a model for predicting deceptive responses, it may be overly simplified. Second, the performances of the trained classifiers were far from perfect, albeit relevant results were statistically significant. Such results possibly pointed out the limitations of the current stimuli and paradigm. In brief, future studies might need to take all the aforementioned issues into account when attesting to deception detection.

## 6 | CONCLUSION

In conclusion, our results not only replicated prior reports on the underlying neural correlates of deception but further controlled for the potential difficulty confound to better localize the critical regions subserving deceptive responses. Furthermore, results from the within-participant cross-task analyses showed that spontaneous lying could be predicted using the classifier trained with the neural patterns elicited by instructed lying, revealing the possibility of applying experimental paradigms used in the laboratory settings to real forensic situations. Next, results from the cross-participant cross-task analyses further demonstrated that such validation was robust across multiple participants. Importantly, results from cross-task analyses may guide future research toward attesting deception ecologically. Bearing the aforementioned factors in mind, future studies in this research line may be advanced using a similar paradigm and analytical rationales introduced in the current study.

## ACKNOWLEDGMENTS

We are grateful for the research funding provided by the Ministry of Education (MOE), Yushan Young Scholar Program (NTU-109 V0202), Ministry of Science and Technology (109-2410-H-002-004-MY3), and National Taiwan University (110L9A00701). There are no conflicts of interest to report.

## DATA AVAILABILITY STATEMENT

The raw data and code used for data analysis are available from the corresponding author upon reasonable request.

## ORCID

Yen-Ju Feng  <https://orcid.org/0000-0002-7265-6709>

## REFERENCES

- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage*, 30(3), 1059–1068. <https://doi.org/10.1016/j.neuroimage.2005.10.026>
- Bhatt, S., Mbwana, J., Adeyemo, A., Sawyer, A., Hailu, A., & VanMeter, J. (2009). Lying about facial recognition: An fMRI study. *Brain and Cognition*, 69(2), 382–390. <https://doi.org/10.1016/j.bandc.2008.08.033>
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12), 2767–2796.
- Chen, Z. X., Xue, L., Liang, C. Y., Wang, L. L., Mei, W., Zhang, Q., & Zhao, H. (2015). Specific marker of feigned memory impairment: The activation of left superior frontal gyrus. *Journal of Forensic and Legal Medicine*, 36, 164–171. <https://doi.org/10.1016/j.jflm.2015.09.008>
- Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta-analyses. *Cerebral Cortex*, 19(7), 1557–1566. <https://doi.org/10.1093/cercor/bhn189>
- Cummine, J., Hanif, W., Dymouriak-Tymashov, I., Anchuri, K., Chiu, S., & Boliek, C. A. (2017). The role of the supplementary motor region in overt Reading: Evidence for differential processing in SMA-proper and pre-SMA as a function of task demands. *Brain Topography*, 30(5), 579–591. <https://doi.org/10.1007/s10548-017-0553-3>

- Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughhead, J. W., ... Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage*, 28(3), 663–668.
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70(5), 979–995. <https://doi.org/10.1037//0022-3514.70.5.979>
- Fonov, V., Evans, A. C., Botteron, K., C. R. Almi+, McKinstry, R. C., D. L. Collins, and Brain Development Cooperative Group. (2008). Unbiased average age-appropriate atlases for pediatric studies. *Bone*, 23(1), 1–7. <https://doi.org/10.1021/nn2045246.Multifunctional>
- Freeman, G. L. (2006). Development of neural mechanisms. *Physiological Psychology*, 6(6), 8–20. <https://doi.org/10.1037/11170-002>
- Gamer, M., Klimecki, O., & Bauermann, T. (2009). fMRI-activation patterns in the detection of concealed information rely on memory-related effects. *Social cognitive and affective neuroscience*, 7(5), 506–515.
- Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex*, 13(8), 830–836. <https://doi.org/10.1093/cercor/13.8.830>
- Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., & Schendan, H. E. (2011). Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *NeuroImage*, 55(1), 312–319. <https://doi.org/10.1016/j.neuroimage.2010.11.025>
- Gerlach, C., Aaside, C. T., Humphreys, G. W., Gade, A., Paulson, O. B., & Law, I. (2002). Gerlach 2002.Pdf, 40, 1254–1267.
- Hakun, J. G., Seelig, D., Ruparel, K., Loughhead, J. W., Busch, E., Gur, R. C., & Langleben, D. D. (2008). fMRI investigation of the cognitive structure of the concealed information test. *Neurocase*, 14(1), 59–67. <https://doi.org/10.1080/13554790801992792>
- Ito, A., Abe, N., Fujii, T., Hayashi, A., Ueno, A., Mugikura, S., ... Mori, E. (2012). The contribution of the dorsolateral prefrontal cortex to the preparation for deception and truth-telling. *Brain Research*, 1464, 43–52. <https://doi.org/10.1016/j.brainres.2012.05.004>
- Jiang, W., Liu, H., Liao, J., Ma, X., Rong, P., Tang, Y., & Wang, W. (2013). A functional MRI study of deception among offenders with antisocial personality disorders. *Neuroscience*, 244, 90–98. <https://doi.org/10.1016/j.neuroscience.2013.03.055>
- Kireev, M., Korotkov, A., Medvedeva, N., Masharipov, R., & Medvedev, S. (2017). Deceptive but not honest manipulative actions are associated with increased interaction between middle and inferior frontal gyri. *Frontiers in Neuroscience*, 11(AUG), 1–12. <https://doi.org/10.3389/fnins.2017.00482>
- Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., & George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry*, 58(8), 605–613. <https://doi.org/10.1016/j.biopsych.2005.07.040>
- Kozel, F. A., Padgett, T. M., & George, M. S. (2004). A replication study of the neural correlates of deception. *Behavioral Neuroscience*, 118(4), 852–856. <https://doi.org/10.1037/0735-7044.118.4.852>
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10), 3863–3868. <https://doi.org/10.1073/pnas.0600244103>
- Langleben, D. D., Loughhead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, 26(4), 262–272. <https://doi.org/10.1002/hbm.20191>
- Lee, T. M. C., Leung, M. K., Lee, T. M. Y., Raine, A., & Chan, C. C. H. (2013). I want to lie about not knowing you, but my precuneus refuses to cooperate. *Scientific Reports*, 3, 1–5. <https://doi.org/10.1038/srep01636>
- Lee, T. M. C., Liu, H. L., Chan, C. C. H., Ng, Y. B., Fox, P. T., & Gao, J. H. (2005). Neural correlates of feigned memory impairment. *NeuroImage*, 28(2), 305–313. <https://doi.org/10.1016/j.neuroimage.2005.06.051>
- Lee, T. M. C., Liu, H. L., Tan, L. H., Chan, C. C. H., Mahankali, S., Feng, C. M., ... Gao, J. H. (2002). Lie detection by functional magnetic resonance imaging. *Human Brain Mapping*, 15(3), 157–164. <https://doi.org/10.1002/hbm.10020>
- Lissek, S., Peters, S., Fuchs, N., Witthaus, H., Nicolas, V., Tegenthoff, M., ... Brüne, M. (2008). Cooperation and deception recruit different subsets of the theory-of-mind network. *PLoS One*, 3(4), e2023. <https://doi.org/10.1371/journal.pone.0002023>
- Lykken, D. T. (1981). *A tremor in the blood: Uses and abuses of the lie detector*, 94, 8(1925–1927). Harvard Law Review. <https://doi.org/10.2307/1340741>
- Mason, R. A., Just, M. A., Keller, T. A., & Carpenter, P. A. (2003). Ambiguity in the brain: What brain imaging reveals about the processing of syntactically ambiguous sentences. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29(6), 1319–1338. <https://doi.org/10.1037/0278-7393.29.6.1319>
- Núñez, J. M., Casey, B. J., Egner, T., Hare, T., & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *NeuroImage*, 25(1), 267–277. <https://doi.org/10.1016/j.neuroimage.2004.10.041>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Encyclopedia of Clinical Neuropsychology*, 9(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVA: Multi-modal multivariate pattern analysis of neuroimaging data in matlab/GNU octave. *Frontiers in Neuroinformatics*, 10(JUL), 1–27. <https://doi.org/10.3389/fninf.2016.00027>
- Phan, K. L., Magalhaes, A., Ziemlewick, T. J., Fitzgerald, D. A., Green, C., & Smith, W. (2005). Neural correlates of telling lies: A functional magnetic resonance imaging study at 4 tesla. *Academic Radiology*, 12(2), 164–172. <https://doi.org/10.1016/j.acra.2004.11.023>
- Sánchez, N., Masip, J., & Gómez-Ariza, C. J. (2020). Both high cognitive load and transcranial direct current stimulation over the right inferior frontal cortex make truth and lie responses more similar. *Frontiers in Psychology*, 11, 1–14. <https://doi.org/10.3389/fpsyg.2020.00776>
- Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing. Scientific Analysis and Public Controversy. *American Psychologist*, 40(3), 355–366. <https://doi.org/10.1037/0003-066X.40.3.355>
- Seghier, M. L. (2013). The angular gyrus: Multiple functions and multiple subdivisions. *The Neuroscientist*, 19(1), 43–61. <https://doi.org/10.1177/1073858412440596>
- Spence, S. A. (2004). The deceptive brain. *Journal of the Royal Society of Medicine*, 97(1), 6–9. <https://doi.org/10.1258/jrsm.97.1.6>
- Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*, 12(3), 2849–2853. <https://doi.org/10.1097/00001756-200109170-00019>
- Spence, S. A., Kaylor-Hughes, C., Farrow, T. F. D., & Wilkinson, I. D. (2008). Speaking of secrets and lies: The contribution of ventrolateral prefrontal cortex to vocal deception. *NeuroImage*, 40(3), 1411–1418. <https://doi.org/10.1016/j.neuroimage.2008.01.035>
- Steinbrook, R. (1992). The polygraph test—A flawed diagnostic method. *New England Journal of Medicine*, 327(2), 122–123. <https://doi.org/10.1056/NEJM199207093270212>
- Tang, H., Lu, X., Cui, Z., Feng, C., Lin, Q., Cui, X., ... Liu, C. (2018). Resting-state functional connectivity and deception: Exploring individualized deceptive propensity by machine learning. *Neuroscience*, 395(15), 101–112. <https://doi.org/10.1016/j.neuroscience.2018.10.036>
- Vartanian, O., Kwantes, P. J., Mandel, D. R., Bouak, F., Nakashima, A., Smith, I., & Lam, Q. (2013). Right inferior frontal gyrus activation as a neural marker of successful lying. *Frontiers in Human Neuroscience*, 7, 1–6. <https://doi.org/10.3389/fnhum.2013.00616>
- Vartanian, O., Kwantes, P., & Mandel, D. R. (2012). Lying in the scanner: Localized inhibition predicts lying skill. *Neuroscience Letters*, 529(1), 18–22. <https://doi.org/10.1016/j.neulet.2012.09.019>

- Volz, K. G., Vogeley, K., Tittgemeyer, M., Von Cramon, D. Y., & Sutter, M. (2015). The neural basis of deception in strategic interactions. *Frontiers in Behavioral Neuroscience*, 9, 1–12. <https://doi.org/10.3389/fnbeh.2015.00027>
- Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24(4), 239–263. <https://doi.org/10.1023/A:1006610329284>
- Wang, J., Baucom, L. B., & Shinkareva, S. V. (2013). Decoding abstract and concrete concept representations based on single-trial fMRI data. *Human brain mapping*, 34(5), 1133–1147. <https://doi.org/10.1002/hbm.21498>
- Wang, X., Hutchinson, R., & Mitchell, T. M. (2004). Training fMRI classifiers to discriminate cognitive states across multiple subjects *Proceedings of 17th Annual Conference on Neural Information Processing Systems*, Vancouver and Whistler, Canada.
- Yin, L., Reuter, M., & Weber, B. (2016). Let the man choose what to do: Neural correlates of spontaneous lying and truth-telling. *Brain and Cognition*, 102, 13–25. <https://doi.org/10.1016/j.bandc.2015.11.007>
- Yin, L., & Weber, B. (2019). I lie, why don't you: Neural mechanisms of individual differences in self-serving lying. *Human Brain Mapping*, 40(4), 1101–1113. <https://doi.org/10.1002/hbm.24432>
- Zhang, M., Liu, T., Pelowski, M., & Yu, D. (2017). Gender difference in spontaneous deception: A hyperscanning study using functional near-infrared spectroscopy. *Scientific Reports*, 7(1), 1–13. <https://doi.org/10.1038/s41598-017-06764-1>
- Zheltayakova, M., Kireev, M., Korotkov, A., & Medvedev, S. (2020). Neural mechanisms of deception in a social context: An fMRI replication study. *Scientific Reports*, 10(1), 1–12. <https://doi.org/10.1038/s41598-020-67721-z>

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Feng, Y.-J., Hung, S.-M., & Hsieh, P.-J. (2022). Detecting spontaneous deception in the brain. *Human Brain Mapping*, 43(10), 3257–3269. <https://doi.org/10.1002/hbm.25849>