RESEARCH ARTICLE

WILEY

Mixed-effects multilevel analysis followed by canonical correlation analysis is an effective fMRI tool for the investigation of idiosyncrasies

Sungman Jo[1] | Hyun-Chul Kim[1,2] | Niv Lustig[1] | Gang Chen[3] | Jong-Hwan Lee[1]

[1]Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

[2]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

[3]Scientific and Statistical Computing Core, NIMH/NIH/DHHS, Bethesda, Maryland

Correspondence
Jong-Hwan Lee, Department of Brain and Cognitive Engineering, Korea University, Anam-ro 145, Seongbuk-gu, Seoul 02841, Republic of Korea.
Email: jonghwan_lee@korea.ac.kr

Abstract

We report that regions-of-interest (ROIs) associated with idiosyncratic individual behavior can be identified from functional magnetic resonance imaging (fMRI) data using statistical approaches that explicitly model individual variability in neuronal activations, such as mixed-effects multilevel analysis (MEMA). We also show that the relationship between neuronal activation in fMRI and behavioral data can be modeled using canonical correlation analysis (CCA). A real-world dataset for the neuronal response to nicotine use was acquired using a custom-made MRI-compatible apparatus for the smoking of electronic cigarettes (e-cigarettes). Nineteen participants smoked e-cigarettes in an MRI scanner using the apparatus with two experimental conditions: e-cigarettes with nicotine (ECIG) and sham e-cigarettes without nicotine (SCIG) and subjective ratings were collected. The right insula was identified in the ECIG condition from the χ^2-test of the MEMA but not from the t-test, and the corresponding activations were significantly associated with the similarity scores ($r = -.52$, $p = .041$, confidence interval [CI] $= [-0.78, -0.17]$) and the urge-to-smoke scores ($r = .73$, $p < .001$, CI $= [0.52, 0.88]$). From the contrast between the two conditions (i.e., ECIG > SCIG), the right orbitofrontal cortex was identified from the χ^2-tests, and the corresponding neuronal activations showed a statistically meaningful association with similarity ($r = -.58$, $p = .01$, CI $= [-0.84, -0.17]$) and the urge to smoke ($r = .34$, $p = .15$, CI $= [0.09, 0.56]$). The validity of our analysis pipeline (i.e., MEMA followed by CCA) was further evaluated using the fMRI and behavioral data acquired from the working memory and gambling tasks available from the Human Connectome Project.

KEYWORDS
canonical correlation analysis, electronic cigarette, functional MRI, Human Connectome Project, mixed-effects multilevel analysis, nicotine craving

1 | INTRODUCTION

Recently, there has been growing interest in the use of functional magnetic resonance imaging (fMRI) data to investigate the relationship between the brain and behavior (Hubbard, Cyrus Arman, Ramachandran, & Boynton, 2005; Kohn, Coen-Cagli, Kanitscheider, & Pouget, 2016; Lamichhane, Westbrook, Cole, & Braver, 2020; Rousselet & Pernet, 2012; Zhou, Li, Zhou, Li, & Cui, 2018). In this approach, to identify the regions-of-interest (ROIs) in the brain that are potentially associated with behavioral data, statistical parametric maps are estimated using a general linear model (GLM) for individual fMRI data at the first level. Group inference is then conducted at the second level by applying statistical models such as Student's t-tests and analysis of variance (ANOVA) to the fMRI data obtained from a group of subjects (Mumford & Nichols, 2009). Consequently, the ROIs that show statistically significant neuronal activations across groups are identified from possible contrasts between experimental conditions. ROIs are also used to evaluate the links between the corresponding neuronal activations and non-neuronal data such as behavioral data and/or individual traits (Hubbard et al., 2005; Kohn et al., 2016). However, these conventional statistical approaches (i.e., t-tests and ANOVA) do not explicitly include a variable to model between-subject variability despite the fact that between-subject variance within the same experimental group/condition is potentially much larger than within-subject variance (Chen, Saad, Nath, Beauchamp, & Cox, 2012; Lindquist, Spicer, Asllani, & Wager, 2012).

It is possible that individual traits and behavioral patterns lead to subject-specific variability in neuronal activations at the first level in addition to producing an estimation error for the effect size. This subject-specific variability can contribute to significant variability in effect sizes across subjects within the same group and/or experimental condition. Thus, the between-subject variability in neuronal activations is key to predicting individual-specific non-neuronal data (Lindquist et al., 2012). In this context, mixed-effects models that consider both within- and between-subject variability (Beckmann, Jenkinson, & Smith, 2003; Friston et al., 2002; Mumford & Nichols, 2009; Woolrich, Behrens, Beckmann, Jenkinson, & Smith, 2004) are useful for acquiring valid and precise estimates of group inference (Lindquist et al., 2012). Thus, we were motivated to evaluate a mixed-effects multilevel analysis (MEMA) model based on the homogeneity of the effect estimates defined by comparing the between-subject variance with the within-subject variance (Chen et al., 2012) used to identify brain regions associated with the idiosyncrasies of the participants.

A variety of regression models have been employed to examine the brain–behavior relationship. Traditional simple or multiple regression analyses have been used to test the relationship between neuronal activations captured by fMRI and related behavioral data (Anders, Lotze, Erb, Grodd, & Birbaumer, 2004; Clark, 2002; de Hollander, Forstmann, & Brown, 2016; Lindquist, 2008; Valente,

Castellanos, Vanacore, & Formisano, 2014). Recently, regression approaches based on multivariate neuronal activation patterns rather than conventional univariate activations have been used to predict behavior and/or cognitive data (Dosenbach et al., 2010; Haynes, 2015; Norman, Polyn, Detre, & Haxby, 2006). More recently, a growing number of neuroimaging studies have focused on the prediction of individual behavior and/or cognitive scores using machine-learning approaches (Arbabshirani, Plis, Sui, & Calhoun, 2017; Cui & Gong, 2018; Dosenbach et al., 2010; Gabrieli, Ghosh, & Whitfield-Gabrieli, 2015; Sui et al., 2018). An example of these multivariate machine-learning-based regression approaches is canonical correlation analysis (CCA), which enables two sets of multivariate variables to be investigated by maximizing the correlation between them (Hotelling, 1935; Ter Braak, 1986). For example, the CCA approach has been suggested as a viable option to determine the link between neuronal activations estimated from fMRI data and subjective behavioral and/or (pre-)clinical data (Smith et al., 2015; Wang et al., 2020; Zhuang et al., 2019).

In this study, we hypothesized that the use of a MEMA model to identify ROIs from fMRI data based on both between-subject and within-subject variability followed by the use of CCA could be employed to examine the possible association between brain ROIs and idiosyncratic data such as individual behavioral data. To evaluate our hypothesis using the data acquired in a real-world scenario, fMRI and associated behavioral data were collected in the context of nicotine craving. Previous studies on the neuronal circuitry of nicotine craving have identified representative brain regions such as the insula (Benowitz, 2010; Moran et al., 2018; Naqvi, Rudrauf, Damasio, & Bechara, 2007) and the orbitofrontal cortex (Franklin et al., 2007; Jasinska, Zorick, Brody, & Stein, 2014). For example, Naqvi et al. (2007) reported that people with lesions of the insula were able to quit tobacco smoking easily with no reported conscious urges. Furthermore, it appears that the neuronal response to nicotine in the insula region is highly heterogeneous across subjects due to individual traits, the strength of nicotine craving in participants on the day of the experiment (Gilbert et al., 1998), and the amount of nicotine absorbed (Farsalinos, Yannovits, Sarri, Voudris, & Poulas, 2018).

To collect fMRI data and associated behavioral data related to nicotine craving, we designed an MRI-compatible apparatus for the smoking of an electronic cigarette (e-cigarette) so that we could systematically switch between nicotine-present and nicotine-absent conditions while simultaneously acquiring the fMRI data. We evaluated the efficacy of our analysis approach (i.e., MEMA followed by CCA) in the context of nicotine craving using the neuronal activations obtained from the fMRI data and behavioral data represented by the subjective ratings collected from the participants while they used the e-cigarette smoking apparatus. We also applied our analysis approach to working memory (WM) and gambling (GB) task fMRI data from the Human Connectome Project (HCP) dataset (Barch et al., 2013; Van Essen et al., 2012) to evaluate the efficacy of our approach for a large public dataset.

2 | MATERIALS AND METHODS

2.1 | Overview

Figure 1 illustrates the overall process for our study using e-cigarette data. Once the fMRI and behavioral data were acquired using our MRI-compatible e-cigarette apparatus, the fMRI data were preprocessed and analyzed using a GLM for each of the experimental conditions at the first level. At the second level, the MEMA was applied utilizing both beta weights (i.e., the estimated voxel-wise effect sizes from the GLM) and their variance across participants to estimate the between-subject variability compared to within-subject variability as a proxy for the homogeneity of the effect sizes. The homogeneity was measured using χ^2-tests, and the ROIs were identified accordingly based on voxel clusters with a high level of significance. CCA was then employed to investigate the relationship between the neuronal activations in each of the ROIs and the behavioral data across subjects in a leave-one-subject-out cross-validation (LOOCV) framework. The ROIs identified from the t-tests were also evaluated using CCA in the LOOCV framework for comparison. We also applied our approach to the fMRI data in the HCP dataset to further validate our method (MEMA followed by CCA) in a 2×5 nested cross-validation framework (see Section 2.9).

2.2 | Participants

The entire study protocol was approved by the Institutional Review Board at Korea University and all participants provided signed written

FIGURE 1 Overall process of the study. More detailed information can be found in the Methods, including "Overview." ECIG, e-cigarette with nicotine; SCIG, sham e-cigarette without nicotine; α, familywise error rate; LOOCV, leave-one-subject-out cross-validation; MEMA, mixed-effects multilevel analysis

consent. Healthy right-handed current e-cigarette users were recruited using word-of-mouth, flyers distributed across campus, and the internet. Volunteers were prescreened for eligibility by phone and then interviewed in person. To be included in the study, they had to meet the following criteria: aged >19 years, continuous use of e-cigarettes for at least the last 3 months with a minimum nicotine concentration in the e-liquid of 12 mg/ml (Hobkirk et al., 2018), and a minimum e-liquid consumption of 30 ml/month. Potential participants were excluded if they had a history of brain trauma, any neurological condition, a Fagerström Test of Nicotine Dependence (FTND) score of <4 (Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991; Kim, Yoo, Tegethoff, Meinlschmidt, & Lee, 2015; Lee, Kim, & Kim, 2012), a Penn State Electronic Cigarette Dependence Index (PSECDI; Foulds et al., 2015) of <4, or any contraindication on an MRI scan. Thirty-two male participants were recruited and participated in a non-MRI session in the laboratory to evaluate the efficacy of our MRI-compatible e-cigarette apparatus. Nineteen of the subjects also participated in an MRI session to investigate their neuronal responses to e-cigarette use during another visit.

Before attending the non-MRI and MRI sessions, participants were asked to abstain from smoking both tobacco cigarettes and e-cigarettes for at least 3 hr to increase their nicotine craving (Schuh & Stitzer, 1995). Carbon monoxide (CO) levels in their exhaled breath were assessed using a piCO Smokerlyzer (Bedfont Scientific, Ltd., Rochester, UK) both when the subjects arrived at the experiment site to confirm that they had abstained from smoking (CO level <10 ppm) and after each of the non-MRI and MRI sessions to confirm that e-cigarette smoking does not increase CO levels. All participants reported that they had abstained from smoking for at least 3 hr prior to arrival. Table 1 presents the demographic and smoking information for the participants.

A set of questionnaires were given to participants before and after the non-MRI and MRI experiments to evaluate their craving for nicotine. These included the Questionnaire on Smoking Urges (QSU; Cox, Tiffany, & Christen, 2001), the Minnesota Nicotine Withdrawal Scale (MNWS; Hughes & Hatsukami, 1986), and a short version of the Shiffman–Jarvik Withdrawal Scale (SJWS; Shiffman & Jarvik, 1976). A craving scale (CRS) was also used as an intuitive measure of subjective craving, with a visual analogue scale between 0 and 10 (Fregni, Fecteau, & Pascual-Leone, 2008). Additional measures were used to evaluate the participants' mood, including the Perceived Stress Scale (PSS; Cohen, Kamarck, & Mermelstein, 1994), the Beck Depression Inventory (BDI; Beck, Steer, & Carbin, 1988), the Beck Anxiety Inventory (BAI; Julian, 2011), and the Multidimensional Mood State Questionnaire (MDBF; Steyer, Eid, & Schwenkmezger, 1997). Refer to the Supporting Information for more details on the experiment ("MRI-compatible e-cigarette apparatus" and "Experimental paradigm").
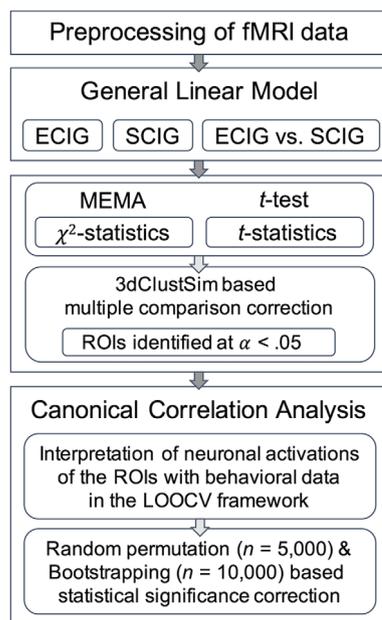
2.3 | Imaging parameters

Blood-oxygenation-level-dependent (BOLD) fMRI data were acquired using a 3-T MRI scanner (Tim Trio, Siemens, Erlangen, Germany) with a 12-channel head coil and a standard gradient-echo echo-planar-

TABLE 1 Demographic information and nicotine dependence for the participants (mean ± *SD*)

		Non-MRI experiment (n = 32)	MRI experiment (n = 18)	p-value from two-sample t-test[a]
Age (years)		25.54 ± 2.64	25.42 ± 2.71	.87
E-cigarette use (months)		13.35 ± 9.67	13.52 ± 8.64	.95
FTND		4.71 ± 1.23	4.68 ± 1.29	.67
PSECDI		9.62 ± 3.57	9.31 ± 3.76	.79
CO level (ppm)	Pre-experiment	8.12 ± 5.48	6.92 ± 5.53	.13
	Post-experiment	7.51 ± 4.80	6.37 ± 4.86	.35

Abbreviations: CO, carbon monoxide; FTND, Fagerström Test of Nicotine Dependence; PSECDI, Penn State Electronic Cigarette Dependence Index; SD, Standard Deviation.
[a]The participants who participated in the MRI experiment were a subset of the participants who participated in the non-MRI experiment.

imaging pulse sequence (repetition time = 2,000 ms, echo time = 30 ms, flip angle = 90°, field of view = 240 × 240 mm^2, voxel size = 3.75 × 3.75 × 4.0 mm^3, 36 interleaved axial slices with no gap). The cardiac and respiratory signals were simultaneously acquired with the BOLD fMRI data. The T1-weighted structural MRI volume was acquired as anatomical images using a magnetization-prepared rapid gradient-echo (MP-RAGE) pulse sequence (repetition time = 1,900 ms, echo time = 2.52 ms, flip angle = 9°, field of view = 256 × 256 mm^2, voxel size = 1.0 × 1.0 × 1.0 mm^3, 176 sagittal slices with no gap).

2.4 | Preprocessing of the fMRI data

The standard preprocessing pipeline available in the "afni_proc.py" script of the Analysis of Functional Neuroimages (AFNI) software (afni.nimh.nih.gov/afni) was used in this study. This pipeline included despiking (using "3dDespike"), slice timing correction ("3dTshift"), physiological noise reduction ("3dretroicor"), co-registration of EPI volumes to the anatomical images of the corresponding subject using the local Pearson's correlation cost function ("align_epi_anat.py" with an "lpc" option), mapping to the Montreal Neurological Institute (MNI) coordinate space with the MNI_avg152T1 + tlrc template ("@auto_tlrc"), censoring the EPI volumes with frame-wise displacement (FD) of over 0.5 with severe head motion (Power et al. 2019), spatial smoothing using a Gaussian kernel with a blur size of 6 mm full-width at half-maximum (FWHM), and regressing out low-frequency fluctuations in the BOLD fMRI data using nuisance regressors from the 0th to the highest order polynomials, which were estimated from the length of a run varying depending on the smoking duration ("3dDetrend").

2.5 | Individual-level analysis using a general linear model

The preprocessed BOLD fMRI volumes were concatenated across the two runs and analyzed using a GLM for individual-level analysis (using "3dREMLfit" for generalized least squares-based estimation in AFNI). The ANATICOR was used to de-noise the BOLD signal from the

cardiorespiratory data using the "@ANATICOR" script in AFNI (Jo, Saad, Kyle Simmons, Milbury, & Cox, 2010), and the six motion parameters and their first-order derivatives (i.e., task-unrelated nuisance regressors in the GLM) were also regressed out of the BOLD signal. The contrasts in the effect sizes between the task-related inhaling/exhaling periods with smoke versus without smoke (e.g., T4 − T2 vs. T2 − T1 for ECIG and T10 − T8 vs. T8 − T7 for SCIG in Figure S2) were used to obtain the effect sizes (i.e., beta values) of the BOLD responses associated with smoking an e-cigarette. The hemodynamic responses of the tasks (i.e., task-related regressors) were modeled using these task periods convolved with the canonical hemodynamic response function (using the option, "-stim_times WAV" in the 3dDeconvolve command of AFNI). The length of the fMRI runs varied depending on the inhaling/exhaling periods of the participants (mean ± *SD*: 380.5 ± 54.4 TRs), in which the length of the ECIG and SCIG blocks was 43.0 ± 8.8 TRs and 37.0 ± 8.6 TRs, respectively.

2.6 | Group inference using mixed-effects multilevel analysis

Using the individual beta-value maps, MEMA was employed for group inference. In the MEMA framework, the individual effect size is expressed at the second level as follows:

$$y_i = \alpha_0 + \delta_i + \varepsilon_i, \ y_i \sim N(\alpha_0, \tau^2 + \sigma_i^2), \tag{1}$$

where y_i is the effect size (i.e., the beta value β) of an experimental condition or the contrast of experimental conditions from the GLM for subject i ($i = 1, 2, ..., P$), α_o is the group-level effect across all P subjects, δ_i is the deviation of subject i from the group effect, which follows a normal distribution with mean α_o and variance τ^2, that is, $N(\alpha_0, \tau^2)$, ε_i is the estimation error (i.e., residual) of β for the ith subject, which follows $N(0, \sigma_i^2)$, and σ_i^2 represents the within-subject variability for an individual GLM. To test the homogeneity of the effect sizes across subjects, Q-statistics (Cochran, 1954), which represent an approximate χ^2 distribution with $(P - 1)$ degrees of freedom and are often referred to as Cochran's χ^2 test (Viechtbauer, 2010), can be used (Chen et al., 2012):

$$Q = \sum_{i=1}^{P} \frac{(y_i - \mu)^2}{\sigma_i^2} \qquad (2)$$

where μ is the weighted mean of the individual effect sizes divided by the inverse of the within-subject variability, assuming no between-subject variability ($\tau^2 = 0$). Thus, the Q score represents the sum of squares of the deviation of the individual-level effect size in comparison to the group-level effect size normalized by within-subject variability. As a result, the null hypothesis can be defined as the between-subject variability not being substantially greater than the within-subject variability:

$$H_0 : \tau^2 = 0, \qquad (3)$$

where $\tau^2 = \max\left(0, \dfrac{Q-(P-1)}{\sum_i \frac{1}{\sigma_i^2} - \sum_i \left(\frac{1}{\sigma_i^2}\right)^2 / \sum_i \frac{1}{\sigma_i^2}}\right)$ (see "Appendix A. Mixed-effects multilevel analysis for fMRI data" for more details). The MEMA model reduces to a fixed-effects model in this null hypothesis (Chen et al., 2012). This null hypothesis is rejected and an alternative hypothesis is accepted when there is substantial between-subject variability compared to within-subject variability (Higgins & Thompson, 2002). In this scenario, a brain region where τ^2 is substantially large indicates the presence of heterogeneity across subjects due to substantial differences in their effect sizes. Therefore, it is believed that Q-statistics provide a statistical framework for defining an ROI whose neuronal activations (i.e., effect sizes) are potentially linked to subjective behavioral data and/or psychological processes of interest that vary between subjects (Chen et al., 2012; Lindquist et al., 2012).

The MEMA implemented in AFNI (3dMEMA) was used in our study to identify ROIs with substantial heterogeneity in their neuronal activations across all 18 subjects associated with the ECIG and SCIG blocks independently and with the contrast between them. Multiple comparison correction was conducted using Monte-Carlo simulations ("3dClustSim" in AFNI) with a voxel-wise uncorrected p-value threshold that resulted in a voxel cluster with a corrected p-value (α) of <.05 at the cluster level. More specifically, the cluster correction was performed for all ROIs identified using the t-tests and χ^2-tests. The uncorrected p-value used for the cluster-based multiple comparison correction was selected by considering whether the uncorrected p-value would still provide a statistical power greater than .8 given our sample size (Hintze, 2008; Suresh & Chandrashekara, 2012). The ROIs (i.e., the clusters of brain regions) that remained from this multiple comparison correction were subject to subsequent analysis using CCA.

2.7 | CCA of the link between brain and behavioral data

CCA (Hotelling, 1935; Ter Braak, 1986) is a machine-learning method that can be used to investigate the relationship between two sets of multivariate variables, such as brain data and behavioral data (Smith

et al., 2015; Wang et al., 2020). Suppose that X ($P \times l$) and Y ($P \times m$) are two sets of l- and m-dimensional variables from P subjects; the CCA finds the tunable parameter sets (i.e., canonical weights) $\boldsymbol{a}$ ($l \times 1$) and $\boldsymbol{b}$ ($m \times 1$) by maximizing the correlation between the two sets of multivariate variables as follows:

$$\rho = \underset{\boldsymbol{a},\boldsymbol{b}}{\arg\max}(X\boldsymbol{a}, Y\boldsymbol{b}) \qquad (4)$$

Once $\boldsymbol{a}$ and $\boldsymbol{b}$ are found, the multivariate variables $U \triangleq X\boldsymbol{a}$ ($P \times 1$) and $V \triangleq Y\boldsymbol{b}$ ($P \times 1$) are the first pair of canonical variates. In subsequent CCA, another set of parameters for $\boldsymbol{a}$ and $\boldsymbol{b}$ can also be found with the constraint that the second pair of canonical variates are orthogonal to the first pair. This process of finding another set of canonical variates can continue until the maximum number of sets, $\min(l,m)$, of the canonical variates and the canonical weights are found. See "Appendix B. Canonical correlation analysis (CCA)" for more details.

2.8 | CCA in a leave-one-subject-out cross-validation framework

A LOOCV approach (Esterman, Tamber-Rosenau, Chiu, & Yantis, 2010; Varoquaux et al., 2017) was adopted for the CCA framework to avoid potential overfitting when the dataset for all subjects was used (Dinga et al., 2019; Le Floch et al., 2012). Thus, the data for one subject were used as validation to evaluate the performance of the CCA model trained using the data from the remaining subjects (i.e., $n = P - 1$). Refer to the Supporting Information "CCA in leave-one-subject-out cross-validation (LOOCV) for e-cigarette data" for more details. The CCA function "canoncorr.m" implemented in MATLAB R2018a was used, and the p-values of the Pearson's linear correlations between pairs of canonical variates were corrected using a null distribution obtained from 5,000 random permutations with pseudo-randomly shuffled subject indices for the brain data (Nichols & Holmes, 2002). In addition, bootstrapping was conducted 10,000 times using subject indices with replacement, and a 95% confidence interval (CI) for the correlation coefficients was used to determine significant relationships between the brain and behavioral data (Kim et al., 2019; Pernet, Wilcox, & Rousselet, 2013; Terhune, Russo, Near, Stagg, & Kadosh, 2014). The median absolute deviation (MAD) was calculated for each of the canonical variates for the neuronal activations and behavioral data to exclude potential outliers (Kim et al., 2019; Leys, Ley, Klein, Bernard, & Licata, 2013; Seo, 2006).

2.9 | Evaluation of our analysis pipeline using the HCP dataset

In the analysis using our collected fMRI data acquired from our MRI-compatible e-cigarette apparatus, the same dataset used for the MEMA was used for the CCA due to the relatively small sample size. Thus, there was the potential for circular analysis even though the data for the CCA model training and the test data were separated by

an LOOCV framework. In order to mitigate this potential issue and the weak form of cross-validation (i.e., LOOCV) (Varoquaux et al., 2017; Vu et al., 2020), we also evaluated our analysis pipeline using the task fMRI (tfMRI) data available from the HCP dataset. Of the seven tasks in the HCP, the WM and GB tasks were adopted because we thought that the between-subject heterogeneity compared to within-subject homogeneity in their neuronal activations may be substantial for the working memory performance and gambling outcomes compared to more straightforward tasks such as motor tasks. The selection of the two tasks was also partly because there are two distinct conditions in each of the two tasks (i.e., 2-back vs. 0-back in WM; reward vs. punishment in GB) similar to our e-cigarette data (i.e., the ECIG vs. SCIG conditions).

The analysis pipeline of the MEMA followed by CCA was conducted in a 2×5 nested CV framework (with two outer folds and five inner folds) and consisted of two stages: (1) application of the MEMA to one of the two folds in an outer loop in order to identify the ROIs; (2) application of the CCA to one remaining fold in order to learn and evaluate the relationship between the neuronal activations and the behavioral data based on the five-fold CV in an inner loop (i.e., 4 folds were used to train the CCA model parameters and the one remaining fold was used to test the trained CCA model). Please refer to the Supporting Information for more details on the HCP dataset ("WM and GB tasks, and the associated behavioral data") and the preprocessing and individual-level analysis used to obtain neuronal activations ("Preprocessing of the tfMRI data and estimation of neuronal activations").

2.10 | ROIs from MEMA

The participants for each of the WM ($n = 537$) and GB ($n = 874$) tasks were pseudo-randomly divided into two folds. The MEMA was applied to each of the two folds using both their beta-valued maps and the corresponding t-statistics maps from the individual-level GLM. The statistically significant clusters of brain regions from the Student's t-test (p-value $<10^{-4}$ for WM and p-value $<10^{-10}$ for GB with a minimum of 40 connected voxels) and from the χ^2-test (p-value $<.01$ for WM and p-value $<.005$ for GB with a minimum of 40 connected voxels) were then identified. The anatomical label of the significant clusters was obtained from the 116 regions in the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002).

This application of MEMA to each of the two randomly divided participant subgroups (i.e., two folds) and the identification of significant clusters with their AAL information was repeated 10 times. The reproducibility of the clusters identified from the t-test and χ^2-test was then evaluated using the 20 sets of significant clusters for each of the statistical tests. More specifically, if any of the 116 regions in the AAL were identified in over 80% of the 20 sets of t-test and χ^2-test from MEMA (i.e., at least 16 sets), the corresponding AAL regions were considered to be reproducible from

the corresponding statistical test and were defined as ROIs. Consequently, the significant clusters in each of the reproducible ROIs were subjected to the CCA using the beta-values of the clusters and behavioral data of the participants that belong to the one remaining fold that was not used for the MEMA.

The application of the CCA only to the reproducibly significant clusters for each of the statistics from the MEMA was because we wanted to reduce the overall computational load. This approach can prevent exhaustive analyses when we (a) apply the CCA with five-fold CV in an inner loop to each of the clusters identified from the MEMA using the dataset in an outer loop, (b) perform the analysis using switched inner and outer folds, and (c) repeat the 2×5 nested CV 10 times.

2.11 | CCA to identify the link between the brain and behavioral data

In an inner loop of the 2×5 nested CV using two randomly divided sub-groups, the CCA model was trained and tested in the five-fold CV using half of all of the subjects in the inner loop. In detail, the brain and behavioral data from four folds were used to train the CCA model and the data from the one remaining fold were used to evaluate the trained CCA model as an independent set. Principal component analysis (PCA) was applied to multivoxel beta values across the four folds of the CCA training data to reduce the dimensions of the brain data for each of the clusters with strong statistical evidence (ranging from 153 to 1,134 voxels). This was because the number of voxels in the clusters was substantially higher than the two-dimensional behavioral data (i.e., accuracy and RT for WM; percentage of rewards in comparison to punishment and RT for GB), which can potentially cause the CCA model to overfit the data. Four thresholds for the number of principal components (PCs; eigenvectors from eigen-decomposition using the beta-values of the voxels in the cluster across the participants) were considered: (a) the number of PCs that corresponds to a cumulative explained variance of ≈90%, (b) 50 PCs, (c) 100 PCs, and (d) 150 PCs.

In the test phase, the PCs obtained using the training data were applied to the data in the one remaining test fold, and consequently the trained CCA model was tested. This was repeated five times using each of the five folds as test data and the remaining four folds as training data. Correlation analysis using the canonical variates of the neuronal activations and canonical variates of the behavioral data was conducted for the training and test data. Correlational analysis using the canonical variates of neuronal activations and each of the two types of behavioral data was also performed. The resulting p-values were corrected based on a null distribution obtained from random permutations ($n = 10,000$). When the correlation coefficients between the canonical variates of the neuronal activations in the identified ROIs and the canonical variates of the behavioral data were consistently significant over three of the four scenarios for dimension reduction, the corresponding associations were reported.

3 | RESULTS

3.1 | Behavioral and psychological data

Table 2 shows that smoking an e-cigarette using our MRI-compatible apparatus alleviated the nicotine craving ($p_{corrected}$ <10^{-3}) and enhanced the mood of the participants, possibly because of the effect of nicotine absorption. The similarity, urge-to-smoke, and smoking duration scores obtained during the fMRI runs were comparable with those from the non-MRI sessions (Figure 2), which demonstrates the validity of our e-cigarette smoking apparatus in an MRI environment. The significant differences in the behavioral and psychological data between the ECIG and SCIG conditions indicate that our apparatus correctly delivered the nicotine-containing smoke or nicotine-free smoke to participants depending on the experimental condition.

3.2 | ROIs from Student's *t*-tests and χ^2-tests

Figure 3 presents the ROIs identified using Student's *t*-tests and χ^2-tests across the experimental conditions. The uncorrected *p*-values for the cluster-based multiple comparison correction of the χ^2-tests and the Student's *t*-test were $p = 10^{-8}$ and $p = .001$, respectively, and the corresponding statistical power was .98 and .99, respectively. In the ECIG condition, the left cerebellum, left inferior frontal gyrus, right middle temporal gyrus, right angular gyrus, and precuneus were

identified using Student's *t*-tests, whereas the left temporal pole, right cerebellum, right insula, left hippocampus, and left thalamus were identified using χ^2-tests. In the SCIG condition, the left angular gyrus and right middle temporal gyrus were identified with *t*-tests, and the right thalamus, left inferior frontal gyrus, and left thalamus were identified with χ^2-tests (α <.05 or equivalently, *t*-score >3.97, *p* <10^{-3} with a minimum of 63 connected voxels for the Student's *t*-tests; χ^2-score = 71.93, degrees of freedom [*d.f.*] = 17, *p* <10^{-8} with a minimum of 40 connected voxels for the χ^2-tests). In the contrast of the two conditions (i.e., ECIG > SCIG), the supplementary motor area from the *t*-tests and the left middle temporal gyrus, right orbitofrontal cortex (OFC), and right middle temporal gyrus from the χ^2-tests were identified (α <.05 or equivalently, *t*-score >3.22, *p* <.005 with a minimum of 123 connected voxels for the *t*-tests; χ^2-score >35.72, *d.f.* = 17, *p* <.005 with a minimum of 63 connected voxels for the χ^2-tests). No clusters were found for the opposite contrast (i.e., ECIG < SCIG) even with a marginal threshold (uncorrected *p* <.05). Table 3 summarizes the details of the identified ROIs.

3.3 | Interpreting the ROIs based on the homogeneity of the effect sizes

Figure 4a presents the variability of the beta-value estimates (blue) across subjects (defined as between-subject variability) for each of the voxels within an ROI found from the contrast of the conditions

TABLE 2 Summary statistics for the questionnaires obtained from the participants (mean ± *SD*)

		Non-MRI experiment (*n* = 32)			MRI experiment (*n* = 18)		
		Pre-experiment	Post-experiment	*p*-value	Pre-experiment	Post-experiment	*p*-value
CRS		6.52 ± 1.98	3.49 ± 2.49	**<10^{-3}**	7.30 ± 1.66	2.77 ± 1.99	**<10^{-3}**
MNWS		9.09 ± 5.21	3.66 ± 3.10	**<10^{-3}**	11.37 ± 4.99	3.79 ± 2.35	**<10^{-3}**
QSU		3.70 ± 1.54	1.95 ± 1.06	**<10^{-3}**	4.46 ± 1.09	2.53 ± 1.07	**<10^{-3}**
SJWS	Craving	5.46 ± 1.25	3.09 ± 1.33	**<10^{-3}**	5.83 ± 0.91	3.29 ± 1.16	**<10^{-3}**
	Physical symptoms	2.09 ± 1.13	1.83 ± 1.12	.82	2.12 ± 1.03	1.75 ± 0.99	.67
	Stimulation/sedation	3.28 ± 1.95	2.55 ± 1.59	.76	3.26 ± 1.66	2.47 ± 1.35	.18
	Psychological symptoms	3.68 ± 0.87	2.55 ± 0.73	**<10^{-3}**	3.98 ± 0.91	2.59 ± 0.88	**<10^{-3}**
	Appetite	2.34 ± 1.82	2.03 ± 1.53	.72	2.68 ± 2.00	2.10 ± 1.59	.66
	Overall average	3.89 ± 0.77	2.58 ± 0.65	**<10^{-3}**	4.09 ± 0.61	2.61 ± 0.69	**<10^{-3}**
BDI		7.39 ± 4.66	3.32 ± 2.71	**<10^{-3}**	6.95 ± 4.35	3.47 ± 3.36	**<10^{-3}**
BAI		6.71 ± 4.36	3.97 ± 2.99	**<10^{-3}**	5.42 ± 4.35	2.63 ± 2.75	**<10^{-3}**
PSS		16.39 ± 5.59	14.10 ± 4.61	.64	18.16 ± 4.97	14.58 ± 4.63	.06
MDBF	Good–bad	15.22 ± 3.38	18.97 ± 2.33	**<10^{-3}**	13.57 ± 2.83	18.26 ± 2.83	**<10^{-3}**
	Awake–tired	16.56 ± 4.43	19.59 ± 3.49	**.03**	14.32 ± 2.85	18.21 ± 3.34	**<10^{-3}**
	Calm–nervous	18.03 ± 4.92	22.69 ± 3.24	**<10^{-3}**	17.74 ± 4.39	22.68 ± 3.56	**<10^{-3}**
	Total	49.81 ± 10.63	61.25 ± 7.25	**<10^{-3}**	45.63 ± 7.60	59.16 ± 8.34	**<10^{-3}**

Note: Values in bold represents *p*-value less than .05. The scores before and after the experiment were compared using paired *t*-tests and the corresponding *p*-values were Bonferroni-corrected.
Abbreviations: BAI, Beck Anxiety Inventory; BDI, Beck Depression Inventory; CRS, Craving Scale; MNWS, Minnesota Nicotine Withdrawal Scale; MDBF, Multidimensional Mood State Questionnaire; PSS, Perceived Stress Scale; QSU, Questionnaire of Smoking Urges; *SD*, standard deviation; SJWS, Shiffman–Jarvik Withdrawal Scale-Short Version.
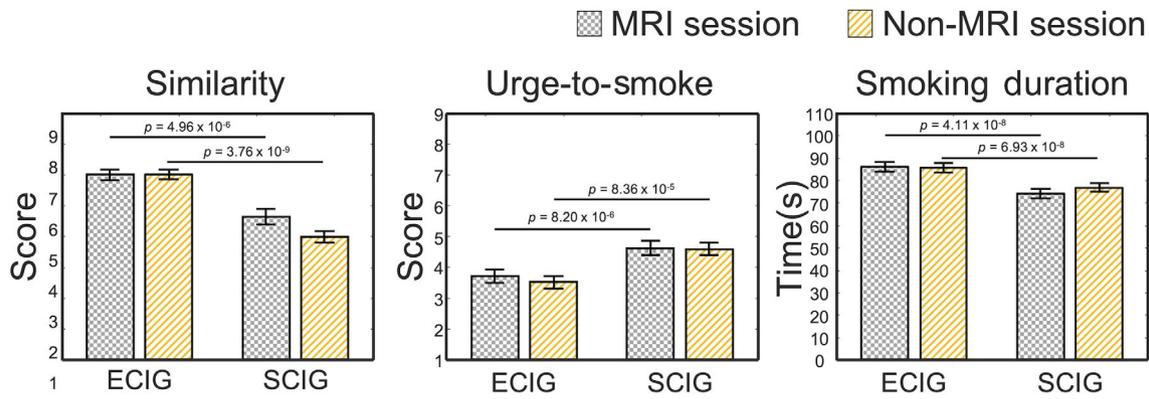
FIGURE 2 Similarity, urge-to-smoke, and smoking duration scores obtained from the non-MRI and MRI sessions stratified by experimental condition (bars and error bars represent mean and standard error of the mean, respectively; *p*-values are Bonferroni-corrected)
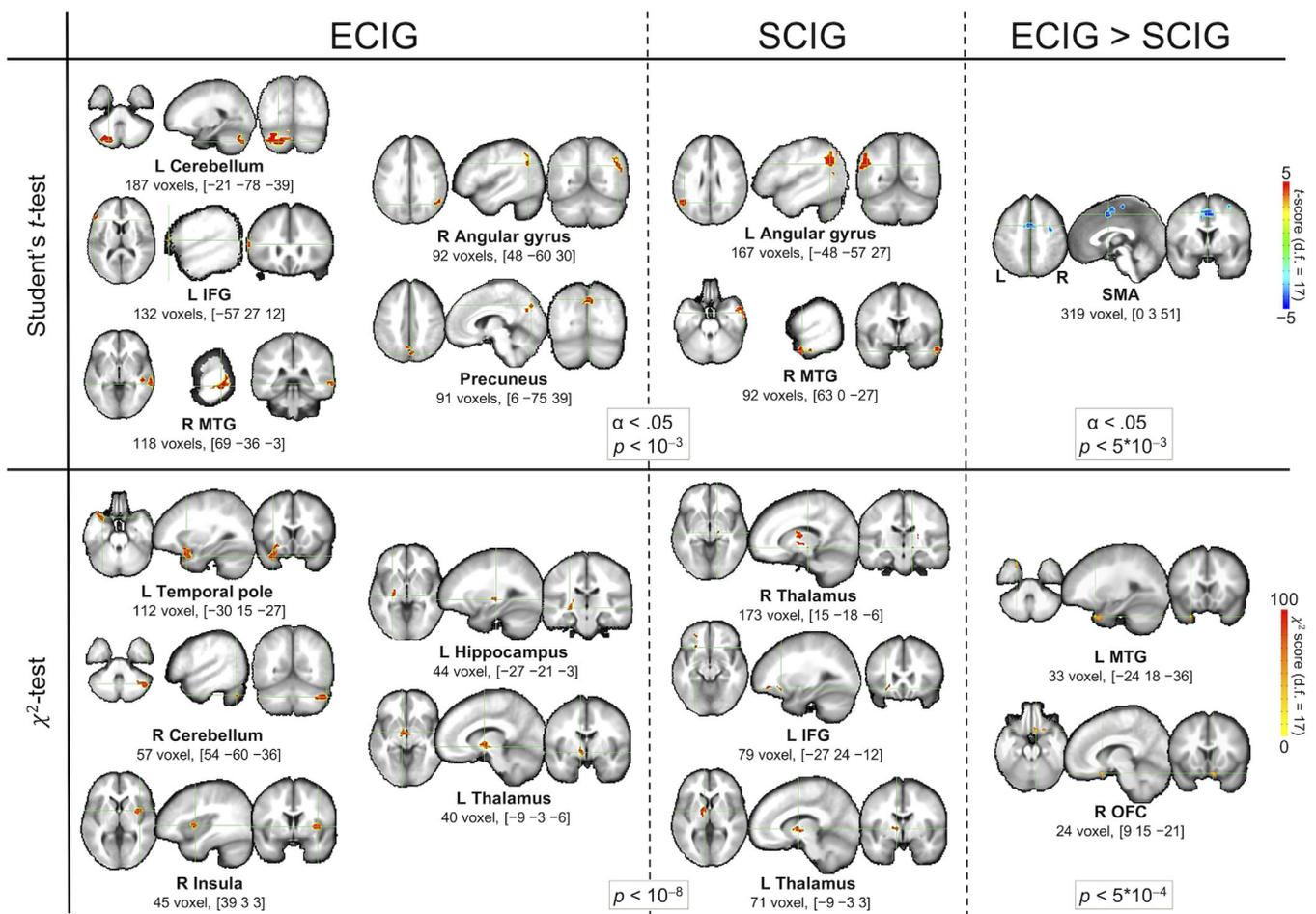
FIGURE 3 ROIs identified from the *t*-tests and χ^2-tests of the MEMA for the experimental conditions ECIG, SCIG, and ECIG > SCIG. The ROIs were identified from cluster-based multiple comparison correction (α <.05) using the uncorrected *p*-values (*t*-test: 10^{-3}; χ^2-test: 10^{-8}, 10^{-8}, and .0005 for ECIG, SCIG, and ECIG > SCIG, respectively) from the conditions to achieve a power of >0.8 given our sample size. The three numbers in each bracket indicate the MNI coordinates (i.e., [*x, y, z*]) of the maximum *t*-score or χ^2-score within the ROI. No cluster was found in the ECIG < SCIG contrast. ECIG, e-cigarette with nicotine; SCIG, sham e-cigarette without nicotine; L, left; R, right; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; MTG, middle temporal gyrus; OFC, orbitofrontal cortex; SMA, supplementary motor area; *d.f.*, degree-of-freedom

(i.e., ECIG > SCIG) and the corresponding *t*-scores (red; the marker and whiskers denote the mean and two standard deviations, respectively). It is evident that all of the voxels of the three ROIs found using

the χ^2-tests had a substantial level of variability in their beta values across subjects (i.e., τ^2) compared to the mean beta value, which resulted in low *t*-scores overall (red). In contrast, all of the voxels of

TABLE 3 Brain regions identified from t-tests and χ^2-tests by applying the MEMA to fMRI data obtained from the ECIG and SCIG conditions and their contrast

Statistics	Condition	Brain region	Voxel count	t-score ($d.f. = 17$)	χ^2 score ($d.f. = 17$)	MNI coordinate (x, y, z [mm])
Student's t	ECIG (α <.05; p <.001)	L cerebellum	187	7.70	47.04	−21, −78, −39
		L IFG	132	7.04	55.97	−57, 27, 12
		R MTG	118	6.37	44.16	69, −36, −3
		R angular gyrus	92	5.43	38.17	48, −60, 30
		Precuneus	91	5.61	27.90	6, −75, 39
	SCIG (α <.05; p <.001)	L angular gyrus	167	5.66	43.83	−48, −57, 27
		R ITG	92	7.96	37.48	63, 0, −27
	ECIG > SCIG (α <.05; p <.005)	SMA	319	−5.50	12.78	0, 3, 51
χ^2	ECIG (p <10^{-8})	L temporal pole	112	3.11	73.24	−30, 15, −27
		R cerebellum	57	1.41	73.73	54, −60, −36
		R insula	45	2.32	72.23	39, 3, 3
		L hippocampus	44	1.94	74.99	−27, −21, −3
		L thalamus	40	1.60	76.67	−9, −3, −6
	SCIG (p <10^{-8})	R thalamus	173	2.50	91.72	15, −18, −6
		L IFG	79	2.99	80.91	−27, 24, −12
		L thalamus	71	1.89	85.68	−9, −3, 3
	ECIG > SCIG (p <5×10^{-4})	L MTG	33	−1.79	59.77	−24, 18, −36
		R OFC	24	2.20	72.69	9, 15, −21

Note: The χ^2 score ranges between 0 and 100.

Abbreviations: $d.f.$, degrees of freedom; ECIG, e-cigarette with nicotine; IFG, inferior frontal gyrus; L, left; ITG, inferior temporal gyrus; MNI, Montreal Neurological Institute; MTG, middle temporal gyrus; OFC, orbitofrontal cortex; SMA, supplementary motor area; R, right; SCIG, sham e-cigarette without nicotine.

the SMA identified in the t-tests had substantially lower variability in the beta values across subjects (and thus lower τ^2 values), which resulted in higher t-scores from the t-test compared to the voxels found in the χ^2-tests. Figure 4b displays the beta values and their variability (i.e., the residual variance of the GLM, σ_i^2) in the peak voxels that exhibited the highest effect size for each of the χ^2-test and t-test scores for each of the subjects (defined as within-subject variability). Interestingly, the homogeneity of the beta values, which was defined as the inverse of the total variance (i.e., the sum of the between-subject variability, τ^2 and within-subject variability, σ_i^2) was substantially lower for the ROIs identified in the χ^2-tests than for those identified in the t-tests.

3.4 | Relationship between the neuronal activations in the ROIs and behavioral data

Using the CCA, significant associations between the neuronal activations and behavioral data were found from the ROIs identified only from the χ^2-tests. Figure 5 presents the scatter plots for the canonical variates of the behavioral data and the neuronal activations from the right insula, an ROI identified under the ECIG condition. The two canonical variates exhibited a significant association ($\rho = 0.78$, $p_{corrected}$ <.001, 95% CI = [0.66, 0.88]). The insula activations had a

significant negative association with the similarity score ($r = -.52$, $p_{corrected} = .041$, 95% CI = [−0.78, −0.17]) and a significant positive association with the urge-to-smoke score ($r = .73$, $p_{corrected} = .001$, 95% CI = [0.52, 0.88]) only in the ECIG condition. Smoking duration had no significant association with the neuronal activations of the right insula.

Figure 6 shows that the right OFC, an ROI found in the contrast of the conditions, had a significant association with the canonical variates of the behavioral data ($\rho = 0.73$, $p_{corrected} = .002$, 95% CI = [0.56, 0.91]). The neuronal activations in the right OFC were negatively correlated with the similarity score ($r = -.58$, $p_{corrected} = .01$, 95% CI = [−0.84, −0.17]) and demonstrated a weakly positive correlation with the urge-to-smoke score ($r = .34$, $p_{corrected} = .15$, 95% CI = [0.09, 0.56]). Smoking duration had no meaningful association with the neuronal activations of the right OFC ($r = -.38$, $p_{corrected} = .13$, 95% CI = [−0.70, 0.15]).

3.5 | ROIs from the tfMRI data in the HCP dataset

Figure 7 shows the brain regions identified for the WM task using (a) t-test and (b) χ^2-test applied to all 537 subjects. Table S1 provides detailed information on the identified clusters. Using 2 × 5 nested CV across the subjects, 27 reproducible ROIs were identified from the
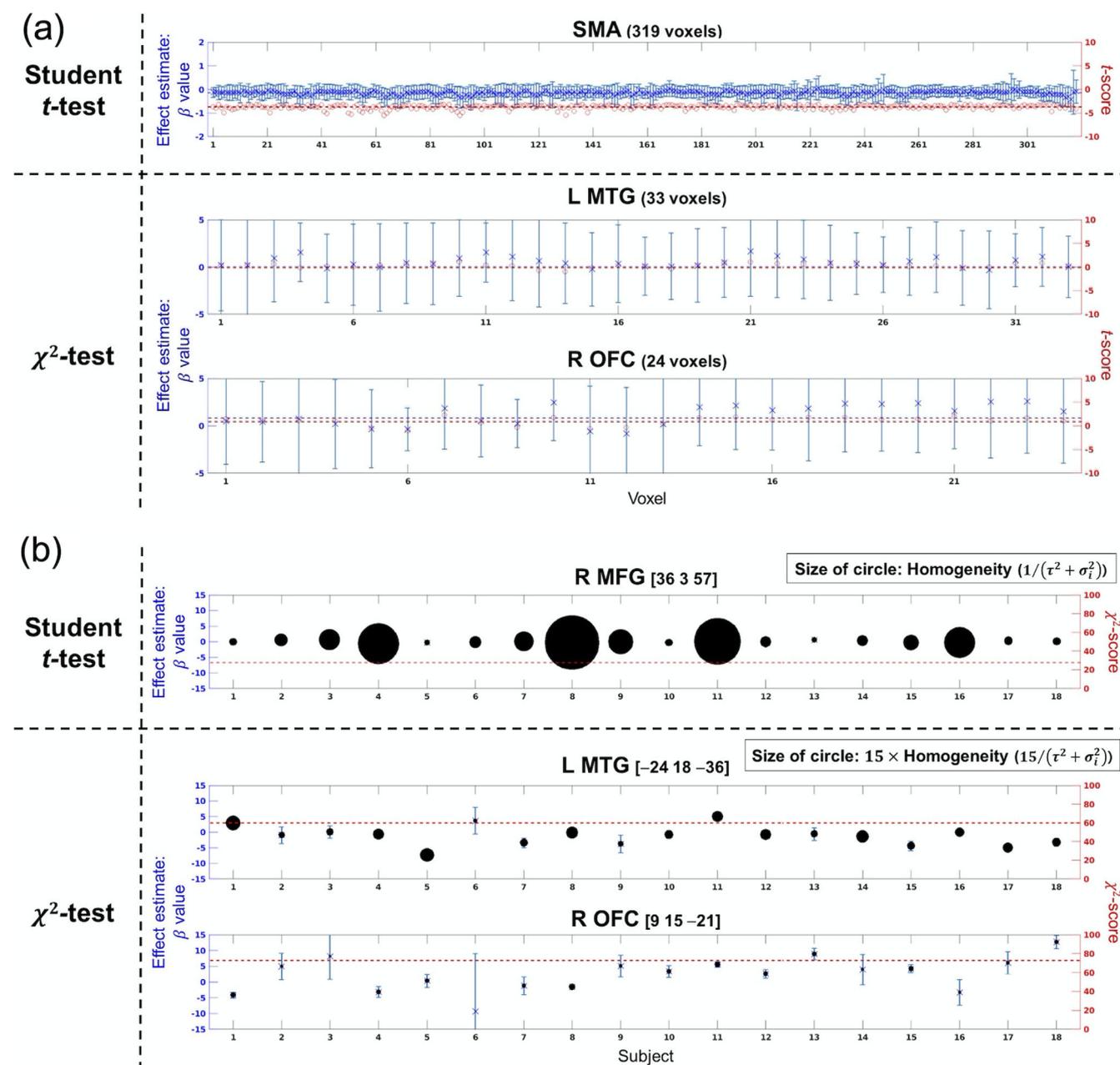
FIGURE 4 Evaluation of the identified ROIs from the contrast of the experimental conditions (ECIG > SCIG) based on (a) the estimate of the effects across the subjects at each voxel of an ROI (i.e., between-subject variability) and (b) residual variance (σ_i^2) at the peak voxel (maximum χ^2 score) of an ROI for each subject. The corresponding t-scores or χ^2-scores are also shown (red). The mean and two standard deviations of the values are presented as a marker and the whiskers in the error bar plot, respectively (blue). In (a), the blue dashed line is the average effect estimate across all voxels within an ROI, and the red dashed line is the average t-score across all voxels. In (b), the red dashed line is the χ^2-score at the peak voxel. In addition, the homogeneity of the effect sizes for the ith subject (i.e., $1/(\tau^2 + \sigma_i^2)$, where τ^2 and σ_i^2 are the between-subject and within-subject variabilities, respectively) modeled in the χ^2 statistics is represented as the radius of the circle in (b). Note that the radius size for the ROIs found from the χ^2-tests in (b) was 15 times larger than the original level of homogeneity. The right MFG exhibited the maximum χ^2 score among the voxels in the SMA cluster found from the t-tests (the bottom row in b). L, left; R, right; MTG, middle temporal gyrus; OFC, orbitofrontal cortex; SMA, supplementary motor area; MFG, middle frontal gyrus

t-test ($p < 10^{-4}$, $t[268] > 3.95$) across the motor area, visual area, parietal areas, and cerebellum regions (Figure 7c). From the χ^2-test ($p < .01$, $\chi^2(268) > 324.78$), four reproducible ROIs (i.e., the right orbital area of the superior frontal gyrus [oSFG], the orbital area of the inferior frontal gyrus [oIFG], the rectus gyrus, and insula) were identified over 80% of the analyses (Figure 7d). Table S2 summarizes these ROIs.

Figure 8 presents the brain regions identified for the GB task using (a) t-test and (b) χ^2-test applied to all 874 subjects. Detailed
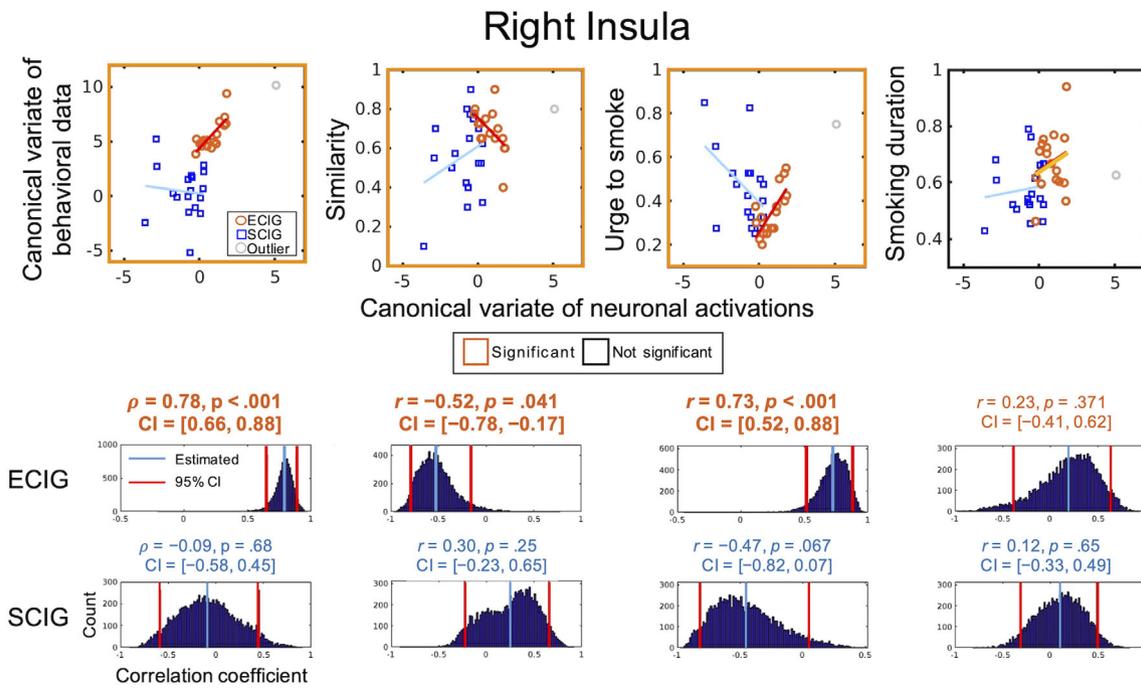
Right Insula

FIGURE 5 Relationship between the neuronal activations of the right insula identified from the ECIG condition using the χ^2 statistics from the MEMA and behavioral data from the CCA. (Top row) Relationship between the canonical variates of the two datasets (leftmost column) and between the canonical variates of the neuronal activations and each of the behavioral scores (orange box indicates a significant relationship between them, that is, $p_{corrected}$ <.05 and 0 ∉ 95% CI). (Bottom two rows) Bootstrapping results (95% CI in red; estimated correlation coefficient in blue) and the p-value for the estimated correlation coefficients corrected with a random permutation (significant cases denoted in bold). Refer to the Section 2.8 for more details. ECIG, e-cigarette with nicotine; SCIG, sham e-cigarette without nicotine; CI, confidence interval

Right Orbitofrontal cortex
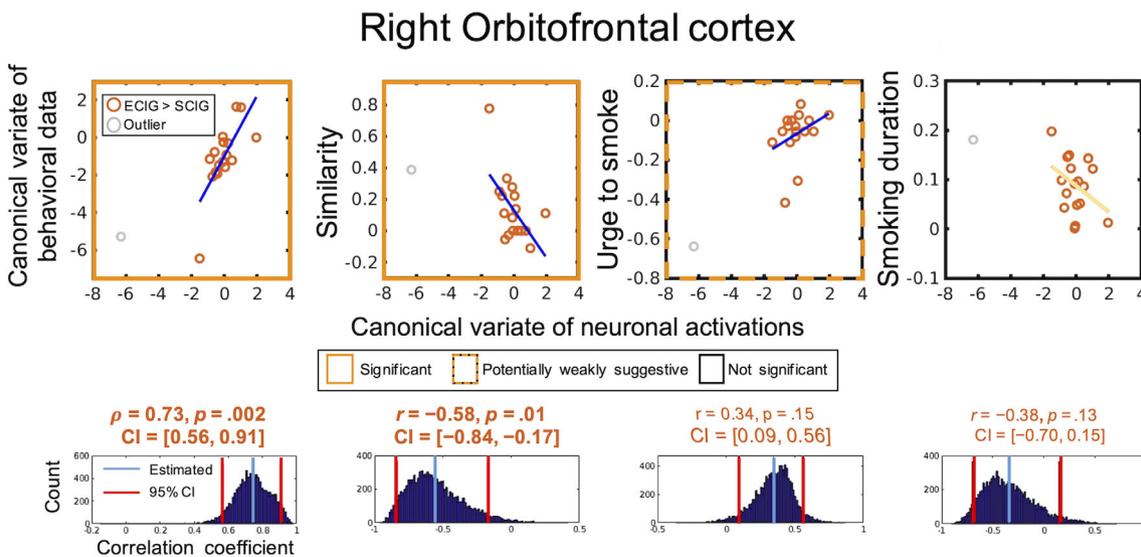
FIGURE 6 Relationship between the neuronal activations of the right orbitofrontal cortex identified from the ECIG > SCIG condition using the χ^2 statistics from the MEMA and behavioral data from the CCA. (Top row) Relationship between the two canonical variates (leftmost column) and between the canonical variates of the neuronal activations and each of the behavioral scores (an orange box indicates a significant relationship, that is, $p_{corrected}$ <.05 and 0 ∉ 95% CI and a dashed orange box indicates a weak relationship, $p_{corrected}$ >.05 and 0 ∉ 95% CI). (Bottom row) Bootstrapping results (95% CI in red; estimated correlation coefficients in blue) and the p-value of the estimated correlation coefficients corrected with a random permutation (significant cases denoted in bold and weak associations denoted in italics). Refer to the Section 2.8 for more details. ECIG, e-cigarette with nicotine; SCIG, sham e-cigarette without nicotine; CI, confidence interval

information on these brain regions is provided in Table S1. Using 2×5 nested CV across the subjects, 30 reproducible ROIs were identified from the t-test (p <10^{-10}, $t[436]$ >6.63), including the right SFG,

the middle frontal gyrus (MFG), oSFG, MFG, and the bilateral occipital gyrus (Figure 8c). From the χ^2-test (p <.005, $\chi^2(437)$ >515.81), 18 reproducible ROIs were found, including the left MFG, IFG, insula,
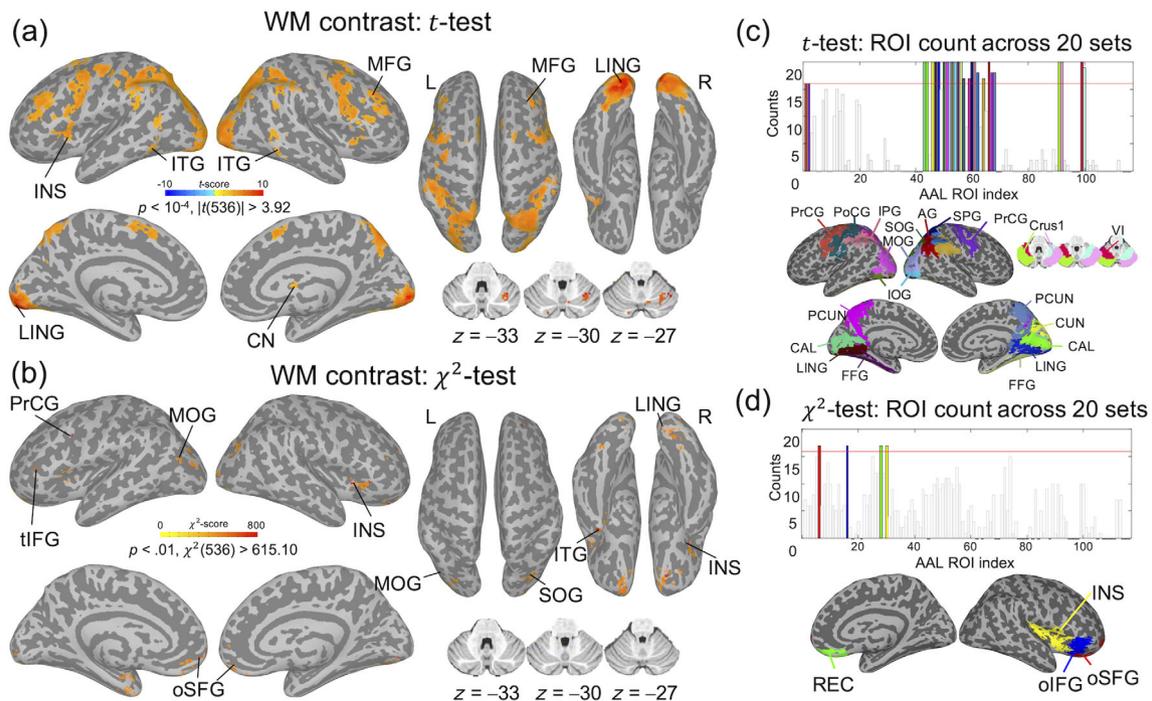
FIGURE 7 Regions-of-interest (ROIs) for the working memory (WM) task from the HCP tfMRI data: (a) Student's t-test ($p <10^{-4}$, $t[536]$ = 3.92) for working memory (WM) contrast (2-back > 0-back [correct > error]). (b) χ^2-test ($p <.01$, $\chi^2(536) >615.10$) for WM contrast (refer to Table S2 for more details). (c) The overlapped ROIs identified with t-test ($p <10^{-4}$, $t(268) >3.95$) across 20 holdout sets consisting of half of the total number of subjects; 27 regions (AAL atlas) were identified in over 80% of the sets. (d) The overlapped ROIs identified with χ^2-test ($p <.01$, $\chi^2(268) > 324.78$) across 20 different sets (refer to Table S3 for more details). The orbital area of the superior frontal gyrus, the orbital area of the inferior frontal gyrus, the rectus, and the insula were frequently identified in over 80% of the sets. The red line indicates 80% of all sets. L, left; R, right; AG, angular gyrus; CAL, calcarine; CN, caudate nucleus; CUN, cuneus; FFG, fusiform gyrus; tIFG, triangular part of inferior frontal gyrus; INS, insula; ITG, inferior temporal gyrus; IOG, inferior occipital gyrus; IPG, inferior parietal gyrus; LING, lingual gyrus; MFG, middle frontal gyrus; MOG, middle occipital gyrus; oIFG, orbital part of inferior frontal gyrus; oSFG, orbital part of superior frontal gyrus; SPG, superior parietal gyrus; SOG, superior occipital gyrus; PCUN, precuneus; PrCG, precentral gyrus; PoCG, postcentral gyrus; REC, rectal gyrus

and bilateral occipital gyrus (Figure 8d). Table S3 summarizes these ROIs.

3.6 | Association between brain and behavioral data identified from CCA

Table S2 summarizes the canonical correlation coefficients (CCs; ρ) for each of the reproducible ROIs across the four PC scenarios for the WM task. The canonical CCs in the training phase were almost all 1.0 when all of the voxels in the ROI were used, possibly due to overfitting. However, the canonical CCs decreased as the number of PCs was reduced. Figure 9 presents those CCA results that were consistently found to be significant in the test phase for at least three of the four PC scenarios. Using the t-test, the left superior parietal gyrus (SPG) and the right Crus 1 in the cerebellum exhibited significant positive correlations between the canonical variates of the neuronal activations and the canonical variates of the behavioral data for both the training and test phases. However, a significant association between the behavioral data and the canonical variates of the neuronal activations was found in only one of the four PC scenarios for the SPG, while no significant association was found for the Crus 1 in

the cerebellum. In addition, the corresponding positive/negative association was not consistent between the training and test phase. On the other hand, using the χ^2-test, the right oIFG and the right insula exhibited a significant correlation between the canonical variates of the neuronal activations and the canonical variates of the behavioral data. Moreover, both types of behavioral data showed a significant association (positive for accuracy and negative for RT across the training/test phases) with the canonical variates of the neuronal activations for both ROIs in three of the four PC scenarios ($p_{corrected} <.05$).

Table S3 summarizes the canonical CCs for each of the reproducible ROIs across the four PC scenarios for the GB task. Figure 10 illustrates the CCA results that were consistently found to be significant in the test phase for at least three of the four PC scenarios from the GB task. Using the t-test, the right orbital/medial area of the SFG, the right MFG, and the bilateral anterior cingulate cortex (ACC) showed significant positive correlations between the canonical variates of the neuronal activations and the canonical variates of the behavioral data for both the training and test data (Figure 10a). However, the association between both types of behavioral data and the canonical variates of the neuronal activations was not consistently found to be significant in the test data. On the other hand, using the χ^2-test, the left orbital area of the MFG (oMFG), the triangular area of
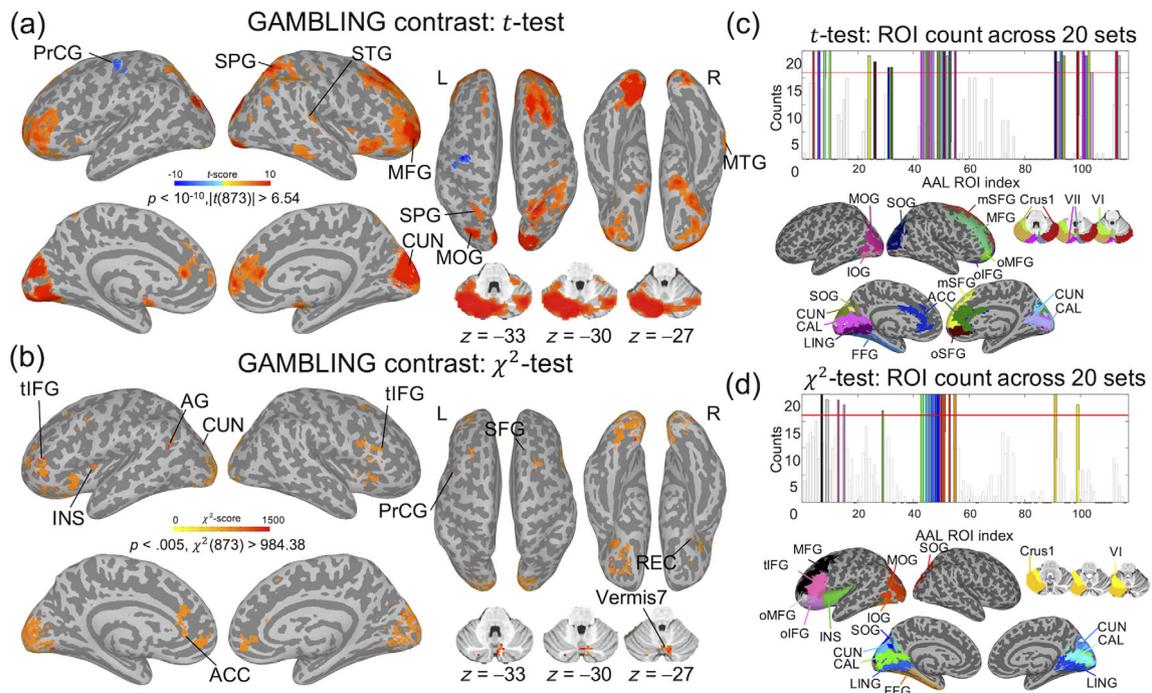
FIGURE 8 Regions-of-interest (ROIs) for the gambling (GB) task from the HCP tfMRI data. (a) Student's t-test ($p <10^{-10}$, $t[873] = 6.54$) for gambling contrast (*reward > punishment*). (b) χ^2-test ($p <.005$, $\chi^2(873) > 984.38$) for gambling contrast (refer to Table S3 for more details). (c) The overlapped ROIs identified with t-test ($p <10^{-10}$, $t[436] > 6.63$) across 20 different sets; 30 regions were identified in over 80% of the sets. (d) The overlapped ROIs identified with χ^2-test ($p <.005$, $\chi^2(436) > 515.81$) across 20 different sets (refer to Table S3 for more details); 18 regions were identified in over 80% of the sets. The red line indicates 80% of all trials. L, left; R, right; AG, angular gyrus; CAL, calcarine; CN, caudate nucleus; CUN, cuneus; FFG, fusiform gyrus; tIFG, triangular part of inferior frontal gyrus; INS, insula; ITG, inferior temporal gyrus; IOG, inferior occipital gyrus; IPG, inferior parietal gyrus; LING, lingual gyrus; MFG, middle frontal gyrus; MOG, middle occipital gyrus; oIFG, orbital part of inferior frontal gyrus; oMFG, orbital part of middle frontal gyrus; mSFG, medial superior frontal gyrus; oSFG, orbital part of superior frontal gyrus; SPG, superior parietal gyrus; SOG, superior occipital gyrus; PCUN, precuneus; PrCG, precentral gyrus; PoCG, postcentral gyrus

the IFG (tIFG), and the oIFG exhibited a significant association between at least one of the two types of behavioral data and the canonical variates of the neuronal activations for both the training and test phases (Figure 10b). The neuronal activations of two ROIs, tIFG, and oIFG, were negatively correlated with the RT and positively correlated with the contrast of the percentage across several PC scenarios ($p_{corrected}$ <.05). In contrast, the neuronal activations of the oMFG were positively correlated with the contrast of RT (i.e., $RT_{Reward} - RT_{Punishment}$) and negatively correlated with the contrast of the reward trial percentage (minus the punishment trial percentage) and across all PC scenarios ($p_{corrected}$ <.05).

4 | DISCUSSION

4.1 | Summary of the study

We presented an analytical pipeline using MEMA followed by CCA to investigate the relationship between brain data and associated behavioral data and demonstrated its efficacy in the context of nicotine craving for a small dataset. We designed an MRI-compatible e-cigarette apparatus that allowed us to deliver smoke with (ECIG) or without (SCIG) nicotine while fMRI data were acquired. The MEMA was applied

to the fMRI data for group inference and the ROIs associated with the experimental conditions (ECIG, SCIG, and ECIG vs. SCIG) were identified based on χ^2-tests and t-tests. The ROIs identified in this manner were further justified by applying CCA using the corresponding neuronal activations and the collected behavioral data. From the CCA results, the right insula was identified in the ECIG condition using the χ^2-tests but not from the t-tests. Furthermore, the neuronal activations of the right insula were significantly associated with the similarity and urge-to-smoke scores. From the contrast of the two conditions (ECIG > SCIG), the right OFC was identified only in the χ^2-tests, and the corresponding neuronal activations had a statistically meaningful association with the similarity and urge-to-smoke scores. We also demonstrated the efficacy of our analysis pipeline using tfMRI data acquired for the WM and GB tasks from the HCP for a large dataset. To our knowledge, our study is the first to show the efficacy of MEMA using real fMRI data in the context of nicotine craving, WM, and GB task processes.

4.2 | Utility of χ^2-tests and t-tests in mixed-effects multilevel analysis

In Equation (2), the Q score increases when the between-subject variability $(y_i - \mu)^2$ of the estimated effect sizes across the subjects starts
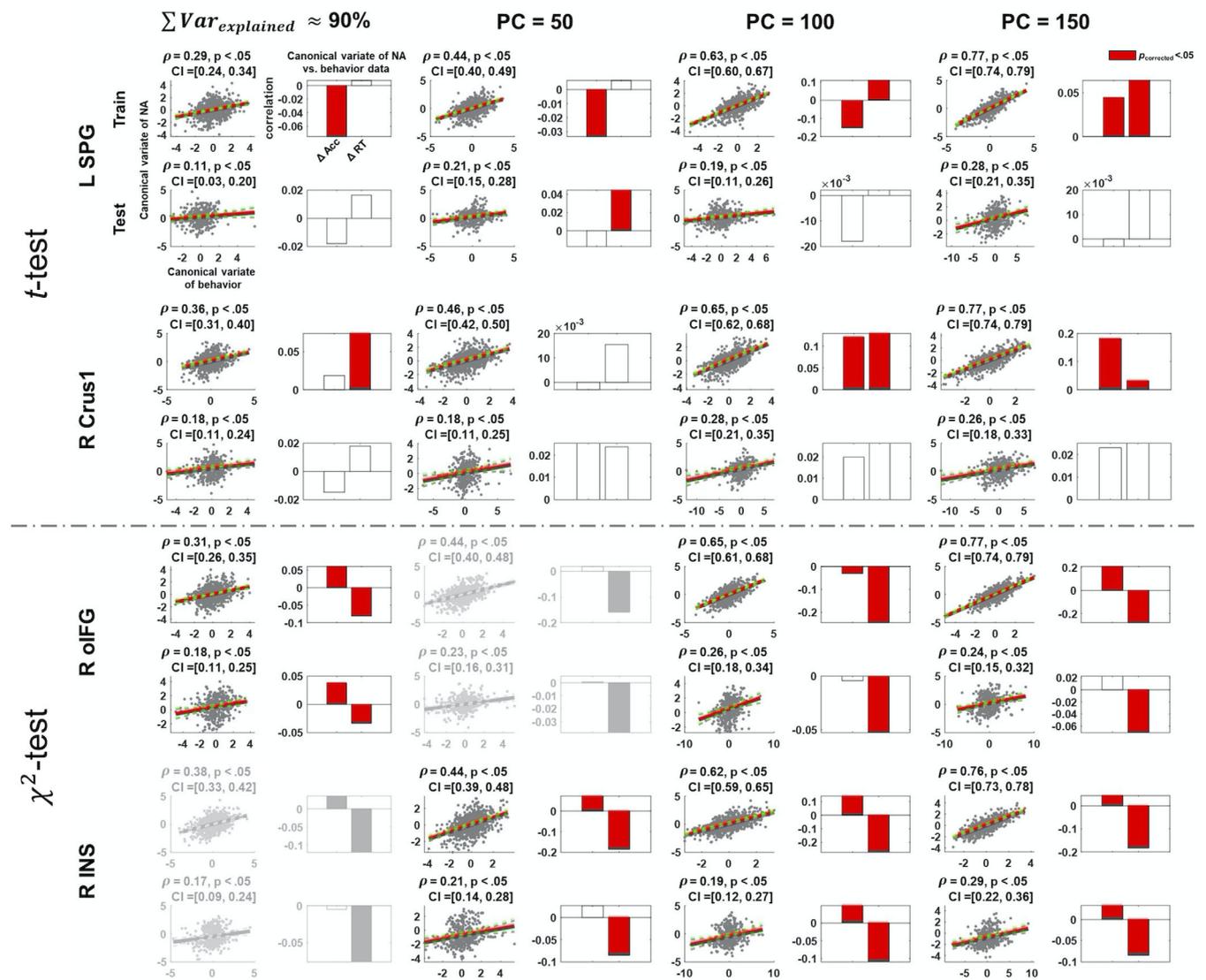
FIGURE 9 HCP working memory (WM) task. Correlation coefficients (a) between the canonical variates of the neuronal activations and the canonical variates of the behavioral data and (b) between the canonical variates of the neuronal activations and both types of behavioral data. Data from the training and test phases of the canonical correlation analysis (CCA) for each of the regions-of-interest (ROIs) that were identified in at least three of the four PC scenarios in the test phase using either t-tests or χ^2-tests are shown. The higher the number of PCs used in the CCA, the higher the canonical correlation coefficient (ρ), particularly for the training phase (red: regression line; green: 95% CI). The dark gray colored subplots indicate that the test ρ in over 90% of the 20 holdout sets were statistically significant (corrected $p < .05$ using 10,000 random permutations). The red bars indicate that the correlation coefficients between the canonical covariates of the neuronal activations and the corresponding behavioral data were significant (corrected $p < .05$ using 10,000 random permutations). Refer to the Section 3.6 for more detail. L, left; R, right; PC, principal component; CI, confidence interval; SPG, superior parietal gyrus; Crus1, cerebellum Crus 1; oIFG, orbital part of inferior frontal gyrus; INS, insula; ΔAcc, difference of accuracy (accuracy of 2-back − accuracy of 0-back); ΔRT, difference in the median reaction time (RT of 2-back − RT of 0-back)

to exceed the amount of within-subject variability σ_i^2 due to measurement error (Chen et al., 2012). Thus, the Q-statistic, which approximates the χ^2-distribution, is inherently advantageous in identifying ROIs that are linked to the indefinite, implicit, and idiosyncratic nature of the subjects (such as behavioral, psychological, and personal traits; for example, the similarity, urge-to-smoke, and smoking duration scores employed in the present study) that are consistent within a subject and vary across subjects. In fact, the Q-statistic has been

suggested as a valid approach for defining an ROI that could be used to find an association between a subject's BOLD responses and their behavioral measures, avoiding the problematic practice of defining ROIs based on the statistical significance of the activation strength alone (Lindquist et al., 2012). Based on our obtained results, the ROIs identified using the t-tests exhibited a substantially higher homogeneity in their beta values across subjects compared to the ROIs from the χ^2-tests (Figure 4). This suggests that the ROIs found using the t-tests
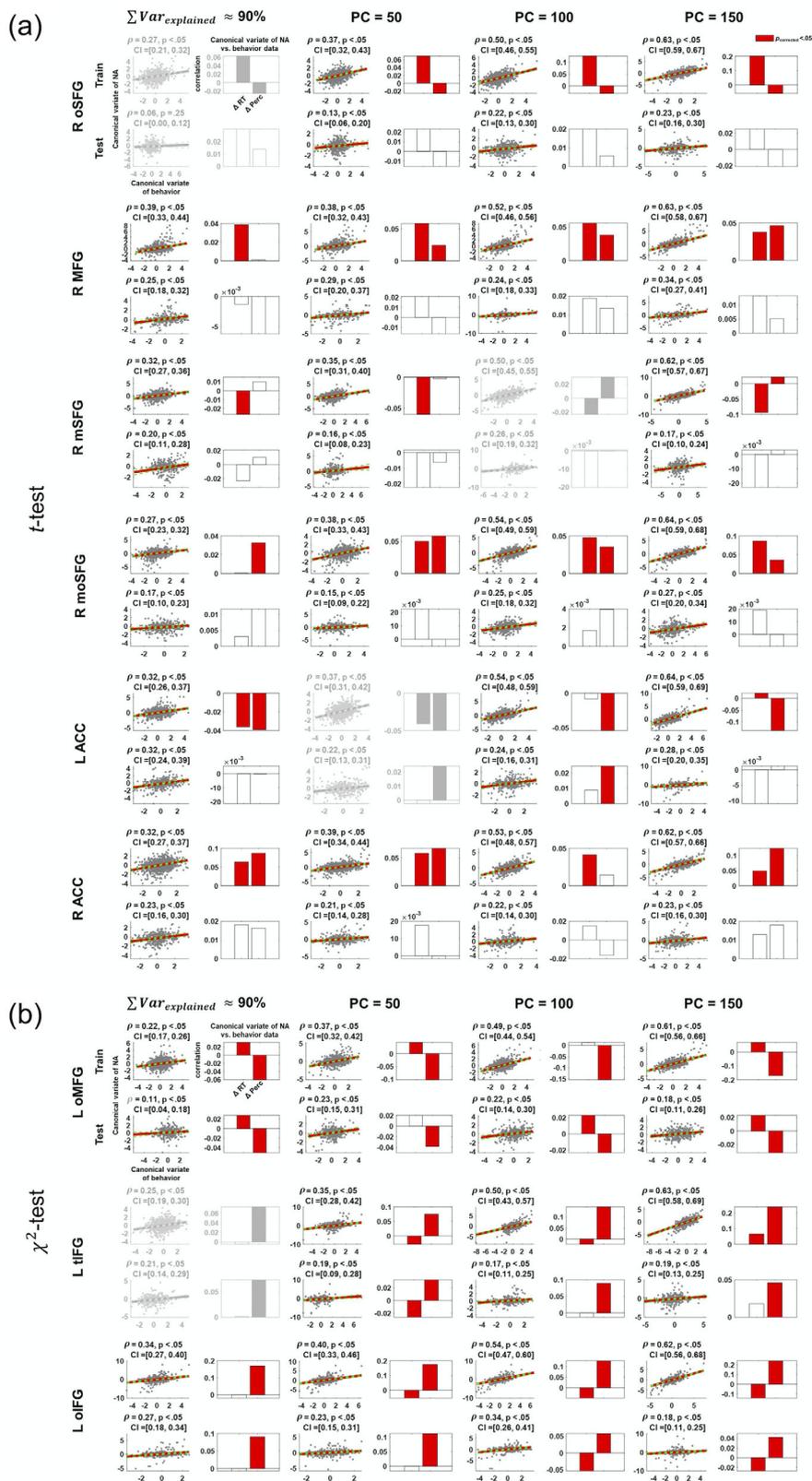
FIGURE 10 HCP gambling (GB) task. (a) CCA results obtained from Student's *t*-tests and (b) χ^2-tests. Refer to the legend for Figure 9 for a description of the figure and the Section 3.6 for an interpretation of the results. L, left; R, right; PC, principal component; CI, confidence interval; ACC, anterior cingulate cortex; moSFG, mid orbital part of superior frontal gyrus; ΔRT, difference of median reaction time ("RT of reward trials" − "RT of punishment trials"); ΔPerc, difference in the percentage of trials ("Percentage of reward trials" − "Percentage of punishment trials")

are advantageous in explaining the effects that are distinct, explicit, and common across all subjects (such as the experimental conditions ECIG and SCIG in the present study).

Consequently, we suggest that MEMA is a viable statistical approach to identifying ROIs that are associated with subjective

behavioral data and potentially alternative nonbrain data such as psychological data with the help of multivariate machine-learning approaches (e.g., CCA). Based on our experimental data on nicotine use, the ROIs identified from the *t*-test did not include the insula, and no ROI was significantly correlated with the behavioral data for

smoking. Interestingly, the ROIs identified based on the χ^2-test in the MEMA included the insula and OFC, which have previously been associated with nicotine craving (Adinoff, 2004; Fedota & Stein, 2015; Jasinska et al., 2014; Naqvi et al., 2007). In the present study, the corresponding neuronal activations were significantly correlated with the behavioral data related to nicotine craving.

The variability in neuronal activations of ROIs can be attributed to the heterogeneity of BOLD responses (Handwerker, Ollinger, & D'Esposito, 2004) or neurophysiological and/or neuropsychological conditions that are potentially associated with behavioral data, such as the degree of nicotine craving (Benowitz, 2010). The tobacco smoking of the participants prior to the MRI sessions was screened using CO levels. However, ensuring that the participants had abstained from e-cigarette use (and nicotine adsorption) prior to the experiment relied on the participants' verbal report without a systematic screening measure. Thus, it is possible that some of the participants did not abstain from e-cigarette smoking for at least 3 hr before the sessions. In addition, the number of hours that the participants abstained from smoking is likely to have differed, leading to variability in their behavioral data.

4.3 | CCA linking brain and behavioral data

The CCA has been instrumental in the interpretation of multimodal datasets, such as analyzing the positive and negative modes of population covariation by linking brain connectivity patterns with corresponding demographic and behavioral information available in the HCP (Smith et al., 2015). In a more recent article, CCA has been suggested as an effective tool for investigating the relationship between brain and behavioral data with optional dimension-reduction steps, such as using independent component analysis and PCA (Wang et al., 2020). In our study, CCA was successfully used to identify links between the neuronal activations of the ROIs and the collected behavior data via the multivariate transformation of both datasets. It is worth noting, however, that meaningful associations were identified from the ROIs obtained from the statistical analysis, enabling brain regions with high inter-subject variability in their neuronal activations in comparison to intra-subject variability (i.e., χ^2-tests in the MEMA) to be identified. On the other hand, this was not the case for conventional statistical analysis (the t-tests in the present study), which focuses on identifying brain regions that are highly associated with experimental conditions across subjects with no explicit consideration of individual-specific variability.

4.4 | Neuronal activations of the insula as a predictor of the level of nicotine craving

We found that the neuronal activations of the right insula were significantly associated with the ECIG condition but not with the SCIG condition. In addition, the level of neuronal activations of the right insula was highly associated with the degree of subjective craving measured using the similarity and urge-to-smoke scores. This is in line with previous reports that the insula is one of the core brain regions associated with nicotine craving (Morales, Ghahremani, Kohno, Hellemann, & London, 2014; Naqvi et al., 2007; Suñer-Soler et al., 2012). For example, it has been reported that nicotine dependence/craving is negatively correlated with the cortical thickness of the right ventral anterior insula (Morales et al., 2014). In addition, the degree of insular damage in stroke patients was positively correlated with success in quitting smoking (Suñer-Soler et al., 2012). Another study reported that people with brain lesions in the insula could quit smoking without reporting conscious urges (Naqvi et al., 2007). In light of these studies, our results suggest that the neuronal activations in the right insula can be used as a predictor to estimate the degree of nicotine craving.

4.5 | Neuronal activations of the orbitofrontal cortex as a proxy for throat-hit

We also found that the neuronal activations in the right OFC were significantly associated with the similarity and urge-to-smoke scores. The OFC is known to be part of the striato-thalamo-orbitofrontal circuit, a dopamine reward pathway that projects from the nucleus accumbens and ventral tegmental area to the prefrontal cortex (PFC) and the amygdala (Volkow & Fowler, 2000). In addition, it has been suggested that the olfactory tubercle, which receives direct input from the olfactory bulb, is activated by the use of a vaporizing device and the subsequent throat-hit (Johnson, 2020). The throat-hit is the burning sensation in the throat caused by the nicotine in smoke when it is inhaled. Due to the large size of the right OFC cluster found in our study, we believe that the olfactory tubercle is included in this cluster because these two areas are anatomically adjacent (Rolls, 2000). Another study reported that high plasma nicotine levels were associated with craving relief, increased satisfaction, and a stronger throat-hit (Etter, 2016; Farsalinos et al., 2015). In light of these studies, we conjecture that the neuronal activations in the right OFC were part of a reward response triggered by the throat-hit when using our e-cigarette apparatus in the ECIG condition, a response that was not present in the SCIG condition due to the absence of nicotine. Interestingly, the OFC was activated when there was insufficient nicotine (represented by a higher urge-to-smoke), leading to a weaker throat-hit response, as represented by a lower similarity (Adinoff, 2004).

4.6 | Data from the debriefing sessions

Based on the verbal reports from the participants after the experiments, most were not aware of the nature of the ECIG and SCIG blocks in the experimental paradigm. Four subjects in the non-MRI session reported that there was a difference between the experimental blocks, whereas one subject in the MRI session appeared to recognize which of the blocks contained nicotine. Interestingly, despite the

fact that most participants were unaware of the nature of each block, their ratings for the similarity, urge-to-smoke, and smoking duration differed significantly between the ECIG and SCIG blocks (Figure 2). This may be because the nicotine in the ECIG inherently alleviated their nicotine craving and caused the participants to feel a greater similarity with their e-cigarette smoking and to extend the smoking duration. At the debriefing session, the participants reported that they assigned a high similarity score if they strongly felt the throat-hit and that they scored their urge to smoke based on how satiated they felt after inhaling/exhaling the smoke and experiencing the lightheadedness associated with smoking.

4.7 | MRI-compatible e-cigarette smoking apparatus

As far as we are aware, there has been no prior research investigating the direct smoking of recent-generation e-cigarettes inside an MRI due to the metallic components found within e-cigarette devices. A previous study did use first-generation e-cigarettes (Wall et al., 2017), but the amount of nicotine delivered by these e-cigarettes was substantially lower than that delivered by second generation e-cigarettes (Farsalinos et al., 2015; Wall et al., 2017). In order to circumvent the MR incompatibility of e-cigarettes while delivering a substantial amount of nicotine, we designed an MRI-compatible e-cigarette apparatus and used a second generation e-cigarette to markedly increase the amount of nicotine in the heated smoke generated from the e-liquid (Farsalinos et al., 2015). Our finding that the insula was highly correlated with the urge-to-smoke scores suggests that our smoking apparatus using a second generation e-cigarette successfully delivered nicotine to the brain (Farsalinos et al., 2018).

4.8 | Limitations of the study using e-cigarette data and further work

The number of subjects ($n = 18$) who participated in the MRI sessions was comparatively low in relation to recent neuroimaging studies. However, we looked to overcome this potential limitation by employing an LOOCV framework during the CCA (Dinga et al., 2019; Le Floch et al., 2012). It is also worth noting that the statistical power was not sufficiently large (>0.8) (Hintze, 2008; Suresh & Chandrashekara, 2012) if the uncorrected p-value was not stringent. In our study, when the uncorrected p-value was .005 for the t-test, the statistical power was .64 and this power increased to .80 when the uncorrected p-value was reduced to .0005. In the clusters remaining from the cluster-based multiple comparison corrections, both the right OFC and the left middle temporal gyrus were identified from an uncorrected p-value of .0005; however, the right temporal gyrus disappeared when the power was >0.8. The importance of stringent uncorrected p-values in enhancing the statistical power has also been stressed in the recent debate on false-positive errors obtained from cluster-based multiple comparison correction

(Cox, Chen, Glen, Reynolds, & Taylor, 2017; Eklund, Nichols, & Knutsson, 2016; Flandin & Friston, 2019).

All of the recruited participants were male, which may have led to potential bias in the identified ROIs from the MEMA and subsequent CCA based on recent reports of distinct functional networks between genders (Clemens et al., 2020; Weis et al., 2020). The use of data from independent subjects that include females would thus provide an unbiased evaluation of our presented approach. Finally, in addition to the insula and OFC, other ROIs are potentially associated with alternative behavioral and/or psychological data that represent the idiosyncratic nature of the participants. Whole-brain functional connectivity analysis (Bastos & Schoffelen, 2016; Rosenberg et al., 2016) may also provide additional information on the functional role of the ROIs.

4.9 | Evaluation of our approach using a larger sample dataset: Working memory task from the HCP

Using WM tfMRI data from the HCP, we found that the neuronal activations of the right oIFG and insula identified from the χ^2-test were positively associated with accuracy and negatively associated with RT. A number of studies have reported the involvement of the PFC in the WM task (Barbey, Koenigs, & Grafman, 2011, 2013; Levens & Phelps, 2010; Miller, Lundqvist, & Bastos, 2018; Wager & Smith, 2003; Yaple, Dale Stevens, & Arsalidou, 2019). In addition, the OFC and insula have been proposed to have a functional role in executive processing during the WM task (Barbey et al., 2011; Levens & Phelps, 2010; Owen, McMillan, Laird, & Bullmore, 2005; Yaple et al., 2019). Our findings are thus in line with previous reports, providing evidence for the executive function of the OFC (Barbey et al., 2011; Schuck, Cai, Wilson, & Niv, 2016) and the insula (Uddin, Nomi, Hébert-Seropian, Ghaziri, & Boucher, 2017), with the neuronal activations in the right oIFG and the insula in our study positively correlated with accuracy from the higher working memory load (i.e., 2-back > 0-back) and negatively correlated with RT. The OFC and insula have also been reported to be part of task-positive networks in which the corresponding neuronal activations increase as the subjects put more effort into a task (Basten, Stelzel, & Fiebach, 2013; Fox et al., 2005). However, the left SPG and cerebellum Crus 1 regions identified from the t-test did not show any consistent association between their neuronal activations and the behavioral data for either the training or test phase (Figure 9). Thus, χ^2-test is more appropriate for identifying ROIs when investigating the brain and behavioral relationship when the CCA is applied (Smith et al., 2015).

4.10 | Evaluation of our approach for a larger sample dataset: Gambling task from the HCP

Using the GB tfMRI data, the neuronal activations of the PFC area including the left oMFG, tIFG, and oIFG were identified from the χ^2-test based on the contrast of the reward versus punishment

conditions. Moreover, the corresponding neuronal activations were significantly associated with the percentage of reward trials in comparison to punishment trials and the corresponding RT. A number of studies have reported a significant association between the ventromedial area of the PFC (vmPFC) and monetary reward (Aharon et al., 2001; Elliott, Friston, & Dolan, 2000; Liu, Hairston, Schrier, & Fan, 2011; O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001; Schneider, Leuchs, Czisch, Sämann, & Spoormaker, 2018) and between the lateral OFC regions and punishment (Elliott, Dolan, & Frith, 2000; Lopez-Persem et al., 2020; O'Doherty et al., 2001; Schoenbaum, Takahashi, Liu, & McDannald, 2011). The reproducible ROIs identified from the t-test did not exhibit any meaningful association between the corresponding neuronal activations and the behavioral data.

The neuronal activations in the left oMFG were positively correlated with RT and negatively with the percentage of reward trials; however, this trend was reversed for the left tIFG and the left oIFG (Figure 10b). This suggests that the functional roles of the oMFG and tIFG/oIFG are distinct from each other. Considering the nature of reward and punishment trials, the participants must not have properly reasoned the reward/punishment of the subsequent trials in the experiment. Nevertheless, the participants may have relied on (a) guessing/instinct based on their gut feeling or (b) their reasoning to predict the number on the card based on previous trials. We believe these two types of brain process coexisted in their brain functions. In this context, based on our findings, the left oMFG is perhaps more relevant to the cognitive process of reasoning, while the left tIFG/oIFG may be associated with guessing.

4.11 | Number of voxels to include in an ROI identified using the cluster-extend threshold

In the analysis using our e-cigarette dataset, we used seven neighboring voxels including a center voxel in the multivoxel pattern for the CCA due to the limited number of participants ($n = 18$). In order to minimize the potential utilization of part of the voxel cluster (Woo, Krishnan, & Wager, 2014), we used stringent p-values of .001 for the t-test and 10^{-8} for the χ^2-test to select the candidate voxel clusters. In the application of our method to the HCP dataset, we considered all voxels in each of the identified clusters for the multivoxel patterns. Because the number of participants was not substantially larger than the number of voxels in the clusters, we adopted a dimension-reduction approach using PCA to alleviate the overfitting associated with CCA (Smith et al., 2015). The CCA in the training phase exhibited virtually perfect canonical CCs (i.e., 1.0) when all of the voxels in the ROI were used, indicating possible overfitting. When using reduced dimensions for the multivoxel patterns, the CCs in the training phase were lower but statistically significant (Tables S2 and S3). Furthermore, the CCs for some of the ROIs maintained their statistical significance in the test phase, particularly for the reduced dimensions (i.e., 50, 100, and 150 PCs) that were greater than the cluster size threshold (i.e., 40). Our findings suggest that our approach is suitable

for both small and large datasets that appropriately consider the number of voxels to include in identified ROIs based on the cluster-extent threshold (Woo et al., 2014).

4.12 | Candidate behavioral data for CCA

We obtained behavioral data from the participants simultaneously with the fMRI data (i.e., similarity, urge-to-smoke, and smoking duration for the e-cigarette data; response time and accuracy/percentage of trials for the HCP data). This was based on our belief that behavioral data obtained during the tfMRI run would be directly associated with the heterogeneity of the neuronal activations across the participants estimated from the tfMRI data. In addition, the HCP dataset also provides 26 subjective items collected from each participant related to their mental and physical processes to characterize individual traits (https://wiki.humanconnectome.org/display/Public Data/HCP-YA+Data+Dictionary-+Updated+for+the+1200+Subject +Release). We also analyzed these 26 items along with the RT and accuracy/percentage in the CCA phase using the ROIs from the MEMA and found preliminary evidence for an interesting relationship between the behavioral data and some of the ROIs identified using the χ^2-test (data not shown). Future work could provide a fine-grained interpretation of the association of the identified ROIs with individual mental/cognitive and physical traits and their stability across the permuted sets of training and test folds.

5 | CONCLUSION

In the present study, we reported the utility of MEMA followed by CCA to investigate the relationship between neuronal activations from fMRI and associated behavioral data. To evaluate our hypothesis as a proof-of-concept using a real dataset, we designed an MRI-compatible e-cigarette apparatus in order to carefully deliver smoke with or without nicotine. fMRI data and behavioral data associated with the smoking experience when using the apparatus were subsequently taken from the participants. Although further investigation is warranted due to the limited number of subjects ($n = 18$), the right insula and right OFC, which are associated with the nicotine-craving pathway, were identified with strong statistical evidence using the χ^2-test from the MEMA but not using conventional t-test. Furthermore, CCA revealed that the BOLD fMRI responses of these ROIs were significantly associated with the degree of nicotine craving and the throat-hit. Our code and sample data are publicly available (https:// github.com/bsplku/MEMA-CCA). Both the code and data comply with the requirements of our funding bodies, institution, and institutional ethics committee. The analytical results of our method obtained from the HCP dataset also suggest that χ^2-test is more appropriate than t-test for modeling the association between the brain and behavioral data across subjects. We believe that our presented analytical framework is potentially useful for a range of basic neuroscientific research and clinical studies for the identification of functional brain regions

from fMRI data whose neuronal activations are associated with the idiosyncrasies of the participants.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

ORCID

Jong-Hwan Lee https://orcid.org/0000-0002-8902-6009

REFERENCES

Adinoff, B. (2004). Neurobiologic processes in drug reward and addiction. *Harvard Review of Psychiatry*, *12*(6), 305–320. https://doi.org/10.1080/10673220490910844

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: FMRI and behavioral evidence. *Neuron*, *32*(3), 537–551.

Anders, S., Lotze, M., Erb, M., Grodd, W., & Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: A response-related FMRI study. *Human Brain Mapping*, *23*(4), 200–209. https://doi.org/10.1002/hbm.20048

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, *145*, 137–165. https://doi.org/10.1016/j.neuroimage.2016.02.079

Barbey, A. K., Koenigs, M., & Grafman, J. (2011). Orbitofrontal contributions to human working memory. *Cerebral Cortex*, *21*(4), 789–795.

Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *Cortex*, *49*(5), 1195–1205.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., ... Feldt, C. (2013). Function in the human connectome: Task-FMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189.

Basten, U., Stelzel, C., & Fiebach, C. J. (2013). Intelligence is differentially related to neural effort in the task-positive and the task-negative brain network. *Intelligence*, *41*(5), 517–528. https://doi.org/10.1016/j.intell.2013.07.006

Bastos, A. M., & Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in Systems Neuroscience*, *9*, 175.

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the beck depression inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, *8*(1), 77–100. https://doi.org/10.1016/0272-7358(88)90050-5

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *NeuroImage*, *20*(2), 1052–1063. https://doi.org/10.1016/S1053-8119(03)00435-X

Benowitz, N. L. (2010). Nicotine addiction. *New England Journal of Medicine*, *362*(24), 2295–2303.

Chen, G., Saad, Z. S., Nath, A. R., Beauchamp, M. S., & Cox, R. W. (2012). FMRI group analysis combining effect estimates and their variances. *NeuroImage*, *60*(1), 747–765. https://doi.org/10.1016/j.neuroimage.2011.12.060

Clark, V. P. (2002). Orthogonal polynomial regression for the detection of response variability in event-related FMRI. *NeuroImage*, *17*(1), 344–363. https://doi.org/10.1006/nimg.2002.1100

Clemens, B., Derntl, B., Smith, E., Junger, J., Neulen, J., Mingoia, G., ... Habel, U. (2020). Predictive pattern classification can distinguish gender identity subtypes from behavior and brain imaging. *Cerebral Cortex*, *30*(5), 2755–2765.

Cochran, W. G. (1954). Some methods for strengthening the common χ2 tests. *Biometrics*, *10*(4), 417–451.

Cohen, S., Kamarck, T., & Mermelstein, R. (1994). Perceived stress scale. In *Measuring stress: A guide for health and social scientists* (Vol. 10, pp. 1–2). Menlo Park, CA: Mind Garden, Inc.

Cox, L. S., Tiffany, S. T., & Christen, A. G. (2001). Evaluation of the brief questionnaire of smoking urges (QSU-brief) in laboratory and clinical settings. *Nicotine & Tobacco Research*, *3*(1), 7–16. https://doi.org/10.1080/14622200020032051

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). FMRI clustering in AFNI: False-positive rates redux. *Brain Connectivity*, *7*(3), 152–171.

Cui, Z., & Gong, G. (2018). The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage*, *178*, 622–637. https://doi.org/10.1016/j.neuroimage.2018.06.001

de Hollander, G., Forstmann, B. U., & Brown, S. D. (2016). Different ways of linking behavioral and neural data via computational cognitive models. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(2), 101–109. https://doi.org/10.1016/j.bpsc.2015.11.004

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188.

Dinga, R., Schmaal, L., Penninx, B. W. J. H., van Tol, M. J., Veltman, D. J., van Velzen, L., ... Marquand, A. F. (2019). Evaluating the evidence for biotypes of depression: Methodological replication and extension of. *NeuroImage: Clinical*, *22*, 101796. https://doi.org/10.1016/j.nicl.2019.101796

Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., ... Lessov-Schlaggar, C. N. (2010). Prediction of individual brain maturity using FMRI. *Science*, *329*(5997), 1358–1361.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why FMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex*, *10*(3), 308–317.

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *Journal of Neuroscience*, *20*(16), 6159–6165.

Esterman, M., Tamber-Rosenau, B. J., Chiu, Y.-C., & Yantis, S. (2010). Avoiding non-independence in FMRI data analysis: Leave one subject out. *NeuroImage*, *50*(2), 572–576.

Etter, J.-F. (2016). Throat hit in users of the electronic cigarette: An exploratory study. *Psychology of Addictive Behaviors*, *30*(1), 93–100. https://doi.org/10.1037/adb0000137

Farsalinos, K. E., Spyrou, A., Tsimopoulou, K., Stefopoulos, C., Romagna, G., & Voudris, V. (2015). Nicotine absorption from electronic cigarette use: Comparison between first and new-generation devices. *Scientific Reports*, *4*(1), 4133. https://doi.org/10.1038/srep04133

Farsalinos, K. E., Yannovits, N., Sarri, T., Voudris, V., & Poulas, K. (2018). Nicotine delivery to the aerosol of a heat-not-burn tobacco product: Comparison with a tobacco cigarette and e-cigarettes. *Nicotine and Tobacco Research*, *20*(8), 1004–1009.

Fedota, J. R., & Stein, E. A. (2015). Resting-state functional connectivity and nicotine addiction: Prospects for biomarker development: Resting-

state functional connectivity and nicotine. *Annals of the new York Academy of Sciences*, 1349(1), 64–82. https://doi.org/10.1111/nyas.12882

Flandin, G., & Friston, K. J. (2019). Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Human Brain Mapping*, 40(7), 2052–2054.

Foulds, J., Veldheer, S., Yingst, J., Hrabovsky, S., Wilson, S. J., Nichols, T. T., & Eissenberg, T. (2015). Development of a questionnaire for assessing dependence on electronic cigarettes among a large sample of ex-smoking E-cigarette users. *Nicotine & Tobacco Research*, 17(2), 186–192. https://doi.org/10.1093/ntr/ntu204

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, Anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27), 9673–9678.

Franklin, T. R., Wang, J., Sciortino, N., Harper, D., Li, Y., Ehrman, R., ... Childress, A. R. (2007). Limbic activation to cigarette smoking cues independent of nicotine withdrawal: A perfusion FMRI study. *Neuropsychopharmacology*, 32(11), 2301–2309.

Fregni, F., Fecteau, S., & Pascual-Leone, A. (2008). Cortical stimulation of the prefrontal cortex with transcranial direct current stimulation reduces Cue-provoked smoking craving: A randomized, sham-controlled study. *Journal of Clinical Psychiatry*, 69(1), 32–40.

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage*, 16(2), 465–483. https://doi.org/10.1006/nimg.2002.1090

Gabrieli, J. D. E., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, 85(1), 11–26.

Ge, R., Jin, C., Netrapalli, P., & Sidford, A.. (2016). Efficient algorithms for large-scale generalized eigenvector computation and canonical correlation analysis. *International Conference on Machine Learning*. pp. 2741–2750.

Gilbert, D. G., Joseph McClernon, F., Rabinovich, N. E., Plath, L. C., Jensen, R. A., & Meliska, C. J. (1998). Effects of smoking abstinence on mood and craving in men: Influences of negative-affect-related personality traits, habitual nicotine intake and repeated measurements. *Personality and Individual Differences*, 25(3), 399–423.

Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21(4), 1639–1651. https://doi.org/10.1016/j.neuroimage.2003.11.029

Haynes, J.-D. (2015). A primer on pattern-based approaches to FMRI: Principles, pitfalls, and perspectives. *Neuron*, 87(2), 257–270.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K.-O. (1991). The Fagerstrom test for nicotine dependence: A revision of the Fagerstrom tolerance questionnaire. *Addiction*, 86(9), 1119–1127. https://doi.org/10.1111/j.1360-0443.1991.tb01879.x

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. https://doi.org/10.1002/sim.1186

Hintze, J. L. (2008). Power Analysis and Sample Size System (PASS) for windows user's guide I. NCSS: Kaysville, UT.

Hobkirk, A. L., Nichols, T. T., Foulds, J., Yingst, J. M., Veldheer, S., Hrabovsky, S., ... Wilson, S. J. (2018). Changes in resting state functional brain connectivity and withdrawal symptoms are associated with acute electronic cigarette use. *Brain Research Bulletin*, 138, 56–63. https://doi.org/10.1016/j.brainresbull.2017.05.010

Hotelling, H. (1935). Canonical correlation analysis (CCA). *Journal of Educational Psychology*, 10.

Hubbard, E. M., Cyrus Arman, A., Ramachandran, V. S., & Boynton, G. M. (2005). Individual differences among grapheme-color Synesthetes: Brain-behavior correlations. *Neuron*, 45(6), 975–985.

Hughes, J. R., & Hatsukami, D. (1986). Signs and symptoms of tobacco withdrawal. *Archives of General Psychiatry*, 43(3), 289–294.

Jasinska, A. J., Zorick, T., Brody, A. L., & Stein, E. A. (2014). Dual role of nicotine in addiction and cognition: A review of neuroimaging studies in humans. *Neuropharmacology*, 84, 111–122.

Jo, H. J., Saad, Z. S., Kyle Simmons, W., Milbury, L. A., & Cox, R. W. (2010). Mapping sources of correlation in resting state FMRI, with artifact detection and removal. *NeuroImage*, 52(2), 571–582. https://doi.org/10.1016/j.neuroimage.2010.04.246

Johnson, Bankole A. (2020). *Addiction medicine: Science and practice*.

Julian, L. J. (2011). Measures of Anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). *Arthritis Care & Research*, 63(S11), S467–S472. https://doi.org/10.1002/acr.20561

Kim, D.-Y., Yoo, S.-S., Tegethoff, M., Meinlschmidt, G., & Lee, J.-H. (2015). The inclusion of functional connectivity information into FMRI-based neurofeedback improves its efficacy in the reduction of cigarette cravings. *Journal of Cognitive Neuroscience*, 27(8), 1552–1572. https://doi.org/10.1162/jocn_a_00802

Kim, H.-C., Tegethoff, M., Meinlschmidt, G., Stalujanis, E., Belardi, A., Jo, S., ... Lee, J.-H. (2019). Mediation analysis of triple networks revealed functional feature of mindfulness from real-time FMRI neurofeedback. *NeuroImage*, 195, 409–432.

Kohn, A., Coen-Cagli, R., Kanitscheider, I., & Pouget, A. (2016). Correlations and neuronal population information. *Annual Review of Neuroscience*, 39, 237–256.

Lamichhane, B., Westbrook, A., Cole, M. W., & Braver, T. (2020). Exploring brain-behavior relationships in the N-Back task. *NeuroImage*, 212, 116683.

Le Floch, E., Édith, V. G., Frouin, V., Pinel, P., Lalanne, C., Trinchera, L., ... Duchesnay, É. (2012). Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse partial least squares. *NeuroImage*, 63(1), 11–24. https://doi.org/10.1016/j.neuroimage.2012.06.061

Lee, J.-H., Kim, D.-Y., & Kim, J. (2012). Mesocorticolimbic hyperactivity of deprived smokers and brain imaging. *Neuroreport*, 23(18), 1039–1043.

Levens, S. M., & Phelps, E. A. (2010). Insula and orbital frontal cortex activity underlying emotion interference resolution in working memory. *Journal of Cognitive Neuroscience*, 22(12), 2790–2803. https://doi.org/10.1162/jocn.2010.21428

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lindquist, M. A. (2008). The statistical analysis of FMRI data. *Statistical Science*, 23(4), 439–464. https://doi.org/10.1214/09-STS282

Lindquist, M. A., Spicer, J., Asllani, I., & Wager, T. D. (2012). Estimating and testing variance components in a multi-level GLM. *NeuroImage*, 59(1), 490–501. https://doi.org/10.1016/j.neuroimage.2011.07.077

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 35(5), 1219–1236.

Lopez-Persem, A., Roumazeilles, L., Folloni, D., Marche, K., Fouragnan, E. F., Khalighinejad, N., ... Sallet, J. (2020). Differential functional connectivity underlying asymmetric reward-related activity in human and nonhuman primates. *Proceedings of the National Academy of Sciences*, 117(45), 28452–28462.

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working memory 2.0. *Neuron*, 100(2), 463–475. https://doi.org/10.1016/j.neuron.2018.09.023

Morales, A. M., Ghahremani, D., Kohno, M., Hellemann, G. S., & London, E. D. (2014). Cigarette exposure, dependence, and craving are related to insula thickness in young adult smokers. *Neuropsychopharmacology*, 39(8), 1816–1822.

Moran, L. V., Stoeckel, L. E., Wang, K., Caine, C. E., Villafuerte, R., Calderon, V., ... Pizzagalli, D. A. (2018). Nicotine increases activation to

anticipatory valence cues in anterior insula and striatum. *Nicotine and Tobacco Research*, 20(7), 851–858.

Mumford, J. A., & Nichols, T. (2009). Simple group FMRI modeling and inference. *NeuroImage*, 47(4), 1469–1475. https://doi.org/10.1016/j.neuroimage.2009.05.034

Naqvi, N. H., Rudrauf, D., Damasio, H., & Bechara, A. (2007). Damage to the insula disrupts addiction to cigarette smoking. *Science*, 315(5811), 531–534.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1), 1–25. https://doi.org/10.1002/hbm.1058

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.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, 4(1), 95–102.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1), 46–59.

Penny, W., & Holmes, A. (2007). Random effects analysis. In *Statistical parametric mapping: The analysis of functional brain images* (pp. 156–165). Amsterdam, The Netherlands: Elsevier.

Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust correlation analyses: False positive and Power validation using a new open source Matlab toolbox. *Frontiers in Psychology*, 3. https://doi.org/10.3389/fpsyg.2012.00606

Power, J. D., Silver, B. M., Silverman, M. R., Ajodan, E. L., Bos, D. J., & Jones, R. M. (2019). Customized head molds reduce motion during resting state fMRI scans. *NeuroImage*, 189, 141–149. http://doi.org/10.1016/j.neuroimage.2019.01.016

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex*, 10(3), 284–294. https://doi.org/10.1093/cercor/10.3.284

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Todd Constable, R., & Chun, M. M. (2016). A Neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, 19(1), 165–171.

Rousselet, G. A., & Pernet, C. R. (2012). Improving standards in brain-behavior correlation analyses. *Frontiers in Human Neuroscience*, 6, 119.

Safo, S. E., Ahn, J., Jeon, Y., & Jung, S. (2018). Sparse generalized eigenvalue problem with application to canonical correlation analysis for integrative analysis of methylation and gene expression data. *Biometrics*, 74(4), 1362–1371.

Schneider, M., Leuchs, L., Czisch, M., Sämann, P. G., & Spoormaker, V. I. (2018). Disentangling reward anticipation with simultaneous Pupillometry/FMRI. *NeuroImage*, 178, 11–22.

Schoenbaum, G., Takahashi, Y., Liu, T.-L., & McDannald, M. A. (2011). Does the orbitofrontal cortex signal value? *Annals of the new York Academy of Sciences*, 1239, 87.

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human orbitofrontal cortex represents a cognitive map of state space. *Neuron*, 91(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Schuh, K. J., & Stitzer, M. L. (1995). Desire to smoke during spaced smoking intervals. *Psychopharmacology*, 120(3), 289–295. https://doi.org/10.1007/BF02311176

Seo, S. (2006). *A review and comparison of methods for detecting outliers in univariate data sets*. University of Pittsburgh.

Shiffman, S. M., & Jarvik, M. E. (1976). Smoking withdrawal symptoms in two weeks of abstinence. *Psychopharmacology*, 50(1), 35–39. https://doi.org/10.1007/BF00634151

Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E. J., Glasser, M. F., ... Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, 18(11), 1565–1567.

Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true Intraindividual change: True change as a latent variable. *Methods of Psychological Research Online*, 2(1), 21–33.

Sui, J., Qi, S., van Erp, T. G. M., Bustillo, J., Jiang, R., Lin, D., et al. (2018). Multimodal neuromarkers in schizophrenia via cognition-guided MRI fusion. *Nature Communications*, 9(1), 1–14.

Suñer-Soler, R., Grau, A., Gras, M. E., Font-Mayolas, S., Silva, Y., Dávalos, A., ... Serena, J. (2012). Smoking cessation 1 year Poststroke and damage to the insular cortex. *Stroke*, 43(1), 131–136.

Suresh, K. P., & Chandrashekara, S. (2012). Sample size estimation and Power analysis for clinical research studies. *Journal of Human Reproductive Sciences*, 5(1), 7.

Ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67(5), 1167–1179.

Terhune, D. B., Russo, S., Near, J., Stagg, C. J., & Kadosh, R. C. (2014). GABA predicts time perception. *Journal of Neuroscience*, 34(12), 4364–4370.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical Parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Uddin, L. Q., Nomi, J. S., Hébert-Seropian, B., Ghaziri, J., & Boucher, O. (2017). Structure and function of the human insula. *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, 34(4), 300.

Valente, G., Castellanos, A. L., Vanacore, G., & Formisano, E. (2014). Multivariate linear regression of high-dimensional FMRI data with multiple target variables: MLR of high-dimensional FMRI data. *Human Brain Mapping*, 35(5), 2163–2177. https://doi.org/10.1002/hbm.22318

Van Essen, D. C., David, C., Ugurbil, K., Auerbach, E., Deanna, B., Behrens, T. E. J., ... Curtiss, S. W. (2012). The human connectome project: A data acquisition perspective. *NeuroImage*, 62(4), 2222–2231.

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, 145, 166–179. https://doi.org/10.1016/j.neuroimage.2016.10.038

Viechtbauer, W. (2010). Conducting meta-analyses in R with the Metafor package. *Journal of Statistical Software*, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Volkow, N. D., & Fowler, J. S. (2000). Addiction, a disease of compulsion and drive: Involvement of the orbitofrontal cortex. *Cerebral Cortex*, 10(3), 318–325.

Vu, H., Kim, H.-C., Jung, M., & Lee, J.-H. (2020). FMRI volume classification using a 3D convolutional neural network robust to shifted and scaled neuronal activations. *NeuroImage*, 223, 117328.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory. *Cognitive, Affective, & Behavioral Neuroscience*, 3(4), 255–274. https://doi.org/10.3758/CABN.3.4.255

Wall, M. B., Mentink, A., Lyons, G., Kowalczyk, O. S., Demetriou, L., & Newbould, R. D. (2017). Investigating the neural correlates of smoking: Feasibility and results of combining electronic cigarettes with FMRI. *Scientific Reports*, 7(1), 1–8.

Wang, H.-T., Smallwood, J., Mourao-Miranda, J., Xia, C. H., Satterthwaite, T. D., Bassett, D. S., & Bzdok, D. (2020). Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. *NeuroImage*, 216, 116745. https://doi.org/10.1016/j.neuroimage.2020.116745

Weis, S., Patil, K. R., Hoffstaedter, F., Alessandra, N., Thomas Yeo, B. T., & Eickhoff, S. B. (2020). Sex classification by resting state brain connectivity. *Cerebral Cortex*, 30(2), 824–835. https://doi.org/10.1093/cercor/bhz129

Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in FMRI analyses: Pitfalls and recommendations. *NeuroImage*, 91, 412–419.

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, 21(4), 1732–1747.

Yaple, Z. A., Dale Stevens, W., & Arsalidou, M. (2019). Meta-analyses of the n-back working memory task: FMRI evidence of age-related changes in prefrontal cortex involvement across the adult lifespan. *NeuroImage*, *196*, 16–31.

Zhou, X., Li, M., Zhou, H., Li, L., & Cui, J. (2018). Item-wise Interindividual brain-behavior correlation in task neuroimaging analysis. *Frontiers in Neuroscience*, *12*, 817.

Zhuang, X., Yang, Z., Sreenivasan, K. R., Mishra, V. R., Curran, T., Nandy, R., & Cordes, D. (2019). Multivariate group-level analysis for task FMRI data with canonical correlation analysis. *NeuroImage*, *194*, 25–41.

SUPPORTING INFORMATION

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

APPENDIX A: MIXED-EFFECTS MULTILEVEL ANALYSIS FOR FMRI DATA

The effect size for each participant estimated from the fMRI data, which is obtained using a GLM at the first level, often deviates from the group effect size as follows (Penny & Holmes, 2007):

$$\theta_i = \alpha_0 + \delta_i, \ i = 1, 2, ..., P \tag{A1}$$

where θ_i is the individual effect size for the ith subject, α_0 is the group effect, which follows a normal distribution $N(0, \tau^2)$, δ_i is the deviation of the effect size for the ith subject compared to the effect size of the associated group, and P is the number of subjects. The estimated individual effect size inherently confounded by estimation error in the first-level GLM is thus

$$y_i = \theta_i + \varepsilon_i, \tag{A2}$$

where ε_i is the sampling error of the ith subject, which follows $N(0, \sigma_i^2)$, and σ_i^2 represents the level of within-subject variability. From Equations (A1) and (A2), the MEMA model is given as

$$y_i = \theta_i + \varepsilon_i = \alpha_0 + \delta_i + \varepsilon_i, \ y_i \sim N(\alpha_0, \ \tau^2 + \sigma_i^2), \ i = 1, 2, ..., P \tag{A3}$$

where y_i is the estimated effect estimate (β) or the linear combination (i.e., the contrast) of the β values for the ith subject from the first-level GLM.

In the second level, a homogeneity test for θ_i is conducted using the Q-statistics as follows (Cochran, 1954; Viechtbauer, 2010):

$$Q = \sum_i \frac{(y_i - \mu)^2}{\sigma_i^2} \tag{A4}$$

The Q-statistics follow the Cochran's χ^2 distribution with $(P-1)$ degrees of freedom and μ is calculated from a weighted mean of the effect estimates ($\mu = \sum_i \frac{1}{\sigma_i^2} y_i / \sum_i \frac{1}{\sigma_i^2}$), in which an individual effect estimate with a small within-subject variability has greater weight in the mean compared to an individual effect with a large within-subject variability (Chen et al., 2012). The expected value of the Q score can thus be expressed as follows (DerSimonian & Laird, 1986):

$$E[Q] = \tau^2 \left(\sum_i \frac{1}{\sigma_i^2} - \sum_i \left(\frac{1}{\sigma_i^2}\right)^2 / \sum_i \frac{1}{\sigma_i^2} \right) + (P-1) \tag{A5}$$

Here, the approximated τ^2 can be represented based on an ensemble average of the Q score:

$$\hat{\tau}^2 = \frac{Q - (P-1)}{\sum_i \frac{1}{\sigma_i^2} - \left(\sum_i \left(\frac{1}{\sigma_i^2}\right)^2 / \sum_i \frac{1}{\sigma_i^2} \right)} \tag{A6}$$

To avoid a negative value, $\tau^2 = \max\left(0, \frac{Q-(n-1)}{\sum_i \frac{1}{\sigma_i^2} - \sum_i \left(\frac{1}{\sigma_i^2}\right)^2 / \sum_i \frac{1}{\sigma_i^2}} \right)$ is considered (Chen et al., 2012). Thus, the null hypothesis can be formulated from τ^2, which follows the χ^2 distribution:

$$H_0 : \tau^2 = 0 \tag{A7}$$

When the null hypothesis holds, the overall variability of the individual effect sizes is dominated by the within-subject variability and thus the MEMA model is reduced to a fixed-effects model (Chen et al., 2012). On the other hand, when an alternative hypothesis holds (i.e., the τ^2 value is sufficiently greater than zero), significant heterogeneity is present across subjects due to large between-subject variability in comparison to within-subject variability.

APPENDIX B: CANONICAL CORRELATION ANALYSIS (CCA)

CCA finds canonical variates to determine the relationship between two sets of multivariate variables (e.g., effect estimates from multiple voxels in an ROI and behavioral data). Suppose X ($P \times l$) and Y ($P \times m$) are two sets of multivariate variables with l and m dimensions ($l > m$), respectively, from P subjects. The CCA can find the two vectors $\boldsymbol{a} \in R^{l \times 1}$ and $\boldsymbol{b} \in R^{m \times 1}$ that maximize the correlation of the two canonical variates $X\boldsymbol{a}$ and $Y\boldsymbol{b}$:

$$\rho = \underset{\boldsymbol{a}, \boldsymbol{b}}{\mathrm{argmax}}(X\boldsymbol{a}, \ Y\boldsymbol{b}) \tag{A8}$$

The two variables $U = X\boldsymbol{a}$ and $V = Y\boldsymbol{b}$ are defined as the first pair of canonical variates. The correlation between them is expressed as follows:

$$\rho = \frac{\mathrm{Cov}(U,V)}{\sqrt{\mathrm{Var}(U)\,\mathrm{Var}(V)}} = \frac{\boldsymbol{a}^T \sum_{XY} \boldsymbol{b}}{\sqrt{\boldsymbol{a}^T \sum_{XX} \boldsymbol{a}}\sqrt{\boldsymbol{b}^T \sum_{YY} \boldsymbol{b}}}, \tag{A9}$$

where $\sum_{XX} = \mathrm{Cov}(X,X) = X^T X, \sum_{YY} = \mathrm{Cov}(Y,Y) = Y^T Y, \sum_{XY} = \mathrm{Cov}(X,Y) = X^T Y$, and Cov and Var denote the covariance matrix and variance, respectively. The canonical weights $\boldsymbol{a}$ and $\boldsymbol{b}$ that maximize the correlation in Equation (A9) can be found from a generalized eigenvalue problem (Ge, Jin, Netrapalli, & Sidford, 2016; Safo, Ahn, Jeon, & Jung, 2018). We can find another set of two vectors for the canonical weights by maximizing the correlation between a pair of second canonical variates that are uncorrelated with the first pair of canonical variates. This process can be repeated to find the maximum number, $\min(l,m)$, of canonical variates and their canonical weights.