NeuroImage 157 (2017) 118–128

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Predicting behavior change from persuasive messages using neural representational similarity and social network analyses

Teresa K. Pegors[a,*], Steven Tompson[b], Matthew Brook O'Donnell[c], Emily B. Falk[c,*]

[a] Department of Psychology Azusa Pacific University, 901 E Alosta Ave., Azusa, CA 91702, USA
[b] Department of Bioengineering University of Pennsylvania, 210 South 33rd St Suit 240 Skirkanich Hall, Philadelphia, PA 19104, USA
[c] Annenberg School for Communication University of Pennsylvania, 3620 Walnut St, Philadelphia, PA 19104, USA

ARTICLE INFO

Keywords:
FMRI
Health behavior
RSA
Motivation
Multivariate analyses
Smoking

ABSTRACT

Neural activity in medial prefrontal cortex (MPFC), identified as engaging in self-related processing, predicts later health behavior change. However, it is unknown to what extent individual differences in neural representation of content and lived experience influence this brain-behavior relationship. We examined whether the strength of content-specific representations during persuasive messaging relates to later behavior change, and whether these relationships change as a function of individuals' social network composition. In our study, smokers viewed anti-smoking messages while undergoing fMRI and we measured changes in their smoking behavior one month later. Using representational similarity analyses, we found that the degree to which message content (i.e. health, social, or valence information) was represented in a self-related processing MPFC region was associated with later smoking behavior, with increased representations of negatively valenced (risk) information corresponding to greater message-consistent behavior change. Furthermore, the relationship between representations and behavior change depended on social network composition: smokers who had proportionally fewer smokers in their network showed increases in smoking behavior when social or health content was strongly represented in MPFC, whereas message-consistent behavior (i.e., less smoking) was more likely for those with proportionally more smokers in their social network who represented social or health consequences more strongly. These results highlight the dynamic relationship between representations in MPFC and key outcomes such as health behavior change; a complete understanding of the role of MPFC in motivation and action should take into account individual differences in neural representation of stimulus attributes and social context variables such as social network composition.

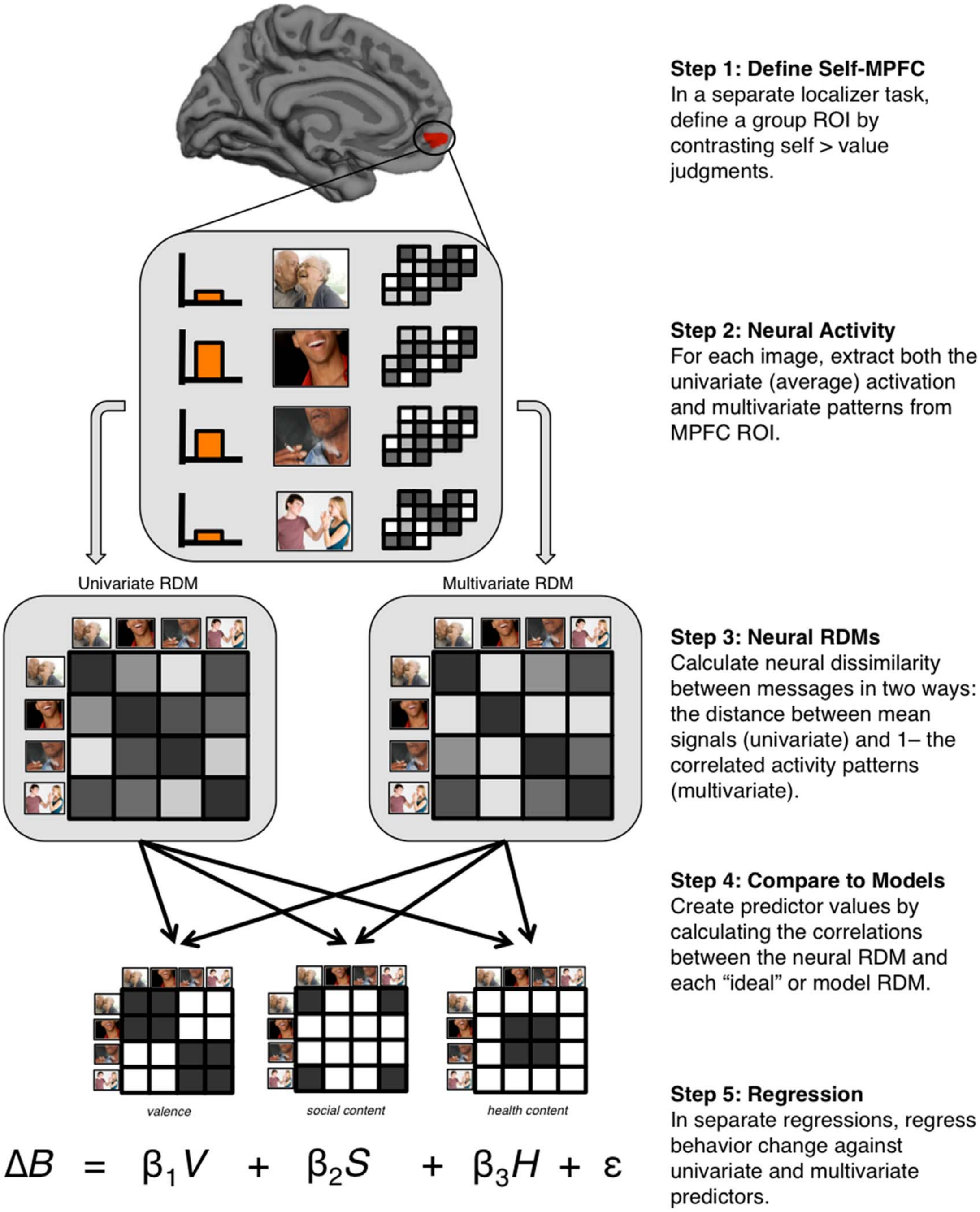
Introduction

Activity in MPFC during exposure to persuasive messages can predict later behavior change, and this predictive capacity has been attributed to MPFC's role in indexing message self-relevance (Cooper et al., 2015; Falk et al., 2010, 2011). Prior neuroscientific research on behavior change has focused on linear mappings between behavior and average activation within the MPFC, but a number of factors can add additional complexity to the relationship between brain and behavior. For example, the same message content may elicit multiple psychological responses which depend on the message recipient's life experience and social context, and thus have different downstream influences on behavior (Tompson et al., 2015).

In the current investigation, we used representational similarity analysis and social network analysis to examine how brain responses in MPFC are moderated by individual differences in ways that map to later behavior change. To do so, we measured MPFC activity in smokers as they viewed messages that varied in the presence or absence of three key features known to affect health decisions and behavior change: portrayal of risk/ negative consequences (Peters et al., 2012; Witte and Allen, 2000a); portrayal of social norms/ consequences (Cialdini et al., 2006; Mead et al., 2014) and portrayal of health outcomes/ consequences (Hammond, 2011) (See fMRI Tasks and Fig. 1 for examples). We focused our analyses on a subregion of MPFC involved in self-related processing (hereafter: *self-MPFC*). Self-relevance is thought to play a large role in the effectiveness of persuasive messages on behavior change (Becker, 1974; Rosenstock, 1974), and a number of researchers have implicated sub-regions of MPFC in self-relevant processing (Amodio and Frith, 2006; Denny et al., 2012; Northoff et al., 2006; Schmitz and Johnson, 2007). Integrating these literatures, a recent report showed that activity in a functionally-defined self-related region of MPFC predicts later behavior

[*] Corresponding authors.
E-mail addresses: tpegors@apu.edu (T.K. Pegors), stevro@seas.upenn.edu (S. Tompson), mbod@asc.upenn.edu (M.B. O'Donnell), falk@asc.upenn.edu (E.B. Falk).

http://dx.doi.org/10.1016/j.neuroimage.2017.05.063
Received 7 January 2017; Accepted 30 May 2017
Available online 31 May 2017
1053-8119/



Step 1: Define Self-MPFC
In a separate localizer task, define a group ROI by contrasting self > value judgments.

Step 2: Neural Activity
For each image, extract both the univariate (average) activation and multivariate patterns from MPFC ROI.

Step 3: Neural RDMs
Calculate neural dissimilarity between messages in two ways: the distance between mean signals (univariate) and 1– the correlated activity patterns (multivariate).

Step 4: Compare to Models
Create predictor values by calculating the correlations between the neural RDM and each "ideal" or model RDM.

Step 5: Regression
In separate regressions, regress behavior change against univariate and multivariate predictors.

$$\Delta B = \beta_1 V + \beta_2 S + \beta_3 H + \varepsilon$$

Fig. 1. Diagram of RSA analysis method.

change (Cooper et al., 2015). This evidence, along with recent work which shows that multivariate representations of both social and valence content are present within ventral MPFC (Chavez and Heatherton, 2014) suggests that representations of these and similar attributes may exist in MPFC and may be associated with later behavior.

We implemented representational similarity analyses (RSA; Kriegeskorte et al., 2008; Nili et al., 2014) to test whether individual differences in the representation of message content features related to later behavior change (i.e., smoking reduction). Rather than measuring overall activity for each message in a given brain region, RSA uses the relative similarity of neural activity between items (in our case, messages) to make inferences about the content encoded in that region. Our first hypothesis, therefore, was that individual differences in the degree to which self-MPFC represented message content (information about valence [regardless of consequence type] and information about social and health consequences [regardless of valence]) would be associated with later behavior change. Specifically, we predicted that stronger representations of each type of message content would be associated with greater reductions in smoking behavior.

Even if all viewers were to show similar levels of message content representation in the brain, differences in individual life experiences may modulate the effectiveness of this content in promoting behavior change. Specifically, social context, including social network composition, may change the processing of health information and downstream behaviors (Christakis and Fowler, 2007). In the context of health messaging, this may make social, health and risk information more or less salient to the self or may change the connotations drawn to mind, changing whether the person enacts message-consistent behavior change (O'Donnell and Falk, 2015). Our second hypothesis, therefore, was that individual differences in social network composition moderates the relationship between neural representations of content and smoking behavior, such that an increased proportion of smokers in one's network would strengthen the impact of neural representations on smoking reduction.

In summary, the fundamental hypotheses of our analyses were 1) that image attribute information relates to message-consistent future behavior by the strength of its representation in the portion of MPFC that engages self-related processing; and 2) that the nature of this relationship will be significantly moderated by the individual's social environment, as indexed by social network composition.

It should be noted that the current dataset uses the same participants included in Cooper et al. (2015), which examined smoker's neural responses to a different task, and Falk et al. (2015), which examined aggregate neural responses to the images used in the current task and linked them to population level outcomes. However, no prior reports have used these data to analyze individual differences in representations of image content in MPFC, nor examined social network composition as a moderator of these effects. Hence, the current analyses substantially extend prior reports.

Materials and methods

Participants

50 smokers took part in the fMRI study, and all participants were consented in accordance with the procedures of the Institutional Review Board of the University of Michigan. Sample size was based on funding availability. 3 subjects were excluded for excessive head motion (movement exceeding 3 mm translation or 1° rotation), 5 subjects were excluded for questionable data quality (distorted field maps and recon "features"), 1 subject could not be reached for their final session, and 1 subject was excluded as a behavioral outlier (see Procedure) leaving us with 40 usable subjects (Females: 15, age mean/SD: 32.9/13.1, age range: 19–64 years old). 25 subjects were White/European, 5 were African American, 5 were Hispanic, and 5 reported a mixed ethnicity. 8 participants reported a bachelor's degree or higher, 3 participants reported an associate's degree, 10 participants were currently in college, and 19 participants had a high school degree or lower. (See Fig. S-1 in Supplementary Materials for a detailed chart of subject exclusion criteria.[1]).

Participants were recruited from the general population using Craigslist and UMClinicalStudies.org. Interested participants completed an eligibility screening phone call. To participate in the study, participants should have been a smoker for at least 12 months, reported smoking at least 5 cigarettes per day for the past month, and been between the ages of 18 and 65. Additionally, participants had to meet standard fMRI eligibility criteria, including having no metal in their body, no history of psychiatric or neurological disorders, and not taking any psychiatric or illicit drugs at the time of the study. Additionally, all participants were required to be right-handed.

Procedure

Enrolled participants completed three separate appointments. During the intake appointment (session 1), participants received consent and completed baseline self-report surveys over the course of approximately an hour, including questions about the level of their intentions to quit smoking and their smoking behavior. During the second session, subjects completed an hour of tasks in the fMRI scanner as well as pre- and post-scan self-report measures that included the baseline smoking behavior value employed here. The overall session lasted approximately 3 h, and took place an average of 4.8 days after the first appointment (SD: 3.3 days). The third and final session was a phone appointment, in which participants gave follow-up self-report measures, including endpoint smoking behavior. This appointment took place an average of 38.8 days after the fMRI appointment (SD: 8.4 days). As a standard procedure in preparing data for regression analyses, we checked for outliers in our behavioral measure of proportional change in smoking behavior (see below for information on the calculation of this measure). Out of the 41 scanned subjects for which we had data from all three sessions, only one subject was identified as a behavioral outlier (> 2.5 SD from the mean on behavior change; the participant in question reported smoking 18 cigarettes per day at time points 1 and 2, and then reported smoking 40 cigarettes per day at time point 3, whereas no other subject at any time point reported smoking more than 30 cigarettes per day). We excluded this subject, as we believe the final behavioral report may have been misrecorded and outlying data is known to have a disproportionate weighting when determining linear fit.

For all three appointments, participants reported the number of cigarettes they smoked per day. As a reference, they were told that a pack contains 20 cigarettes. Self-report measures are significantly correlated with physiological metrics such as expired CO (Falk et al., 2011; Vogt et al., 1977) and saliva, urine, and serum cotinine (Etter et al., 2000; Klebanoff et al., 1998; Pickett et al., 2005; Pokorski et al., 1994). As such, self-report measures are commonly used to track smoking behavior change (Chua et al., 2011a, 2011b; Jasinska et al., 2012).

During each session, participants were asked whether they were currently enrolled in any type of quit-smoking program and/or whether they had a planned quit date. Of the 40 participants, on the day of the scan no participants were enrolled in a quit smoking program nor had a planned quit date. At the follow-up appointment, two participants were enrolled in a quit program, one had stopped smoking, and one other subject had a planned quit date. Given that only two participants were enrolled in quit smoking programs, we inferred that the majority of the change in participants' smoking behavior was not a result of external professional interventions.

In all of our analyses, our behavioral outcome measure of interest was the proportion change in number of reported cigarettes smoked per day. To calculate this number for each participant, we subtracted the cigarettes smoked per day at baseline (session 2) from the cigarettes smoked per day at the endpoint (session 3), and divided this number by the baseline measure (session 2): [(CigarettesEndpoint − CigarettesBaseline) / CigarettesBaseline].

fMRI tasks

The main focus of our study was a task in which participants viewed and rated images that were modeled off of proposed graphic warning labels for tobacco by the U.S. Food and Drug Administration (FDA). Each image was shown with the accompanying text: "Stop Smoking. Start Living." Images were each presented for 4 s, after which participants were prompted with the text "*This ad makes me want to quit.*" Participants had 3 s to make a rating on a Likert 5-point scale in response to the prompt (anchors: *definitely not, definitely does*). A jittered fixation between trials varied between 3 and 7.5 s with a mean of 4.1 s. Our analyses focus on the 60 images that were viewed by all subjects across 2 runs. (20 additional images were personalized for each participant, and were not analyzed here.) Each of the two runs for each subject contained a random sample of 15 neutral-valence images and 15 negative-valence images which were selected using the Python random sample function *sample()* from the standard library random module. 'Risk' information in the negative valence images either related to social or health consequences. Negative-valence social images portrayed scenes such as social exclusion from family, co-

[1] 37 subjects were the same as those used in Cooper et al. (2015) – the groups were slightly different due to differing motion and data quality in the separate scan tasks analyzed.

workers and friends (n = 12). Negative-valence health images portrayed scenes such as individuals in the hospital or a casket, smoking related symptoms such as yellow teeth or a neck stoma (n = 18). Neutral-valence social images portrayed social scenes with family, co-workers, and friends (n = 11). Neutral-valence health images portrayed people being physically active (n = 19). (Fig. S-2 shows all images displayed by category.) The neutral images were matched with the negative images on dimensions including the number and general demographics of people in the images, but did not portray smoking behavior and had "neutral" valence. (Valence as "negative" or "neutral" was based on independent data in which 70 smokers participated in an Amazon Mechanical Turk survey where they rated each image on a 5 point Likert scale according to various emotions [negative: depressing, disgusting, frightening, unpleasant; positive: encouraging, hopeful, inspiring, meaningful]. Given that the positivity for the control images was 2.01 for the degree to which the four positive emotions were represented, with 1 = "not at all" and 5 = "very much," we labeled these images as having "neutral" rather than "positive" valence). During the course of the scan, subjects also completed a self-related processing localizer task used to identify the portion of MPFC most highly associated with self-related processing, and two other tasks which are not the focus of this study (see Cooper et al., 2015 for a description of the banner ads task. Additionally, an "arguments" task presented subjects with statements about what they should do [e.g. "people should eat broccoli"], and subjects were instructed to respond whether they agree with the statement [agree condition], think of reasons in support of doing that task [in favor condition] or think of reasons against doing that task [against condition]).

Social network structure

At the end of Session 1, participants completed a web-based task that measured their social network composition focusing on their friends and the connections between them (i.e. the participant's ego network; Marsden, 2002). Participants first listed up to twenty individuals with whom they had interacted in the past week through each of the following communication mediums: 1. face-to-face interaction, 2. phone calls, 3. text messages and 4. social media (i.e., Facebook interactions),[2] as well as any close friends and family members (who might not have been captured in the previous lists). They then indicated (i) how close they felt to each individual, (ii) whether each individual was a cigarette smoker and (iii) which of those individuals listed knew each other (see O'Donnell and Falk, 2015: 278–279 for more details). We then calculated the ratio of smokers in each network by dividing the number of smokers in each participant's network by the number of non-smokers in each participant's network.

fMRI acquisition

Neuroimaging data were acquired using a 3 T GE Signa MRI scanner. Two functional runs for the self-localizer task (288 volumes total) were collected at the start of the scan and two functional runs of the images task (227 volumes each) were also collected. Functional images were recorded using a reverse spiral sequence (TR = 2000 ms, TE = 30 ms, flip angle = 90°, 43 axial slices, FOV = 220 mm, slice thickness = 3 mm; sequential descending slice acquisition; voxel size = 3.44 × 3.44 × 3.0 mm). The first 5 TRs were not recorded to account for stabilization of the BOLD signal. We also acquired in-plane T1-weighted images (43 slices; slice thickness = 3 mm; voxel size = 0.86 × 0.86 × 3.0 mm) and high-resolution T1-weighted images (SPGR; 124 slices; slice thickness = 1.02 × 1.02 × 1.2 mm) for use in co-registration and normalization.

Imaging data analysis

Preprocessing and data analysis for individual participants was performed using the FMRIB Software Library (Jenkinson et al., 2012). Functional images were pre-processed using FEAT (v6.0.0). Volumes were slice-time corrected, and motion correction (MCFLIRT) was applied by spatially realigning each image with the central image in the run, and these images were then registered to the subject-specific T1-weighted image using Boundary-Based Registration (Greve and Fischl, 2009). High-pass filtering was applied to remove temporal frequencies below 0.02 Hz. Given the nature of our pattern-based analyses, no smoothing was applied.

General linear modeling (GLM) was used to estimate the neural response to each individual image. For each image, we constructed a separate GLM (see Mumford et al., 2012) with the following predictors: 1) a predictor for the 4-second image presentation, 2) a complementing predictor representing all other image presentations in that run, and 3) a predictor representing all feedback periods. Nuisance predictors for global and motion outliers were also included. These outliers were calculated using the Gabrieli Lab's Artifact Detection Tools scripts (global signal threshold: 3 SD, movement threshold: 2 mm; www.nitrc.org/projects/artifact_detect/). All predictors were convolved with FSL's canonical double-gamma HRF.

Unsmoothed parameter estimates for each individual image contrast were registered to the cortical surface using surface templates that had been derived from each participants' T1-weighted anatomical image using Freesurfer's segmentation function (recon-all). The data were then spherically registered to a standard "Buckner40″ brain (fsaverage).

ROI definition

In our analyses we were specifically interested in message content within medial prefrontal cortex (MPFC). We used a self-localizer task to define a group-level MPFC ROI. In this task, participants judged trait words in relation to themselves (e.g., does the participant view herself as "funny"?) and in relation to the word's valence (e.g., is "funny" a positive word?). The same 36 (18 negative and 18 positive) trait words were shown in each condition and were selected from the Anderson trait word list (Anderson, 1968). The task included blocks of six trials with three positive and three negative words in each block, with each block preceded by a three second orientation screen and followed by a two second fixation screen. To maintain a consistent ROI with previous analyses from this dataset, we derived our ROI directly from the ROI reported in Cooper et al. (2015). Therefore, for this contrast, analyses were conducted in SPM and 45 subjects were used in the group contrast of self-related judgments > control judgments. In that analysis, subject level maps were smoothed (8 mm Gaussian kernel) and registered to standard space. The statistical map from the second-level group GLM was corrected with a family-wise error correction at $p < 0.05$, and the resulting ROI was defined as that region that survived correction within the larger anatomical region of MPFC (MNI coordinates: −6, 56, −5; max t-value: 7.23; # voxels: 53). The only difference from Cooper et al. was that we then registered this ROI to standard surface space (fsaverage). This resulting ROI mapped entirely to the left hemisphere (See Fig. 1).

Representational similarity analyses

Our analyses tested two primary hypotheses: 1) that image characteristics are represented in MPFC in a way that relates to future behavior; and 2) that social network composition modulates this relationship. Representational similarity analyses (RSA) have the potential to explore the nature of a region's representational space by using open-ended exploratory techniques and by testing hypothesis-driven models (Kriegeskorte et al., 2008; Mur et al., 2009). Here, our

[2] The web application asked participants to allow our Facebook application to briefly access their Facebook account to make the process of creating the list of friends they had interacted with on Facebook less onerous. For those who allowed this access the most recent 20 people they had interacted with on was automatically added to the list of individuals in their social network. For participants who did not provide access to their Facebook account, they were asked to self-report previous Facebook interactions.

analyses centered on three types of content in the messages: social information, health information, and valence information.

All of our analyses initially required the construction of "model" and neural representational dissimilarity matrices (RDMs). RDMs are matrices that summarize the pairwise similarities between all stimuli in the experiment (in our case, anti-smoking messages). Similarity can be measured by comparing neural activity or the presence or absence of certain types of content. RDMs themselves can be compared to each other in informative ways. Broadly, in our study, we had predictions about what kind of message content would be most relevant to behavior change (e.g. social consequence information), and so constructing an RDM labeling which pairs of messages were similar/dissimilar in their social content allowed us to then compare this model to the actual neural RDM. From this comparison, we could infer that the greater the similarity between the neural and model RDM, the greater degree to which that ROI represents the modeled content (e.g. social information). Importantly, we could then test whether this strength of representation is predictive of later smoking behavior.

More specifically, we first constructed model matrices to signify what the neural data would look like if they perfectly represented our content of interest (valence, social, and health information). Because these RDMs were coding *dissimilarity,* cells were coded as "1" if the two message pairs represented different kinds of content, and cells were coded as "0" if the two message pairs represent the same kind of content. Separate RDMs were constructed for valence (negative consequences and controls), social information, and health information (See Fig. 1 for a representative subset of these RDMs). Given that subjects viewed 60 messages in total, matrices contained 60 rows by 60 columns, such that all pairwise relationships between messages were represented. Cells in the *valence* matrix were coded as "0" if both messages reflected either negative or neutral repercussions of smoking. Cells were coded as "1" if one message was negative and the other neutral. For the *social information* matrix, cells were coded as "0" if both messages reflected social information. Cells were coded as "1" if one message reflected social information and one reflected health information or if both messages reflected health information. For the *health information* matrix, only pairs of health-related messages were coded as "0". (Note that the social and health matrices were not exact inverses of each other, given that both matrices coded as "1" cases in which pairs were not matched on content type).

We then constructed RDMs out of the actual neural response patterns to each message. There were two ways in which we measured neural similarity between messages, resulting in two sets of RDMs. In our "univariate" RDMs, each cell contained the Euclidean distance between the average levels of activity in the ROI for both messages (averaged betas across all vertices[3] in MPFC). If this average overall activity for both messages was either low or high, this distance measure would be low, signifying low dissimilarity (i.e. high similarity). In our "multivariate" RDMs, matrix cells were calculated by taking 1 minus the pairwise correlation between *patterns of* activity (i.e. between two vectors containing all vertex values within MPFC). We used both univariate and multivariate RDMs so that both kinds of content representation (overall ROI activity and vertex-wise patterns of activity) could be used as predictors of later behavior. Most studies using RSA are focused on directly interrogating multivariate patterns of activity contained in specific regions of the brain or relating neural RDMs to item-specific responses, rather than using RSA as a means to measure the relationship between representation of a content set and person-level changes in behavior outside of the scanner. For our purposes, a key benefit in using the RSA method is that we can derive a single number for the degree to which activity within a region of interest reflects specific content features (by correlating RDMs of BOLD activity to each model RDM), and this number can be entered into a regression model to predict behavior. We computed both univariate and multivariate RDMs to bridge the gap between the two research traditions we are bringing together—brain-as-predictor studies that have traditionally focused on univariate activation, and RSA analysis that has traditionally focused on multivariate patterns (See Discussion RSA Methodology for an extended discussion of the use of both univariate and multivariate RDMs).

For both model and neural RDMs, the matrices were symmetric around the diagonal and so subsequent analyses were therefore conducted using only the upper diagonals. All analyses used a combination of scripts from the RSA Toolbox (http://www.mrc-cbu.cam.ac.uk/methods-and-resources/toolboxes/) and custom Matlab scripts.

Message content and behavior change

Our first major goal was to test whether differences in the strength of neural content representation across subjects predicted differences in later behavior change. To do this, we regressed proportion change in smoking behavior against three predictors: the Spearman[4] correlations between the individual's neural RDM and the three model RDMs *valence, social information, health information.* (See Fig. 1 for a schematic of the analysis steps.) One regression used univariate neural RDMs and a second regression used multivariate neural RDMs. The model used is summarized by the following equation:

$$\Delta B = \beta_0 + \beta_1 V + \beta_2 S + \beta_3 H + \varepsilon, \tag{1}$$

where ΔB is the proportion reduction in smoking behavior, and V, S, and H, are the correlations (Fisher r-to-z transformed) between the neural RDM and valence information, social information, and health information, respectively.

Based on significant findings for certain predictors in our univariate and multivariate models, we ran an additional exploratory regression for each, subdividing content into both valence and content type: *negative social information*, *negative health information*, *neutral social information*, and *neutral health information.*

$$\Delta B = \beta_0 + \beta_1 NS + \beta_2 NH + \beta_3 TS + \beta_4 TH + \varepsilon, \tag{2}$$

Where ΔB is the proportion reduction in smoking behavior, and NS, NH, TS, and TH are the correlations (Fisher r-to-z transformed) between the neural RDM and the negative social, negative health, neutral social, and neutral health models, respectively.

Social network structure and behavior change

Our second goal was to test whether the makeup of an individuals' social network influenced the way content motivated behavior. To do this, we measured, for each subject, the proportion of smokers to non-smokers in their recent social interaction network and added this value as an interaction term to each of the three model predictors from Eq. (1). The model used is summarized by the following equation:

$$\Delta B = \beta_0 + \beta_1 R + \beta_2 V + \beta_3 S + \beta_4 H + \beta_5 V{*}R + \beta_6 S{*}R + \beta_7 H{*}R + \varepsilon, \tag{3}$$

Where all of the terms are the same as Eq. (1), and the additional variable R is the proportion of smokers to non-smokers in each subjects' social network.

All regressions were run using the software R (Version 3.1.0; R Core Team, 2014). Diagnostic plots were used for all regressions to confirm normality and homoscedasticity.

[3] Rather than voxels, vertices are the units used in surface-space.

[4] Spearman correlations are used so that a linear match between two matrices does not need to be assumed (Kriegeskorte et al., 2008)

Results

Behavioral results

The major dependent measure in our study was change in smoking habits before and after a scan session where participants viewed anti-smoking messages. Overall, 27 participants reported reducing their smoking behavior (the number of cigarettes smoked per day at follow-up was less than the number reported during the scan session), 7 participants reported no change in behavior, and 6 participants reported an increase in smoking behavior. The mean proportion change was a 25.3% decrease in smoking behavior (SD: 36.5%).

Participants also completed questionnaires that allowed us to conduct a social network analysis of the ratio of smokers to non-smokers in each participants' recent interaction network (i.e., people the participant interacted with in the past week). On average, participants reported interacting in the past week with 25.8 friends (SD: 11.44, min: 5, max: 48). Across subjects, the average ratio of smokers to non-smokers was 0.55, meaning that, on average, there were 0.55 smokers in a participants' recent interaction network for every 1 non-smoker. 4 participants had equal or more smokers than non-smokers in their network, whereas the rest had a greater proportion of non-smokers.

Message content representations predicts behavior change in MPFC

Our first question asked whether the degree to which specific message content was represented in a functionally localized region of interest defined by self-related processing (*self-MPFC*) predicted later smoking behavior change. We defined this self-MPFC ROI in a separate localizer task as the region involved in self-related processing (see ROI Definition). For each image, content-related information could potentially be encoded by either the mean signal across the self-MPFC (i.e., univariate signal) or by a particular pattern of activation within the self-MPFC (i.e., multivariate signal). Therefore, we measured the similarity of neural activity in self-MPFC between all messages by constructing two kinds of representational dissimilarity matrices (RDMs): one comparing univariate activity between all pairwise messages, and one comparing multivariate activity between all pairwise messages. These RDMs were then correlated with RDM models of message content (valence, social, and health information) to determine the degree to which neural activity acted similarly to idealized models of each type of message content (neural RDMs correlated with univariate valence RDM: mean $r = -0.001$, SD = 0.015; with univariate social RDM: mean $r = 0.010$, SD = 0.069; with univariate health RDM: mean $r = -0.008$; SD = 0.081; with multivariate valence RDM: mean $r = -0.007$, SD = 0.016; with multivariate social RDM: mean $r = 0.005$, SD = 0.048; with multivariate health RDM: mean $r = -0.007$, SD = 0.052, see Supplemental Materials S-3 for additional tests and discussion related to these results). These scores were then input into regressions predicting later smoking behavior (See Fig. 1 for an overview of the analysis steps).

Our univariate RSA regression model revealed that the degree to which the self-MPFC represented valence information significantly predicted later behavior change (β = −9.69, t(36) = −2.55, p = 0.015): Subjects who represented valence information in self-MPFC more strongly during exposure to persuasive messages were more likely to reduce their smoking over the next 30 days (See Fig. 2).

Univariate representations of the other two types of content did not significantly predict behavior change. (The overall univariate model significantly predicted behavior: R^2 = 0.195, adjusted R^2 = 0.128, $F(3,36) = 2.91$, $p = 0.048$.) Our multivariate RSA model, on the other hand, revealed that at the pattern level, social information (driven by negative social information, as described below; Table 2) significantly predicted later behavior, such that those who showed greater representation of social information in the multivariate pattern of responses within MPFC showed an increase in smoking behavior (β = 6.20, t(36) = 2.17, p = 0.007) (Fig. 3); In other words, subjects who more strongly represented the (negative) social content of the persuasive messages at the multivariate level were less likely to reduce smoking over the next month. Multivariate representations of valence and health content were not associated with later behavior change. (The overall multivariate model significantly predicted behavior: R^2 = 0.242, adjusted R^2 = 0.178, $F(3,36) = 3.823$, $p = 0.018$.) (See Table 1 for full results of both univariate and multivariate models).

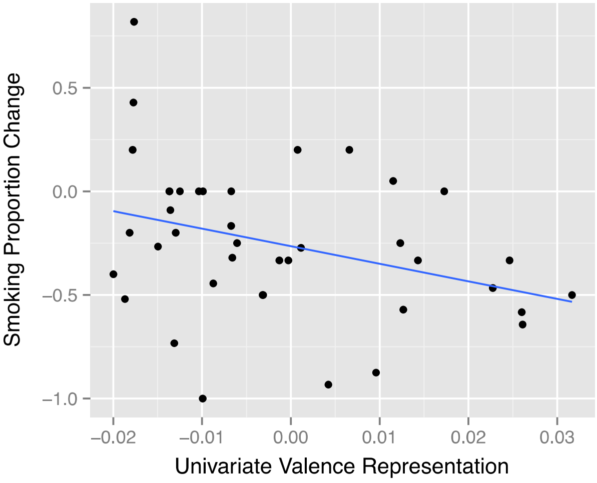


Fig. 2. Univariate Valence Representation Predicts Behavior Change. **Legend:** As univariate representations of valence increase, smoking decreases. Each data point represents one subject (n = 40). The x-axis displays the degree to which a subjects' neural RDM correlated with the valence model RDM (all scores were Fisher r-to-z transformed). The y-axis displays the degree to which a subject changed smoking behavior (negative numbers indicate an overall proportion reduction in smoking.).

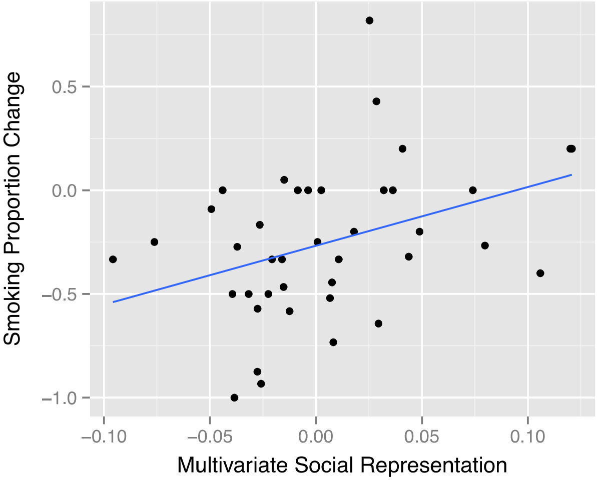


Fig. 3. Multivariate Social Representation Predicts Behavior Change. **Legend:** As multivariate representations of social information increase, participants reduce their smoking less. Each data point represents one subject (n = 40). The x-axis displays the degree to which a subjects' neural RDM correlated with the social model RDM (all scores were Fisher r-to-z transformed). The y-axis displays the degree to which a subject changed smoking behavior (negative numbers indicate an overall proportion reduction in smoking.).

We explored these results further by looking at whether the predictive valence information at the univariate level was more strongly associated with the valence of social or health information, and whether the predictive social information at the multivariate level was more strongly associated with negative or neutral valence. To do this, we ran another set of regressions, this time encoding for 4 subtypes of content: *negative health, negative social, neutral health, and neutral social* (See Eq. (2) in Section Message Content and

Table 1
Content representations predict behavior.

Eq. (1): $\Delta B = \beta_1 V + \beta_2 S + \beta_3 H + \varepsilon$

(N=40)	**β**	**Std. Error**	**t-value**	**95% C.I.**	**P-value**
Using univariate RDMs					
$\beta_1 V$ (valence)	−9.69	3.8	−2.55	−17.40, −1.99	0.015*
$\beta_2 S$ (social consequences)	−5.73	3.16	−1.81	−12.14, 0.68	0.078
$\beta_3 H$ (health consequences)	−4.35	2.67	−1.63	−9.76, 1.07	0.112
Using multivariate RDMs					
$\beta_1 V$ (valence)	5.01	3.37	1.49	−1.83, 11.86	0.146
$\beta_2 S$ (social consequences)	6.18	2.17	2.84	1.77, 10.58	0.007**
$\beta_3 H$ (health consequences)	3.25	2.01	1.62	−0.82, 7.32	0.114

Legend: In self-MPFC, the degree to which brain activity correlated with models of valence, social, and health representation were used as predictors of smoking behavior change (where negative values indicate a reduction in smoking). ($p < 0.05$=*, β = unstandardized parameter estimate, C.I. = confidence interval).

Table 2
Sub-types of content representations predicts behavior.

Eq. (2): $\Delta B = \beta_1 NH + \beta_2 NS + \beta_3 TH + \beta_4 TS + \varepsilon$

(N=40)	**β**	**Std. Error**	**t-value**	**95% C.I.**	**P-value**
Using univariate RDMs					
$\beta_1 NH$ (negative health)	−5.71	2.67	−2.14	−11.14, −0.28	0.040*
$\beta_2 NS$ (negative social)	−4.75	2.57	−1.85	−9.96, 0.47	0.073
$\beta_3 TH$ (neutral health)	−4.71	2.45	−1.92	−9.67, 0.26	0.063
$\beta_4 TS$ (neutral social)	−6.77	3.02	−2.24	−12.91, −0.63	0.032*
Using multivariate RDMs					
$\beta_1 NH$ (negative health)	1.04	1.82	0.57	−2.66, 4.73	0.573
$\beta_2 NS$ (negative social)	6.74	1.56	4.32	3.57, 9.91	0.000***
$\beta_3 TH$ (neutral health)	2.86	1.62	1.76	−0.43, 6.15	0.086
$\beta_4 TS$ (neutral social)	2.28	1.74	1.31	−1.25, 5.82	0.198

Legend: In self-MPFC, representations of social and health information were subdivided into valence types and used to predict smoking behavior change (where negative values mean a reduction in smoking). ($p < 0.05$=*, $p < 0.001$=***, β = unstandardized parameter estimate, C.I. = confidence interval).

Behavior Change). Here, for the univariate encoding of valence, it was negative health information (i.e., risk information) and neutral social information that were significantly predictive of later behavior change (negative health: β = −5.71, t(35) = −2.14, p = 0.040; neutral social: β = −6.77, t(35) = −2.24, p = 0.032) (The overall fit of this univariate model was R^2 = 0.133, adjusted R^2 = 0.034, F(4,35) = 1.338, p = 0.275). Teasing apart the multivariate encoding of social information, we found that negative social information, specifically, was predictive of behavior change (β = 6.74, t(35) = 4.32, p = 0.0001). (The overall fit of this multivariate model was R^2 = 0.367, adjusted R^2 = 0.295, F(4,35) = 5.07, p = 0.0025). (See Table 2 for the full results of these regressions).

Social network structure moderates MPFC prediction of behavior change

We next examined whether the relationship between the strength of representation of specific image attributes and later behavior change was moderated by individual differences in social context. Given the importance of social networks in setting social norms and experiences of consequences, we examined the interaction between the ratio of smokers to non-smokers in each participant's social network and brain activity in predicting behavior change.

As a reminder, in our univariate model without the interaction term, stronger valence representation had been associated with greater reductions in smoking behavior. Adding the social network interaction term did not significantly moderate this relationship with valence information (See Table 3). (The overall fit of this univariate model was R^2 = 0.245, adjusted R^2 = 0.079, F(7,32) = 1.479, p = 0.21).

Our original multivariate model had shown that stronger social representations were associated with message inconsistent behavior change; an interaction analysis suggested that this effect was stronger for those who had a smaller proportion of smokers to non-smokers in their social network (network*social consequences: β = −12.26, t(32) = −2.24, p = 0.032) (See Fig. 4b). Next, although we did not observe a main effect of multivariate representations of health consequences on behavior change, adding the social network term revealed a significant interaction between the proportion of smokers in a participant's network and representations of health information (network*health: β = −6.01, t(32) = −2.61, p = 0.014) (See Table 3). In this case, greater proportions of smokers in the social network were associated with increased message-consistent change in smoking as MPFC representation of health information also increased (See Fig. 4a). (The overall fit of this multivariate model was R^2 = 0.410, adjusted R^2 = 0.281, F(7,32) = 3.179, p = 0.011).

Conversely, for those who had a smaller proportion of smokers to non-smokers in their network, the more strongly health content was represented in MPFC and the less consistent the participant's behavior was with the anti-smoking messages. Stated another way, participants with larger numbers of smokers in their social networks were more likely to maintain or reduce smoking when health information was strongly represented in self-MPFC.

Discussion

Our data provide new insight about how neural representations, in combination with social context, may be associated with future action. Broadly, our results demonstrate a novel method for understanding the nuanced relationship between brain and long-term behavior and substantially extend prior investigations that have linked average brain activity in MPFC during persuasive messages to later behavior change (e.g. Chua et al., 2011a, 2011b; Cooper et al., 2015; Falk et al., 2015; Wang et al., 2013). First, we show that univariate patterns of message valence predict message-consistent behavior change. Second, we show that social network composition influences the relationship between behavior change and self-MPFC multivariate representations of social and health outcomes. Supplemental results confirmed that representations within self-MPFC of the degree to which each image makes a person want to quit smoking were also predictive of actual later smoking reduction. When controlling for this variable, however, our primary results linking content representations to behavior change still hold, suggesting that message content dimensions such as valence and social consequences may not be directly tied to a conscious perception of motivation to quit, but rather serve as additional factors associated with behavior change (See Supplemental Analysis S-2 which uses a model RDM representing self-reports of the degree to which each image makes a person want to quit smoking). The fact that heterogeneity in neural representations of message content was associated with later behavior change also highlights that the same message concepts may carry variable influence across individuals. Further, the relationship between neural representations in MPFC and behavioral outcomes depends both on the specific type of content and message recipient's social context.

Table 3
Social network composition modulates brain-behavior relationship.

Eq. (3): $\Delta B = \beta_1 R + \beta_2 V + \beta_3 S + \beta_4 H + \beta_5 V{*}R + \beta_6 S{*}R + \beta_7 H{*}R + \varepsilon$

(N=40)	**β**	**Std. Error**	**t-value**	**95% C.I.**	**P-value**
Using univariate RDMs					
$\beta_1 R$ (propSmokers)	−0.01	0.19	−0.03	−0.39, 0.38	0.978
$\beta_2 V$ (valence)	−17.32	8.06	−2.15	−33.73, −0.90	0.039*
$\beta_3 S$ (social consequences)	0.39	5.97	0.07	−11.77, 12.55	0.948
$\beta_4 H$ (health consequences)	1.26	5.57	0.23	−10.09, 12.61	0.823
$\beta_5 V{*}R$ (valence*propSmokers)	24.43	19.71	1.24	−15.72, 64.59	0.224
$\beta_6 S{*}R$ (social conseq*propSmokers)	−9.65	7.36	−1.31	−24.65, 5.35	0.199
$\beta_7 H{*}R$ (health conseq *propSmokers)	−20.05	8.28	−1.21	−26.91, 6.81	0.234
Using multivariate RDMs					
$\beta_1 R$ (propSmokers)	−0.37	0.19	−1.93	−0.77, 0.02	0.063
$\beta_2 V$ (valence)	11.34	4.70	2.41	1.77, 20.90	0.022
$\beta_3 S$ (social consequences)	12.92	3.51	3.68	5.77, 20.07	0.0008***
$\beta_4 H$ (health consequences)	7.64	2.59	2.95	2.37, 12.90	0.006**
$\beta_5 V{*}R$ (valence*propSmokers)	−11.71	7.36	−1.59	−26.71, 3.28	0.121
$\beta_6 S{*}R$ (social conseq*propSmokers)	−12.26	5.48	−2.24	−23.42, −1.11	0.032*
$\beta_7 H{*}R$ (health conseq *propSmokers)	−6.01	2.30	−2.61	−10.69, −1.33	0.014*

Legend: For self-MPFC, a social network parameter (proportion smokers to non-smokers, where higher values mean more smokers) was added as an interaction term to each Eq. (1) predictor. ($p < 0.05$=*, $p < 0.01$=**, $p < 0.001$=***, β = unstandardized parameter estimate, C.I. = confidence interval).

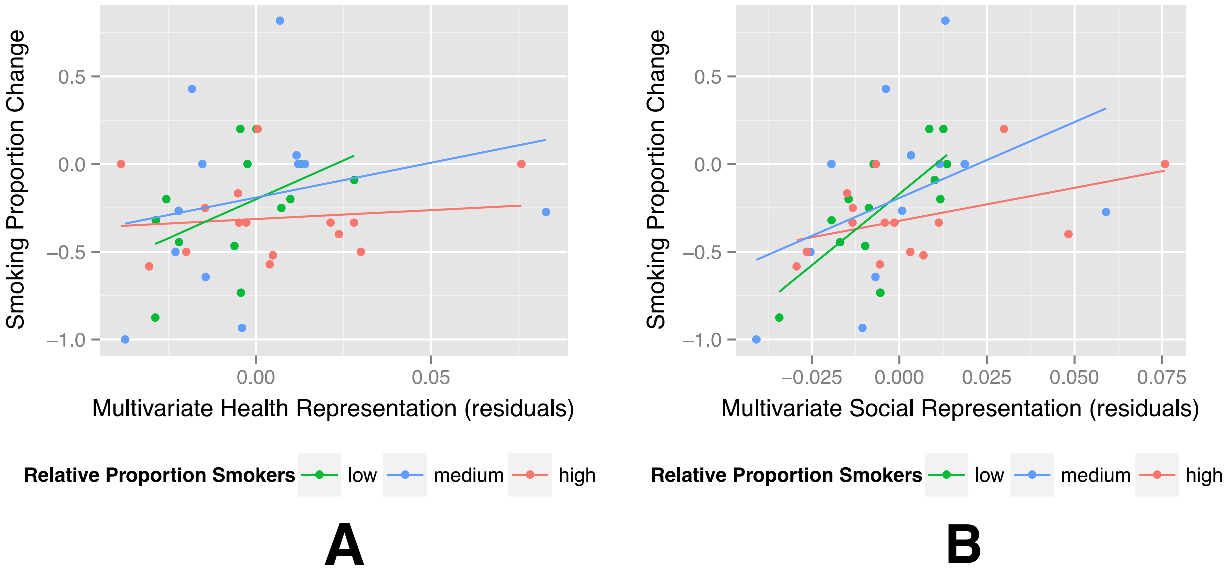


Fig. 4. Interaction Plots for Social Network Analysis (residuals). **Legend:** Low relative proportions of smokers in participant social networks were associated with stronger relationships between multivariate representations of social and health information and smoking change. (For visualization purposes, subjects were divided by network composition into approximately equal groups, though statistics presented in the main paper use continuous variables. Low = less than / equal to 24%; High = greater than 49%).

Message valence and behavior change

During the presentation of persuasive anti-smoking messages, increased univariate representations of valence in a region of MPFC defined by a self-localizer task were associated with decreased smoking behavior over the next 30 days. Stronger representations of negative health content, specifically, were predictive of message consistent behavior change, and the direction of this effect did not depend on individual social context (i.e., social network composition). Prior research has found mixed effects of portraying negative health outcomes and eliciting fear responses (e.g., Earl and Albarracín, 2009 vs. Peters et al., 2013; Witte and Allen, 2000a, 2000b). Our results offer one piece of evidence to help reconcile these disparate findings by suggesting the possibility that the portrayal of strong negative consequences (e.g., fear appeals) is not uniformly effective, but rather individuals who show the strongest representation of valence, and in particular negative health consequences show the greatest behavior change.

Context-dependent effects on behavior change

We also examined how content representations differentially predicted smoking behavior change for smokers depending on the participant's broader social environment (i.e., proportions of other smokers in their social networks). We found that MPFC multivariate representations of social and health information were predictive of later behavior change, but this brain-behavior relationship was context-dependent. Smokers who had recently interacted with few other smokers in their social network were less likely to show message consistent behavior when representing social or health content during persuasive messaging. As the proportion of smokers increased, this tendency was attenuated. By contrast, participants who had recently interacted with more smokers than non-smokers were more likely to show message consistent behavior and reduce smoking when social or health information was strongly represented in self-MPFC. Why would this be the case? One possibility is that a greater presence of smokers in one's daily life offer more vicarious experience of the risks and benefits portrayed in those messages. Another possibility is that smokers who know fewer other smokers show an increased resistance and defensiveness when presented with anti-smoking messages, given a social context in which they may already have built defenses against similar arguments from their non-smoking friends, family and other social contacts. More broadly, these findings are consistent with the idea that persuasive content can affect behavior differently based on socio-cultural context and thus how content is interpreted and received (Tompson et al., 2015).

The MPFC and self-relevance during health messaging

The current data also advance our understanding of the role of MPFC during decision-making. Numerous studies have shown that subregions of the MPFC are more active when making judgments that require self-knowledge (Amodio and Frith, 2006; Denny et al., 2012; Northoff et al., 2006; Schmitz and Johnson, 2007). Different types of self-related processing also appear in major theories of behavior change (Becker, 1974; Rosenstock, 1974). However, "self-related processing" is an umbrella term that can include a number of distinct processes, and different theories of decision-making and behavior change have focused on separate elements. The current results may provide a platform to integrate prior results. In particular, RSA offers a method to handle complex stimuli (like emotionally laden health messages) and, in our study, suggests that variability in how individuals represent the consequences of smoking to themselves and socially may also predict who will be most likely to enact change.

Although tested here in the context of health communication, this method could also be applied to other types of decisions and actions. For example, financial decisions and consumer choices are motivated by a variety of self-related factors (i.e., how valuable is this to me?) where representations of stimulus properties may be able to complement the types of integrated "decision-signals" captured by the average signal in MPFC. In this way, our data extend neuroeconomic views of MPFC by suggesting ways in which different stimulus properties encoded in MPFC can predict complex behaviors.

RSA methodology

Prior neuroscientific research on behavior change has focused on linear mappings between behavior and average activation within the MPFC, but analyses that take into account additional representational and contextual complexity can additionally enhance our understanding of this brain-behavior relationship. Representational similarity analyses, in particular, offer a suite of methods that uses neural and model relationships between items (representational dissimilarity matrices, RDMs) to derive the informational content of neural signals (Kriegeskorte et al., 2008; Nili et al., 2014). Most studies using RSA focus on using multivariate pattern similarities to directly interrogate information contained in specific regions of the brain or to relate neural RDMs to item-specific behavior (these types of behavior measures can also be represented by RDM). Our approach was to bring together two prior research traditions, multivariate RSA analysis of within-subject effects in the scanner, with brain-as-predictor analyses (Berkman and Falk, 2013) that have traditionally focused on average, univariate BOLD activation. Rather than asking about what kind of content is generally represented in MPFC, we examine whether the *degree* to which content is represented predicts later behavior. (In fact, using our own data to average across the group as a whole, we found no significant correlations between either of the multivariate or univariate RDMs and the three content models. [See Analysis S-3 in Supplementary Information.]) In this way, we take RSA one step further by entering into a regression these correlations between brain and model RDMs. To do this, we found value not only in the more standard RDMs which are constructed using similarities in multivariate patterns, but also in creating RDMs using similarities in item-wise univariate activity. By deriving single numbers for the degree to which univariate (or multivariate) activity represents each characteristic of message content (by correlating this RDM to each model RDM), these numbers could then be easily entered into a regression model to predict behavior. In traditional brain-as-predictor approaches, deriving a single value from a standard univariate contrast would require averaging across all items in a given category without preserving those item-specific relationships. With this approach, we believe our method offers a conceptual advance and bridge between two previously unconnected methodologies.

In our study, we used RSA to ask how the brain represented message content in MPFC using both overall BOLD response as well as multivariate patterns of activity. The presence of content contained within MPFC multivariate patterns of activity, specifically, has been demonstrated in other recent literature on stimulus evaluation (Kahnt et al., 2010; Mcnamee et al., 2013; Pegors et al., 2015). Measuring similarities in the patterns of neural responses provided additional information beyond just measuring the overall univariate direction of the response. Among other advantages, multivariate methods may better take into account a property of neuronal populations known as "opponent coding," in which information is coded by some neurons as increased firing and by other neurons as decreased firing. For example, studies in macaques have shown that similar numbers of neurons encode value by either increasing *or* decreasing their firing rates (Kennerley et al., 2011; Padoa-Schioppa, 2009). This property of encoding means that in may cases, the summed signal across the region may appear similar between conditions, even though information is very much present. Multivariate techniques can uncover some of this information not seen using univariate approaches (for an example, see Kahnt et al., 2010). The goal of our method was not to measure the presence of stimulus characteristics in MPFC at this level of representation per se, but to show that the strength of such representations in this region are meaningfully related to later behavior change.

Our results located meaningful (i.e. associated with relevant later behavior) representations within both univariate and multivariate measures of similarity. This highlights the importance of measuring neural responses to content characteristics at different levels of representation. Although multiple other studies have shown mean-level predictions of behavior (Cooper et al., 2015; Falk et al., 2010, 2011), our current results extend these findings by specifying the types of information contained in messaging which are most strongly associated with later behavior change. More broadly, this work also illustrates the role of content representation in decision-making. In other words, our findings highlight the importance of using methods that consider content-specific responses in understanding how the brain makes important, real-world decisions. Furthermore, these results highlight the fact that the underlying assumption that people share common neural representations during an event on average may obfuscate the fact that it is exactly the differences in content representation across individuals that may contain important information related to behavior change. Task-relevant information may be recruited by MPFC on a subject-specific basis for motivating future behavior, and neural models of behavior change should use methods that take this into account.

Why might valence be represented in MPFC at the univariate level and social/health information at the multivariate level of representation? It may be the case that valence is a more fundamental property related to the evaluation of stimuli, and many studies have shown that MPFC activity scales with stimulus value (Bartra et al., 2013). *Types* of content (e.g. social vs. health information), however, may be less associated with evaluation per se, but this content-specific information may still be located in these regions as a way to inform motivation and action. Other studies have shown a univariate/multivariate distinction between the value of a stimuli (represented by overall MPFC activity) and the stimulus type (represented by multivariate patterns) (Mcnamee et al., 2013; Pegors et al., 2015). Our results similarly inform different ways in which different stimulus dimensions might be represented in MPFC.

Conclusions

The same stimulus properties (e.g., negative health information) may be more or less strongly represented in brain regions that compute self-relevance. Our results demonstrate the importance of taking into account individual differences in neural representations of specific content attributes. Furthermore, even at the same strength of representation, different stimulus properties may be differentially associated with action (i.e., behavior change) depending on social context. As such, the current findings shed new light on how stimulus properties and social context variables predict effects of persuasive messaging on behavior change, and open up new avenues for research on real-world decision making.

Funding

Data collection for this project was supported by The Michigan Center of Excellence in Cancer Communication Research through the National Institutes of Health NIH-P50 CA101451; data collection and analysis was supported by the National Institutes of Health NIH 1DP2DA03515601 and by a pilot grant made possible by the National Cancer Institute of the National Institutes of Health and the FDA Center for Tobacco Products (CTP) P50CA179546. We also wish to acknowledge support from support from the Army Research Laboratory through contract number W911NF-10-2-0022, a DARPA Young Faculty Award YFAD14AP00048 and Hope Lab.

Acknowledgments

We acknowledge important contributions to the development of this study from Richard Gonzalez, Sonya Dal Cin, Victor Strecher, Kenneth Michael Cummings, Lawrence An, Matthias Kirch, Michele Demers, Sarah Lillie, Darla Williams, Monte Montgomery at the University of Michigan and Anthony Brown, Seshadri Srinivasa, and Paula Celestino from the New York State Smokers' Quit Line. We thank Francis Tinney Jr., Kristin Shumaker, Li Chen, Nicolette Gregor, Becky Lau, Larissa S Svintsitski, Cole Schaffer for assistance with data collection, J.P. Obley for assistance with code to collect social network data, Joe Bayer for helpful conversations related to social network data, and Nicole Cooper for insightful comments on the methods and results.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuroimage.2017.05.063.

References

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7 (4), 268–277. http://dx.doi.org/10.1038/nrn1884.

Anderson, N.H., 1968. Likableness ratings of 555 personality-trait words. J. Pers. Soc. Psychol. 9 (3), 272–279.

Bartra, O., McGuire, J.T., Kable, J.W., 2013. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage 76, 412–427. http://dx.doi.org/10.1016/j.neuroimage.2013.02.063.

Becker, M.H., 1974. The health belief model and sick role behavior. Health Educ. Behav. 2 (4), 409–419.

Berkman, E.T., Falk, E.B., 2013. Beyond brain mapping: using the brain to predict real-world outcomes. Curr. Dir. Psychol. Sci. 22 (1), 45–55.

Chavez, R.S., Heatherton, T.F., 2014. Representational similarity of social and valence information in the medial pFC. J. Cogn. Neurosci. 27 (1), 73–82. http://dx.doi.org/10.1162/jocn.

Christakis, N. a., Fowler, J.H., 2007. The spread of obesity in a large social network over 32 years. N. Engl. J. Med. 357 (4), 370–379. http://dx.doi.org/10.1056/NEJMsa066082.

Chua, H.F., Ho, S.S., Jasinska, A.J., Polk, T.A., Welsh, R.C., Liberzon, I., Strecher, V.J., 2011a. Self-related neural response to tailored smoking-cessation messages predicts quitting. Nat. Neurosci. 14 (4), 426–427, (http://doi.org/nn.2761)(pii)(10.1038/nn.2761).

Chua, H.F., Ho, S.S., Jasinska, A.J., Polk, T.A., Welsh, R.C., Liberzon, I., Strecher, V.J., 2011b. Self-related neural response to tailored smoking-cessation messages predicts quitting. Nat. Neurosci. 14 (4), 426–427. http://dx.doi.org/10.1038/nn.2761.

Cialdini, R.B., Demaine, L.J., Sagarin, B.J., Barrett, D.W., Rhoads, K., Winter, P.L., 2006. Managing social norms for persuasive impact. Soc. Influ. 1 (1), 3–15. http://dx.doi.org/10.1080/15534510500181459.

Cooper, N., Tompson, S., O'Donnell, M.B., Falk, E.B., 2015. Brain activity in self- and value-related regions in response to online antismoking messages predicts behavior change. J. Media Psychol. 27 (3), 93–109. http://dx.doi.org/10.1027/1864-1105/a000146.

Denny, B.T., Kober, H., Wager, T.D., Ochsner, K.N., 2012. A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. J. Cogn. Neurosci. 24 (8), 1742–1752. http://dx.doi.org/10.1162/jocn_a_00233.

Earl, A., Albarracín, D., 2009. Nature, decay, and spiraling of the effects of fear-inducing arguments and HIV counseling and testing: a meta-analysis of the short- and long-term outcomes of HIV-prevention interventions. Health Psychol. 26 (4), 496–506. http://dx.doi.org/10.1037/0278-6133.26.4.496.Nature.

Etter, J.F., Vu Duc, T., Perneger, T.V., 2000. Saliva cotinine levels in smokers and nonsmokers. Am. J. Epidemiol. 151 (3), 251–258. http://dx.doi.org/10.1093/oxfordjournals.aje.a010200.

Falk, E.B., Berkman, E.T., Mann, T., Harrison, B., Lieberman, M.D., 2010. Predicting persuasion-induced behavior change from the brain. J. Neurosci. 30 (25), 8421–8424. http://dx.doi.org/10.1523/JNEUROSCI.0063-10.2010.

Falk, E.B., Berkman, E.T., Whalen, D., Lieberman, M.D., 2011. Neural activity during health messaging predicts reductions in smoking above and beyond self-report. Health Psychol. 30 (2), 177–185. http://dx.doi.org/10.1037/a0022259.

Falk, E., O'Donnel, M., Tompson, S., Gonzalez, R., Cin, S., Strecher, V., An, L., 2015. Functional brain imaging predicts public health campaign success. Soc. Cogn. Affect. Neurosci., 1–11. http://dx.doi.org/10.1093/scan/nsv108, *e-ahead of.*

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. NeuroImage 48 (1), 63–72. http://dx.doi.org/10.1016/j.neuroimage.2009.06.060.

Hammond, D., 2011. Health warning messages on tobacco products: a review. Tob. Control 20 (5), 327–337. http://dx.doi.org/10.1136/tc.2010.037630.

Jasinska, A.J., Chua, H.F., Ho, S.S., Polk, T.A., Rozek, L.S., Strecher, V.J., 2012. Amygdala response to smoking-cessation messages mediates the effects of serotonin transporter gene variation on quitting. NeuroImage 60 (1), 766–773. http://dx.doi.org/10.1016/j.neuroimage.2011.12.064.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. NeuroImage 62 (2), 782–790. http://dx.doi.org/10.1016/j.neuroimage.2011.09.015.

Kahnt, T., Heinzle, J., Park, S.Q., Haynes, J.-D., 2010. The neural code of reward anticipation in human orbitofrontal cortex. Proc. Natl. Acad. Sci. USA 107 (13), 6010–6015. http://dx.doi.org/10.1073/pnas.0912838107.

Kennerley, S.W., Behrens, T.E.J., Wallis, J.D., 2011. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat. Neurosci. 14 (12), 1581–1589. http://dx.doi.org/10.1038/nn.2961.

Klebanoff, M.A., Levine, R.J., Clemens, J.D., DerSimonian, R., Wilkins, D.G., 1998. Serum cotinine concentration and self-reported smoking during pregnancy. Am. J. Epidemiol. 148 (3), 259–262.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4. http://dx.doi.org/10.3389/neuro.06.004.2008, (November).

Marsden, P.V., 2002. Egocentric and sociocentric measures of network centrality. Soc. Netw. 24 (4), 407–422.

Mcnamee, D., Rangel, A., O'Doherty, J.P., 2013. Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. Nat. Neurosci. 16 (4), 479–485. http://dx.doi.org/10.1038/nn.3337.

Mead, E.L., Rimal, R.N., Ferrence, R., Cohen, J.E., 2014. Understanding the sources of normative influence on behavior: the example of tobacco. Social. Sci. Med. 115 (0), 139–143. http://dx.doi.org/10.1016/j.micinf.2011.07.011.Innate.

Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R. a., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. NeuroImage 59 (3), 2636–2643. http://dx.doi.org/10.1016/j.neuroimage.2011.08.076.

Mur, M., Bandettini, P.A., Kriegeskorte, N., 2009. Revealing representational content with pattern-information fMRI–an introductory guide. Soc. Cogn. Affect. Neurosci. 4 (1), 101–109, (http://doi.org/nsn044)(pii)(10.1093/scan/nsn044).

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A toolbox for representational similarity analysis. PLoS Comput. Biol. 10 (4), e1003553. http://dx.doi.org/10.1371/journal.pcbi.1003553.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., Panksepp, J., 2006. Self-referential processing in our brain—A meta-analysis of imaging studies on the self. NeuroImage 31 (1), 440–457. http://dx.doi.org/10.1016/j.neuroimage.2005.12.002.

O'Donnell, M.B., Falk, E.B., 2015. Big data under the microscope and brains in social context: integrating methods from computational social science and neuroscience. Ann. Am. Acad. Polit. Soc. Sci. 659 (1), 274–289. http://dx.doi.org/10.1177/0002716215569446.

Padoa-Schioppa, C., 2009. Range-adapting representation of economic value in the orbitofrontal cortex. J. Neurosci.: Off. J. Soc. Neurosci. 29 (44), 14004–14014. http://dx.doi.org/10.1523/JNEUROSCI.3751-09.2009.

Pegors, T.K., Kable, J.W., Chatterjee, A., Epstein, R.A., 2015. Common and unique representations in pFC for face and place attractiveness. J. Cogn. Neurosci. 27 (5), 959–973. http://dx.doi.org/10.1162/jocn.

Peters, G.-J.Y., Ruiter, R.A.C., Kok, G., 2012. Threatening communication: a critical reanalysis and a revised meta-analytic test of fear appeal theory. Health Psychol. Rev. 7, 1–24. http://dx.doi.org/10.1080/17437199.2012.703527.

Peters, G.-J.Y., Ruiter, R.A.C., Kok, G., 2013. Threatening communication: a critical reanalysis and a revised meta-analytic test of fear appeal theory. Health Psychol. Rev. 7 (sup 1), S8–S31. http://dx.doi.org/10.1080/17437199.2012.703527.

Pickett, K.E., Rathouz, P.J., Kasza, K., Wakschlag, L.S., Wright, R., 2005. Self-reported smoking, cotinine levels, and patterns of smoking in pregnancy. Paediatr. Perinat. Epidemiol. 19 (5), 368–376.

Pokorski, T.L., Chen, W.W., Bertholf, R.L., 1994. Use of urine cotinine to validate smoking self-reports in U.S. Navy recruits. Addict. Behav. 19 (4), 451–454.

Rosenstock, I.M., 1974. *The h*ealth belief model and preventive health behavior. Health Educ. Behav. 2 (4), 354–386.

Schmitz, T.W., Johnson, S.C., 2007. Relevance to self: a brief review and framework of neural systems underlying appraisal. Neurosci. Biobehav. Rev. 31 (4), 585–596.

Tompson, S., Lieberman, M., Falk, E., 2015. Grounding the Neuroscience of Behavior Change in the Sociocultural Context. *Current Opinion in Behavioral Sciences*.

Vogt, T.M., Selvin, S., Widdowson, G., Hulley, S.B., 1977. Expired air carbon monoxide and serum thiocyanate as objective measures of cigarette exposure. Am. J. Public Health 67 (0090–0036 (Print)), 545–549. http://dx.doi.org/10.2105/AJPH.67.6.545.

Wang, A.-L., Ruparel, K., Loughead, J.W., Strasser, A.A., Blady, S.J., Lynch, K.G., Langleben, D.D., 2013. Content matters: neuroimaging investigation of brain and behavioral impact of televised anti-tobacco public service announcements. J. Neurosci. 33 (17), 7420–7427. http://dx.doi.org/10.1523/JNEUROSCI.3840-12.2013.

Witte, K., Allen, M., 2000a. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ. Behav. 27 (5), 591–615. http://dx.doi.org/10.1177/109019810002700506.

Witte, K., Allen, M., 2000b. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ. Behav. 27 (5), 591–615.