Brief Communications

Predicting Persuasion-Induced Behavior Change from the Brain

Emily B. Falk,[1] **Elliot T. Berkman,**[1] **Traci Mann,**[2] **Brittany Harrison,**[1] **and Matthew D. Lieberman**[1]

[1]University of California, Los Angeles, Department of Psychology, Los Angeles, California 90095-1563, and [2]University of Minnesota, Department of Psychology, Minneapolis, Minnesota 55455-0344

Although persuasive messages often alter people's self-reported attitudes and intentions to perform behaviors, these self-reports do not necessarily predict behavior change. We demonstrate that neural responses to persuasive messages can predict variability in behavior change in the subsequent week. Specifically, an a priori region of interest (ROI) in medial prefrontal cortex (MPFC) was reliably associated with behavior change ($r = 0.49$, $p < 0.05$). Additionally, an iterative cross-validation approach using activity in this MPFC ROI predicted an average 23% of the variance in behavior change beyond the variance predicted by self-reported attitudes and intentions. Thus, neural signals can predict behavioral changes that are not predicted from self-reported attitudes and intentions alone. Additionally, this is the first functional magnetic resonance imaging study to demonstrate that a neural signal can predict complex real world behavior days in advance.

Introduction

People are exposed to an almost endless stream of persuasive messages each day: television and magazines are full of explicit advertisements, friends and family induce us to see things their way, and even educators fill their teachings with rhetorical devices meant to persuade. Commonsense suggests that when others report that their attitudes and intentions have changed in response to a persuasive message, actual behavior change is likely to follow. Indeed, self-reported attitudes and intentions have served as dominant constructs in psychological models used to predict behavioral responses to persuasive messages (Ajzen and Fishbein, 1980).

Although self-reported responses to persuasive communications predict some variability in future behavior, they do not predict it as accurately as scientists would like and nonscientists would expect (Webb and Sheeran, 2006); our word is only sometimes as good as our deed (LaPiere, 1934; Wicker, 1969). It is possible that implicit processes that are present during the receipt of persuasive communications but inaccessible to conscious awareness, as well as conscious processes that are simply not captured by summary ratings following message receipt, might explain variability in behavior change that is not explained by self-reported measures like attitudes or intentions. Neuroimaging methods, such as functional magnetic resonance imaging (fMRI), are uniquely positioned to assess processes that are introspectively opaque (Morris et al., 1998) with the added advantage that neural responses can be recorded in the moment that persuasion occurs, without re-

quiring behavioral assessments that are likely to contaminate the message recipient's natural responses by imposing a concurrent cognitive task. Furthermore, controlling for self-report variables, it is possible to determine what additional variance in behavior is accounted for by brain activity.

In this investigation, we explored the utility of fMRI in predicting and understanding behavior change in response to persuasive messages. More specifically, we measured neural activity in a priori regions of interest—medial prefrontal cortex (MPFC) and precuneus (Fig. 1a)—while people were exposed to persuasive messages regarding the value of regular sunscreen use. We then used those values to predict future behavior change in the same individuals (i.e., increased sunscreen use). We chose these neural regions to examine for both theoretical and data-driven reasons. These regions are reliably coactivated across a host of "self" processes (Kelley et al., 2002; Lieberman, 2007) and the extent to which people perceive persuasive messages to be self-relevant has long been thought to play a part in attitude and behavior change (McGuire, 1969; Briñol and Petty, 2005). Additionally, these regions have been previously observed in multiple studies of persuasion and attitude change (Chua et al., 2009; Jarcho et al., 2010). Finally, activity in these regions has been used to predict spontaneous motor behavior several seconds before reportable motor intentions formed (Soon et al., 2008) indicating that these regions may be involved in the formation of behavioral intentions that are not accessible to conscious self-report. We chose to measure changes in sunscreen use because successful persuasion-induced behavior change has been observed in this domain in several studies (Detweiler et al., 1999; Webb and Sheeran, 2006).

Materials and Methods

Participants

Twenty participants (10 female; mean age, 22.8 years; SD, 3.6 years) were recruited from the UCLA subject pool and through mass emails and
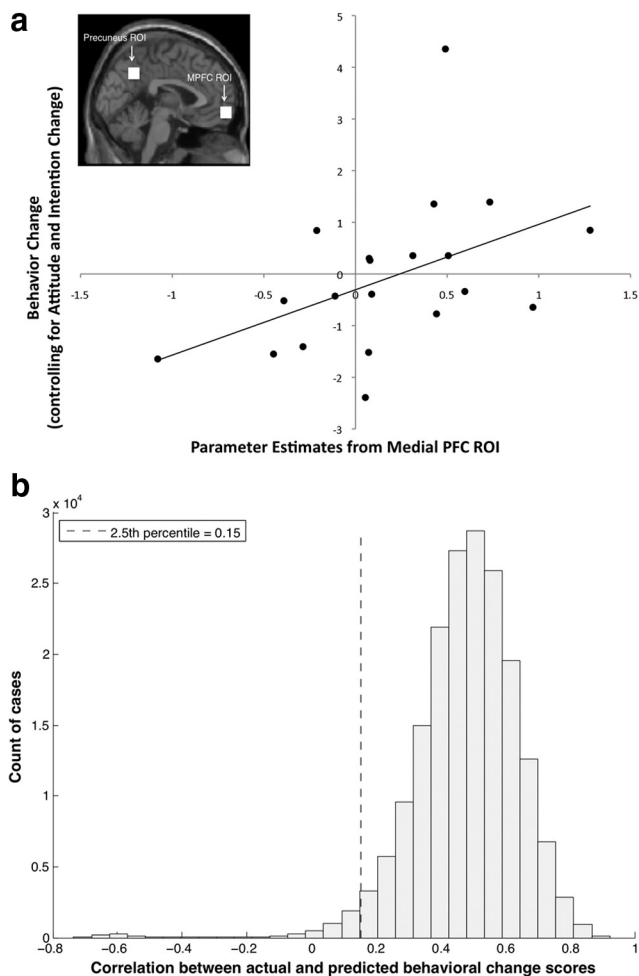
Figure 1. Neural prediction of behavior change. **a,** Scatterplot between activity in MPFC ROI and behavior change in the week following the persuasive messages (controlling for changes in self-reported attitudes and intentions) with an inset sagittal image displaying the location of the ROIs. **b,** Distribution of correlations between actual and predicted values of behavior change using a twofold cross-validation over 184,756 iterations (controlling for changes in self-reported attitudes and intentions).

posted fliers, and received either course credit or financial compensation for their participation. All participants were right-handed, European American, born and raised in the United States, spoke English as their first language, and met standard MRI safety criteria. Participants met the following criteria related to fMRI safety: (1) were not claustrophobic; (2) had no metal in their bodies (other than tooth fillings); (3) were not pregnant/breast-feeding. Potential participants were excluded if they were currently taking any psychoactive medication. Data were collected in accordance with the policies of the UCLA Institutional Review Board.

Materials
Slides pertaining to the importance of sunscreen use were created based on information from reputable organizations. The slides contained text and images designed to educate participants about the need to wear sunscreen every day, about proper application of sunscreen, about the relationship between sun exposure and skin cancer, as well as cosmetic reasons to protect the skin (see supplemental material, available at www.jneurosci.org).

Procedure
On day 1 of the experiment, before the scanning session, each participant indicated their sunscreen use over the prior week, their intentions to use sunscreen in the next week and their attitudes toward sunscreen. A number of distracter questions were also included relating to other activities.

Next, while in an fMRI scanner, participants viewed text and image-based persuasive slides from expert sources regarding sunscreen use. The text of the slides was also read aloud through earphones to control participant reading speed. Before the task, participants were instructed to read along silently, consider each slide carefully, and were told that they would be asked some questions about the slides later.

The primary task consisted of six blocks in which participants viewed slides about the need to wear sunscreen. Blocks ranged from 15 to 37 s, and were separated by a variable length fixation-cross rest period (1.8–3 s) with longer rest periods (15 s) before and after the main series of sunscreen messages. Following scanning, participants again indicated their attitudes toward sunscreen and their intentions to use sunscreen in the next week. Before leaving the scanning facility, participants were given a "thank you" bag that included sunscreen towelettes. Although most adults living in southern California own sunscreen, these towelettes were provided to ensure that all participants had easy access to sunscreen without having to purchase it. Participants were unaware on day 1 that we would be measuring their subsequent sunscreen use. One week following the scanner session, participants were contacted via E-mail and reported the number of days that sunscreen was used in the week following the scanner session.

Data acquisition and analysis
Behavioral data. Intention change was calculated as postscan minus prescan number of days that a participant indicated he or she intended to use sunscreen during the postscan week. Attitude change was calculated as postscan minus prescan feeling thermometer ratings toward sunscreen use. Behavior change was calculated as the number of days that sunscreen was reported to have been used in postscan minus prescan weeks.

fMRI imaging data acquisition parameters. Imaging data were acquired using a Siemens Allegra 3 Tesla head-only MRI scanner. Head motion was minimized using foam padding and surgical tape; goggles were also fixed in place using surgical tape connecting to the head coil and scanner bed. A set of high-resolution structural T2-weighted echo-planar images were acquired coplanar with the functional scans [spin-echo; repetition time (TR) = 5000 ms; echo time (TE) = 33 ms; matrix size = 128×128; 36 axial slices; field of view (FOV) = 20 cm; 3 mm thick; voxel size = $1.6 \times 1.6 \times 3.0$ mm]. One functional run was recorded (echo-planar T2-weighted gradient-echo, TR = 2000 ms, TE = 25 ms, flip angle = 90°, matrix size = 64×64, 36 axial slices, FOV = 20 cm, 3 mm thick; voxel size = $3.1 \times 3.1 \times 3.0$ mm).

fMRI data analysis. The fMRI data were analyzed using Statistical Parametric Mapping (SPM5). Images were realigned to correct for motion, normalized into standard stereotactic space, and smoothed with an 8 mm Gaussian kernel full-width half-maximum. The task was modeled for participants at the single subject level, comparing activity while watching persuasive messages to activity at rest. A random effects model was constructed, averaging over these single subject results at the group level.

Regions of interest. Regions of interest (ROIs) were constructed based on coordinates reported by Soon et al. (2008) in MPFC and precuneus, regions that also appeared in a study of persuasive messaging (Chua et al., 2009). To construct the MPFC ROI, MPFC coordinates reported by Soon et al. (2008) were averaged, and a 10 mm cube was constructed centered on the midline at the *y* and *z* coordinates reported by Soon et al. (2008) (0 60 −9). To construct the precuneus ROI, the same procedure was followed, centering the 10 mm cube on the midline at the *y* and *z* precuneus coordinate reported by Soon et al. (2008) (0 −57 39). Parameter estimates of activity, averaged across all voxels in each ROI, were extracted for each participant.

Cross-validation. To assess the predictive validity of the ROI estimates, we also conducted a twofold cross-validation (Stone, 1974). In this analysis, independent halves of the data were used for training and prediction purposes, respectively. For each of the possible combinations of 10 training and 10 test subjects (n = 184,756), a correlation was computed between the predicted behavior change scores of the test subjects (based on a regression model developed on the training subjects) and actual behavior change scores of the test subjects (Fig. 1*b*).
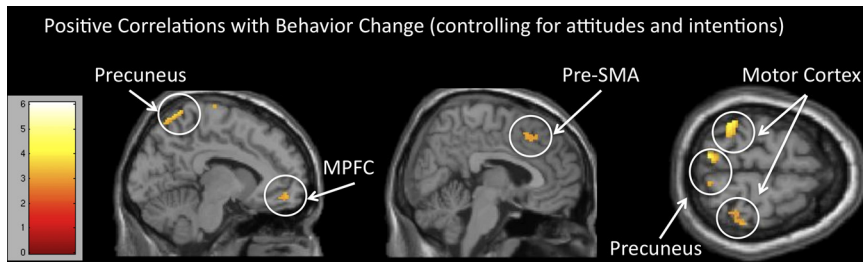
Figure 2. Regions associated with behavior change in a whole brain analysis. Significant correlations ($p < 0.005$; $k = 10$) were observed between behavior change, controlling for self-report measures, and MPFC, $r = 0.64$, $t_{cluster} = 3.52$; precuneus, $r = 0.71$, $t_{cluster} = 4.93$; pre-SMA, $r = 0.59$, $t_{cluster} = 3.10$; motor cortex, $r = 0.59$, $t_{cluster} = 3.10$. These regions were all observed by Soon et al. (2008) as predictors of spontaneous motor behavior, before and independent of consciously reportable behavioral intentions.

Whole brain searches. To explore neural regions that were associated with behavior change, but that we did not choose in our a priori ROI analysis, we regressed behavior change scores onto neural activity (during the presentation of the persuasive messages, compared with rest) at the group level. We performed this analysis both before and after controlling for self-reported attitudes and intentions.

Results

Behavioral data
Participants used sunscreen on more days in the postscan than the prescan week [mean prescan ($M_{prescan}$) = 1.10, $M_{postscan}$ = 1.65, $M_{increase}$ = 0.55, $t_{(19)}$ = 2.01, p = 0.059, d = 0.37]. This effect size is similar to other studies using persuasive messaging to affect sunscreen behavior (average d = 0.30) (Webb and Sheeran, 2006). Significance in this sample may be marginally significant due to the smaller sample sizes typical in fMRI experiments compared with behavioral experiments. Changes in actual sunscreen use were positively, but nonsignificantly, correlated with changes in intentions to use sunscreen (r = 0.17), and uncorrelated with attitude change scores (r = −0.08).

ROI analyses
Activity in the a priori MPFC ROI during the presentation of persuasive messages, compared with rest, was reliably associated with behavior changes in sunscreen use from pre- to postscan ($t_{(18)}$ = 2.30, r = 0.49, p = 0.030). After controlling for self-reported attitude and intention changes, activity in the MPFC ROI was still reliably associated with behavior change from pre- to postscan ($t_{(18)}$ = 2.20, r = 0.46, p = 0.041) (Fig. 1a). Activity in the precuneus ROI was positively, but not significantly, correlated with behavior change, both before and after controlling for attitude and intention change ($r_{uncontrolled}$ = 0.34, p = 0.147; $r_{controlled}$ = 0.33, p = 0.152). For complementary whole-brain analyses, see Figure 2 and supplemental material, available at www.jneurosci.org.

Cross-validation analysis
The association of activation in an a priori MPFC ROI at one time point with behavior at a later time point (controlling for attitudes and intentions) suggests that this neural activity has predictive power beyond self-report in the context of persuasive messaging. To explore the degree of predictive validity of our MPFC ROI, we conducted an iterative cross-validation on independent halves of the data. In each iteration, 10 "training" subjects were used to create a predictive regression model, which was then applied to the 10 remaining "test" subjects. In this analysis, the average correlation between actual behavior change scores in a set of test subjects, and their predicted behavior change scores based on that iteration's predictive regression model, was 0.48 across all

184,756 combinations of training and test subjects (Fig. 1b). Of these correlations, 99.3% were >0, suggesting that the correlation between actual behavior change and predicted behavior change of independent subsets is significantly likely to be positive (α < 0.05, two-tailed). This cross-validation analysis suggests that on average, approximately one quarter of the variability in behavior change scores in a new sample of 10 individuals can be inferred based on a regression equation constructed from the MPFC activity in the first 10 individuals, above and beyond their self-reported attitudes and intentions.

Whole brain searches
We performed an exploratory whole-brain search to find other neural regions that were associated with behavior change, above and beyond self-reported attitudes and intentions. In addition to finding significant relationships between behavior change (controlling for attitude and intention change) and activity in our hypothesized regions (MPFC and precuneus), we also found associations with activity in regions associated with considering the mental states of others (posterior superior temporal sulcus, pSTS; temporal parietal junction, TPJ; temporal pole), memory encoding (hippocampus), attention (supplementary motor cortex, inferior parietal cortex), visual imagery (occipital cortex), motor control and imitation (motor cortex) and affective experience (insula). A complete table of results for this analysis (as well as table of whole-brain searches exploring behavior change, intention change and attitude change separately), can be found in the supplemental material, available at www.jneurosci.org.

Discussion
Using an a priori ROI approach, we found that activity in MPFC was significantly related to persuasion-induced behavior change over the course of 2 weeks, above and beyond self-reported attitudes and intentions. Furthermore, in a cross-validation analysis, the average correlation between actual behavior change scores in a set of test subjects, and their predicted behavior change scores based on that iteration's predictive regression model, was 0.48 (Fig. 1b).

These findings are significant for at least two reasons. First, whereas some have suggested that time between self-reported persuasion and actual behavior-change allows other variables to intervene, thus limiting the maximum predictive utility of self-report measures at the time of message encoding (Randall and Wolff, 1994), our data suggests that the brain's real-time response to persuasive communications may be able to complement self-report measures and provide additional predictive capacity. As an initial proof of concept, by measuring MPFC activity in a preselected region while participants viewed persuasive messages, we were able to predict the behavioral efficacy of the persuasive messages above and beyond what participants' own self-reported attitude and intention change could predict. If MPFC responses reliably predict an additional quarter of the variance in persuasion-induced behavior over and above self-report, this would represent a major step forward in the predictive capacity of our persuasion models. It would also suggest that MPFC responses may represent psychological responses to persuasive communications that index future behavior while being either consciously inaccessible or not captured by summary reports of

attitudes and intentions following the receipt of a persuasive message. In either case, understanding the psychological bases of activity in this region during persuasion attempts may help to update the substance of our persuasion models as well. As noted earlier, this region has been associated with self-referential processing, but our ROI also overlaps with a more ventral portion of MPFC that has been associated with complementary process of implicit valuation (Kawasaki et al., 2001; Milne and Grafman, 2001), which could also play a key part in persuasion outcomes. Neither self-reference nor implicit valuation figure prominently in dominant models of persuasion; our results suggest it may be fruitful to investigate their potential role further.

Our results are also significant in light of a growing trend in neuroimaging studies, moving from statistical association to actual prediction and classification. Traditional fMRI studies typically use behavioral outcomes or self-reported experiences as regressors predicting responses in different brain regions. In these studies, observable outcomes are treated as known variables and their ability to account for variance in neural responses is assessed. Recent machine learning approaches have reversed this (Bandettini, 2009), using neural activity to predict, for instance, whether a participant is currently looking at a house or a face (Haxby et al., 2001). Soon et al. (2008) took this a step further by using neural activity to predict future, rather than concurrent, behavior. Using the findings of Soon et al. (2008) as a starting point, our results represent another order of magnitude in the scale of prediction. Soon et al. (2008) predicted behavior seconds before it occurred, while we were able to predict changes in behavior spanning weeks.

One potential caveat is that our measure of behavior change was self-reported rather than directly observed. Thus there is the possibility of self-report bias (Schwarz, 2007). Although direct observation of behavior would be preferred in future studies, our measurement of behavior through self-report is unlikely to have artificially enhanced our results. If a self-report bias were present, it would have likely led to an artificial increase in the correlation between self-reported intentions and self-reported behavior, as people are motivated to be consistent and follow through on their public commitments (Cialdini and Goldstein, 2004). Because we used neural responses to predict changes in behavior controlling for changes in attitudes and intentions, such a self-report bias would work against our ability to predict behavior changes from the ROIs. Thus, our findings represent a more conservative test of predictive ability, in light of this concern.

Of theoretical interest, in addition to our hypothesized regions of interest (MPFC, precuneus), we also observed correlations between behavior change and neural activity in regions involved in taking the perspective of others. This finding is consistent with prior research on the neural bases of persuasion (Falk et al., 2009), which emphasizes the role of social factors in the persuasion process (Cialdini and Goldstein, 2004). In addition, we observed increased activity in regions involved in memory encoding, attention, visual imagery, motor execution and imitation, and affective experience with increased behavior change. Combined with the primary results of this paper, these associations are consistent with theories of social learning and persuasion that posit that behavior change can result from encoding information about social norms, incorporating those norms into one's own self-concept, and planning to execute the relevant behaviors. Thus, our data do not suggest an exclusive role for MPFC

in predicting behavior, but rather provide a starting point for developing increasingly precise predictive models.

In conclusion, we believe these results represent an important step forward in using neural responses to predict the impact of persuasive messages in a way that complements and extends the use of self-report indicators. A full-scale research program identifying more ROIs in multisample cross-validation studies could usher in an era of renewed progress in persuasion models, both in terms of pure ability to predict behavior and greater insights into the subtle psychological processes that occur in response to persuasive messages and affect our future behavior.

References

Ajzen I, Fishbein M (1980) Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall.

Bandettini P (2009) What's new in neuroimaging methods? Ann N Y Acad Sci 1156:260–293.

Briñol P, Petty RE (2005) Individual differences in attitude change. Mahwah, NJ: Lawrence Erlbaum Associates.

Chua HF, Liberzon I, Welsh RC, Strecher VJ (2009) Neural correlates of message tailoring and self-relatedness in smoking cessation programming. Biol Psychiatry 65:165–168.

Cialdini RB, Goldstein NJ (2004) Social influence: compliance and conformity. Annu Rev Psychol 55:591–621.

Detweiler JB, Bedell BT, Salovey P, Pronin E, Rothman AJ (1999) Message framing and sunscreen use: gain-framed messages motivate beach-goers. Health Psychol 18:189–196.

Falk E, Rameson L, Berkman E, Liao B, Kang Y, Inagaki T, Lieberman M (2009) The neural correlates of persuasion: a common network across cultures and media. J Cogn Neurosci. Advance online publication. doi:10.1162/jocn.2009.21363.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.

Jarcho JM, Berkman ET, Lieberman MD (2010) The neural basis of rationalization: cognitive dissonance reduction during decision making. Soc Cognitive Affect Neurosci, in press.

Kawasaki H, Kaufman O, Damasio H, Damasio AR, Granner M, Bakken H, Hori T, Howard MA 3rd, Adolphs R (2001) Single-neuron responses to emotional visual stimuli recorded in human ventral prefrontal cortex. Nat Neurosci 4:15–16.

Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF (2002) Finding the self? An event-related fMRI study. J Cogn Neurosci 14:785–794.

LaPiere R (1934) Attitudes vs. actions. Social Forces 13:230–237.

Lieberman MD (2007) Social cognitive neuroscience: a review of core processes. Annu Rev Psychol 58:259–289.

McGuire WJ (1969) The nature of attitudes and attitude change. In: The handbook of social psychology (Lindzey G, Aronson E, eds), Reading, MA: Addison-Wesley.

Milne E, Grafman J (2001) Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. J Neurosci 21:RC150(1–6).

Morris JS, Ohman A, Dolan RJ (1998) Conscious and unconscious emotional learning in the human amygdala. Nature 393:467–470.

Randall D, Wolff J (1994) The time interval in the intention-behavior relationship: meta-analysis. Br J Social Psychol 33:405–418.

Schwarz N (2007) Retrospective and concurrent self-reports: the rationale for real-time data capture. In: The science of real-time data capture: self-reports in health research (Stone A, Shiffman S, Atienza A, Nebeling L, eds), pp 11–26. New York: Oxford UP.

Soon CS, Brass M, Heinze HJ, Haynes JD (2008) Unconscious determinants of free decisions in the human brain. Nat Neurosci 11:543–545.

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J R Stat Soc 36:111–147.

Webb TL, Sheeran P (2006) Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 132:249–268.

Wicker A (1969) Attitudes versus actions: the relationship of verbal and overt behavioral responses to attitude objects. J Social Issues 25:41–78.