OXFORD

Social Cognitive and Affective Neuroscience, 2017, 1902–1915

doi: 10.1093/scan/nsx126
Advance Access Publication Date: 11 November 2017
Original article

The persuasion network is modulated by drug-use risk and predicts anti-drug message effectiveness

Richard Huskey,[1] J. Michael Mangus,[2] Benjamin O. Turner,[3] and René Weber[2]

[1]School of Communication, Cognitive Communication Science Lab, The Ohio State University, OH 43210, USA, [2]Department of Communication, Media Neuroscience Lab, University of California, Santa Barbara, CA 93106, USA, and [3]Wee Kim Wee School of Communication and Information, Nanyang Technological University, 637718, Singapore

Correspondence should be addressed to René Weber, Department of Communication, University of California Santa Barbara, 4405 Social Sciences & Media Studies, Santa Barbara, CA 93106-4020, USA. E-mail: renew@comm.ucsb.edu.

Abstract

While a persuasion network has been proposed, little is known about how network connections between brain regions contribute to attitude change. Two possible mechanisms have been advanced. One hypothesis predicts that attitude change results from increased connectivity between structures implicated in affective and executive processing in response to increases in argument strength. A second functional perspective suggests that highly arousing messages reduce connectivity between structures implicated in the encoding of sensory information, which disrupts message processing and thereby inhibits attitude change. However, persuasion is a multi-determined construct that results from both message features and audience characteristics. Therefore, persuasive messages should lead to specific functional connectivity patterns among a priori defined structures within the persuasion network. The present study exposed 28 subjects to anti-drug public service announcements where arousal, argument strength, and subject drug-use risk were systematically varied. Psychophysiological interaction analyses provide support for the affective-executive hypothesis but not for the encoding-disruption hypothesis. Secondary analyses show that video-level connectivity patterns among structures within the persuasion network predict audience responses in independent samples (one college-aged, one nationally representative). We propose that persuasion neuroscience research is best advanced by considering network-level effects while accounting for interactions between message features and target audience characteristics.

Key words: persuasion; functional connectivity; elaboration likelihood model; fMRI; public service announcements

Introduction

By most accounts, the modern scientific investigation of persuasion dates back to the Second World War when scholars such as Lasswell, Hovland and Lazarsfeld sought to understand how propaganda influenced attitudes and motivated behavior change (Rogers, 1994). Since, eight decades of research have yielded a nuanced understanding of the mechanisms that underlie the persuasion process. Indeed, a large body of work has illuminated how structural message characteristics and argumentation contribute to message persuasiveness among different audiences (e.g. Petty and Cacioppo, 1986).

However, only recently have investigators sought to understand the neural basis of persuasive message processing. One of the earliest studies used functional magnetic resonance imaging (fMRI) to evaluate the ways in which American and Korean audiences evaluated text- and video-based persuasive messages (Falk *et al.*, 2009). The results demonstrated that structures commonly implicated in social, affective and executive

Received: 6 July 2017; **Revised:** 21 August 2017; **Accepted:** 16 October 2017

processing are recruited while evaluating persuasive messages. Today, the literature has developed such that a structural consensus is emerging. A synthesis of 20 persuasion-neuroscience studies implicated structures involved in reward processing, self-reflection, social pain, executive processing, and salience detection as key components of the so-called *persuasion network* (Kaye *et al.*, 2016).

Despite reference to a persuasion network, the majority of work in this domain has in fact ignored network structure and has instead used the traditional mass-univariate approach that addresses questions of how much activity is elicited within different brain regions (but not how those regions interact with one another; Friston, 1994; 2011).

However, cognitive neuroscience has increasingly come to recognize that a complete understanding of how psychological processess (including persuasion) are instantiated in the brain requires studying the brain at the network level (Bassett and Gazzaniga, 2011; Bullmore and Sporns, 2012; Petersen and Sporns, 2015; Weber *et al.*, 2015; Bassett and Sporns, 2017; Cooper *et al.*, 2017). Accordingly, prominent calls have forcefully advocated that our understanding of the neural basis of psychological processes can only advance by moving beyond 'blobology' (e.g. Poldrack and Farah, 2015) and adopting a networked perspective. Applying this perspective to the persuasion network requires analytical procedures that illuminate the task-modulated relationship between one network structure and another (Friston, 2011). To date, a small handful of persuasion neuroscience studies have addressed this issue using psychophysiological interaction analyses (PPI; Friston *et al.*, 1997). Among these, two (not necessarily incompatible) functional views of the network connectivity patterns that underpin persuasion are beginning to emerge. Each is briefly summarized below.

Affective-executive hypothesis

Ramsay *et al.* (2013) proposed a mechanism where evaluations of persuasive messages are modulated by affective and executive processing. In their study, among both non-drug and drug users, strong argument strength (AS) persuasive messages yielded increased functional connectivity between left inferior frontal gyrus (lIFG, a putatively 'executive' region) and the amygdala (AMY, a putatively 'affective' region) and insula compared to weak AS messages. A related study among smokers observed a negative relationship between structures commonly implicated in affective and executive processing in response to aversive anti-smoking (and therefore self-relevant) images compared to aversive but not self-relevant images (Dinh-Williams *et al.*, 2014). Specifically, a seed region of interest (ROI) in the lIFG showed negative connectivity with the anterior cingulate cortex (ACC), medial prefrontal cortex (MPFC), precuneus, and insula. This connectivity pattern suggests that issue involvement modulates the way in which audiences process a persuasive message, such that smokers engage in defensive processing in an attempt to down regulate negative feelings associated with a threatening message. Unfortunately, this study did not include a non-smoking comparison group so it is difficult to fully integrate this finding with the study by Ramsay *et al.* Nevertheless, these two results account for key variables known to modulate persuasion, specifically AS and audience issue-involvement (Petty and Cacioppo, 1986) and show that each variable contributes to different neural processing patterns.

Similarly, a study by Cooper *et al.* (2017) examined connectivity patterns in sedentary individuals in response to (individually relevant) messages promoting physical activity and found that connectivity strength between a subregion within the MPFC and ventral striatum (VS) and AMY during message exposure predicted decreases in sedentary behavior. Taken together, these three studies provide converging evidence that affective-executive processing underlies attitude and behavior change.

Encoding-disruption hypothesis

The previous studies evaluated variables such as AS and issue involvement. At the same time, message sensation value (MSV; a measure of message arousingness or intensity) has also been hypothesized as a mechanism that can disrupt cognitive processing, thereby limiting the extent to which a persuasive message is encoded and ultimately recalled (e.g. Lang, 2006). Experimental fMRI work has shown that low-compared to high-MSV persuasive ads are better remembered, elicit increased activity in temporal and prefrontal cortex, and correspond to lower levels of occipital cortex activation (Seelig *et al.*, 2014).

The same study also tested the hypothesis that increases in MSV disrupt message encoding and therefore should result in decreased functional connectivity between structures implicated in sensory processing and message encoding. The results were ambiguous in that a negative relationship was observed between inferior lateral occipital cortex (iLOC) and the middle temporal (MTG) and inferior frontal gyrus (IFG) regardless of variation in MSV, although, this negative relationship was weaker when MSV was low. One possible explanation for this finding is that audience characteristics modulate audience reactions to MSV. For instance, Wang *et al.* (2014) show that increases in MSV lead to approach tendency among high-risk audiences, whereas low-risk audiences show avoidance responses.

The present study

As these studies indicate, and a large body of theoretical and empirical evidence demonstrates, issue involvement modulates persuasive message processing (Petty and Cacioppo, 1986). The literature shows that being at-risk for drug use constitutes a form of issue involvement where those that are at higher-risk for using a drug hold stronger counter-attitudinal views when presented with an anti-drug message compared to those that are at lower-risk (Cappella *et al.*, 2003). This increases the probability that at-risk, and therefore issue-involved, audiences engage in defensive message processing (i.e. counterarguing) when presented with a counter attitudinal message. This defensive processing makes it difficult to use self-report measures to identify which messages are likely to result in attitude or behavior change (Weber *et al.*, 2013, 2014). Identifying what message features contribute to attitude change among this audience is critically important.

Unfortunately, our understanding in this area is incomplete. Of the four fMRI PPI studies to-date, two only evaluate at-risk participants (Dinh-Williams *et al.*, 2014; Cooper *et al.*, 2017). The remaining two studies (Ramsay *et al.*, 2013; Seelig *et al.*, 2014) include both not-at-risk and at-risk participants in their samples, however, they do not consider the different ways each risk group processes persuasive messages. Furthermore, each of these studies has focused narrowly on just one aspect of persuasive messages (e.g. MSV, AS).

Importantly, theoretical and empirical evidence demonstrates that each of these variables interacts with issue involvement which leads to different message evaluations (Weber *et al.*, 2013). These different evaluations correspond to differences in neural processing (Weber *et al.*, 2014) with real-world implications. Specifically, neural activity in the middle frontal gyrus (MFG) and superior temporal gyrus (STG) among issue-involved audiences is predictive of message evaluations among independent samples.

Moreover, not one of these four studies has been subject to empirical replication. In this study, we attempt to replicate and extend the encoding-disruption and affective-executive hypotheses by examining the degree to which they are influenced by audience issue involvement. Should the affective-executive hypothesis be true, we expect that a seed ROI in the MPFC will show network connections with AMY and VS (Cooper *et al.*, 2017) while a seed ROI in the IFG should show connectivity with AMY, ACC, MPFC, precuneus, and insula (Ramsay *et al.* 2013; Dinh-Williams *et al.* 2014), particularly when argument strength or self-relevance is high. By comparison, the network model proposed by the encoding-disruption hypothesis indicates that a seed ROI in the iLOC should show stronger connectivity with the MTG and IFG when MSV is low.

Accordingly, our study interrogates network models proposed in the literature while extending these models to include seed ROIs that have shown to differ by issue involvement (i.e. MFG, STG; Weber *et al.*, 2014).

This extension is motivated by the fact that we currently lack a firm understanding of how message and audience characteristics modulate functional connectivity within the persuasion network and how this network connectivity contributes to message evaluation, let alone attitude or behavior change. Further, and as others have pointed out elsewhere, much of the persuasion neuroscience literature is data-driven in nature (Vezich *et al.*, 2016) which adds additional complications when trying to map neural processing to specific cognitive functions. To address these gaps, we draw on the Elaboration Likelihood Model (ELM), which argues that message features (i.e. MSV, AS) interact with issue-involvement to modulate persuasive message processing (Petty and Cacioppo, 1986). To test these relationships, subjects at high- and low-risk for drug use were exposed to anti-drug public service announcements (PSAs) that varied systematically along MSV and AS dimensions. We show that (i) neural processing among high- and low-risk audiences corresponds to distinct network connectivity patterns and (ii) that these patterns provide support for the affective-executive hypothesis but not also for the encoding-disruption hypothesis. Furthermore, we modify the brain-as-predictor approach (Berkman and Falk, 2013) to identify how brain-network patterns for individual PSAs predict of message perceptions. Results from this network-as-predictor analysis demonstrate that variation in connectivity at the individual video level is predictive of perceived anti-drug messages' effectiveness in two independent samples (one college age, one nationally representative), further suggesting that there is utility in examining the neural underpinnings of persuasion at the network level.

Methods

Previous reporting and general method

The subjects, materials, measures, experimental procedures, functional magnetic resonance imaging (fMRI) scanner parameters, and pre-processing pipeline used in this study are reported by Weber *et al.* (2014). Briefly, the imaging data reported herein were pre-processed using FEAT (fMRI Expert Analysis Tool v6.0) from the Oxford Center for Functional MRI of the Brain (FMRIB) Software Library (FSL v5.0) using a standard pipeline (Weber *et al.*, 2015). The study adopted a 2x2x2 mixed factorial design which manipulated AS (high/low) and MSV (high/low) as within-subjects factors and subject drug-use risk (high/low) as between-subjects factors. A total of 28 subjects, split evenly among high/low drug-use risk (assessed using the risk for marijuana use scale; Cappella *et al.*, 2003), were exposed to 32 30-second anti-drug PSAs (drawn from the Annenberg School for Communication at the University of Pennsylvania antidrug PSA archive) and eight control conditions; four of which featured PSAs where the video and audio had been scrambled (thereby preserving luminosity, auditory levels, and low-level visual features such as cuts, but removing semantic meaning), and four that showed a black screen with no visual or auditory features. PSAs and control conditions were split evenly between two runs in a counterbalanced order.

Subjects in the fMRI sample (henceforth referred to as MRIS) were drawn from the undergraduate research pool at University of California, Santa Barbara M(s.d.)age = 20.3(1.5). Subjects did not show any contraindication to fMRI scanning. All procedures conformed to the Declaration of Helsinki and were approved by the University's Institutional Research Board. Subjects watched each PSA while undergoing fMRI scanning. Participants provided a continuous response measure (CRM; not reported in this manuscript) of message convincingness (not at all convincing—very convincing) during the scanning session. After the fMRI scan was complete, subjects were then exposed to the PSAs for a second time in a session that took place outside of the scanner. Self-report measures were collected during this session. Subjects rated each PSA on perceived argument strength (pAS; Zhao *et al.*, 2011) and perceived message sensation value (pMSV; Kang *et al.*, 2006).

Self-reported ratings of message effectiveness were evaluated using the perceived message effectiveness (PME) scale, which has been shown to correlate highly with actual PSA effectiveness and behavior change (Dillard *et al.*, 2007a,b; Bigsby *et al.*, 2013). These PME ratings were drawn from two independent studies where subjects rated PME for each PSA. The first sample (henceforth referred to as IS1) was among 599 college students M(s.d.)age = 19.55(2.12), 73.8% female (Weber *et al.*, 2013). The second sample (IS2) was nationally representative of adolescents in the United States (Kang *et al.*, 2006). This sample (n = 601, Mage = 15.30, range 12–18) was balanced across males and females.

PPI analysis

Whole-brain functional connectivity patterns for a set of structures in the persuasion network were evaluated using a series of PPI analyses (Friston *et al.*, 1997). As discussed above, the seed ROIs used in this study were selected a priori based on the empirical literature. Seed ROIs were constructed in FSL by drawing a binary 5 mm sphere (in MNI152 space) around coordinates previously reported in the literature and included the medial prefrontal cortex (MPFC, −4, 56, −4; Falk *et al.*, 2016), middle frontal and superior termporal gyri (MFG, 38, 30, 34; STG, 56, 4, −18; Weber *et al.*, 2014), IFG (−46, 28, 12; Ramsay *et al.*, 2013), and inferior lateral occipital cortex (iLOC, −32, −70, −16; Seelig *et al.*, 2014). Time series data (PHYS) were extracted from filtered functional data for each subject for each run within each seed ROI by taking the arithmetic mean of all voxels within the

corresponding binary mask. First-level PPI general linear models (GLMs) were then fit separately using each seed ROI. Following established guidelines (Cohen *et al.*, 2003), each first-level GLM included explanatory variables (EVs) that encoded experimental conditions (i.e. MSV High > MSV Low, AS High > AS Low), control conditions (C1 = scrambled, C2 = black screen), a PHYS EV, and two- and three-way condition by PHYS interaction terms (equation 1).

$$PPI = \beta_1 MSV + \beta_2 AS + \beta_3 MSVxA_S + \beta_4 PHYS + \beta_5 MSVxPHYS + \beta_6 ASxPHYS + \beta_7 MSVxASxPHYS + \beta_8 C1 + \beta 9C2 + \beta_{10} C1xPHYS + \beta_{11} C2xPHYS + \beta_0 + \epsilon. \quad (1)$$

Following a hybrid generalized PPI approach (gPPI; McLaren *et al.*, 2012), contrasts were constructed between interaction EV coefficients (e.g. AS × PHYS > C1 × PHYS). Importantly, while the full model was estimated, we deliberately do not consider the MSV × AS × PHYS > C1 × PHYS contrast. The reason is that the sign of a three-way interaction does not have the same straightforward interpretation as a two-way interaction (where a positive sign indicates an increase and a negative sign a decrease in connectivity). This renders any contrast of a three-way interaction EV coefficient nearly uninterpretable against a two-way interaction EV coefficient. Still, the analysis and results are appropriate given that we are replicating/extending existing hypotheses (i.e. affective-executive, encoding-disruption) and their analytical paradigms.

A second-level analysis then combined runs for each subject in a fixed effects model. Finally, the interaction terms were analyzed in a series of third-level mixed effects models (FLAME 1; Beckmann and Smith, 2004; Woolrich *et al.*, 2004). Subject drug-use risk (high/low) was specified as two groups in FSL's FEAT tool (which lets variance estimates differ between groups), and then encoded using indicator variables. Contrasts at the third-level assessed mean functional connectivity by risk group, as well as in the high-risk > low-risk and low-risk > high-risk contrasts. A cluster-based procedure was applied to correct for multiple comparisons (Worsley, 2001) with a cluster defining threshold of $Z = 2.3$ and a cluster extent threshold of $P < .05$. All structural localizations were identified using FSL's probabilistic atlases and cross referenced with the Neurosynth database (Yarkoni *et al.*, 2011).

Network-as-predictor analysis

The previous section described a classic PPI analysis where task-modulated neural activity in a seed ROI is fit to every other voxel in the brain. Our network-as-predictor analyses had different aims, and a different analytical approach. Here, we wanted to identify network connectivity patterns specific to each PSA and the degree to which variation in these patterns was predictive of PME in two independent samples. To do this, we modified the brain-as-predictor analytical paradigm (Berkman and Falk, 2013). Whereas earlier brain-as-predictor studies used signal change in seed ROIs for each video (e.g. Falk *et al.*, 2012; Weber *et al.*, 2014) to predict message effectiveness in independent samples, we used connectivity parameter estimates (PEs) between seed and target ROIs. Our seed ROIs were the same as those reported above in the PPI Analysis section while target ROIs for the encoding-disruption and affective-executive hypotheses were selected based on previous empirical results and therefore represent connectivity patterns that have been reported in the literature. A full list of seeds and their corresponding target ROIs can be found in Table 1.

Table 1. Seed and Target ROIs used in the network-as-predictor analysis

Structure	Laterality	Coordinates
Middle frontal gyrus	Right	38, 30, 34
Superior parietal lobe	Left	−40, −48, 48
Angular gyrus	Right	46, −46, 38
Inferior parietal lobe	Left	−44, −52, 52
Superior lateral occipital cortex	Left	−30, −66, 42
Superior lateral occipital cortex	Right	34, −62, 50
Supramarginal gyrus	Left	−56, −28, 40
Superior parietal lobe	Medial	0, −74, 52
MFPC	Left	−4, 56, −4
Dorsomedial prefrontal cortex	Left	−10, 52, 44
Posterior superior temporal sulcus	Left	−56, −30, 8
Posterior superior temporal sulcus	Right	−68, −14, −6
Temporal pole	Left	−56, 8, −18
Temporal Pole	Right	60, 4, −16
Ventrolateral prefrontal cortex	left	−52, 20, −2
Inferior frontal gyrus seed	Left	−46, 28, 12
Amygdala	Left	−26, −2, −18
Lateral occipital cortex	Right	48, −68, −16
Temporal occipital cortex	Left	−42, −54, −22
Inferior lateral occipital cortex	Left	−32, −70, −16
Inferior frontal gyrus	Left	−46, 24, 0
Middle temporal gyrus	Left	−54, −26, −4
Middle temporal Gyrus	Right	44, −32, −2
Superior temporal gyrus	Right	56, 4, −18
Posterior cingulate cortex	Right	10, −**56**, 32
Precuneus	Right	8, −54, 48
Primary somatosensory cortex	Right	38, −34, 52
Superior parietal lobe	Right	28, −52, 58

Notes: Seed ROIs are shown in **bold** and target ROIs are listed. A 5 mm binary spherical mask was drawn around each seed and target ROI. Coordinates are shown in MNI152 space.

Said differently, this seed-target ROI-based analysis examined how each PSA video modulated functional connectivity among *a-priori* defined structures within the persuasion network (for subjects in the MRIS sample) and the extent to which this variation was predictive of PME-ratings for each PSA in two large independent samples (IS1, IS2). To that end, a series of gPPI models were constructed for each of the seed ROIs defined above. The first-level models (one model for each subject for each run for each seed ROI) included EVs that encoded each PSA video and control condition (C1 = scrambled, C2 = black screen), a PHYS EV, and then interaction terms (Equation 2). Contrasts at the first-level followed a gPPI logic and were defined as: VideoN × PHYS > C1 × PHYS.

$$PPI = \beta_1 Video1 + \beta_2 Video2 + \ldots + \beta_{16} Video16 + \beta_{17} C1 + \beta_{18} C2 + \beta_{19} PHYS + \beta_{20} Video1xPHYS + \beta_{21} Video2xPHYS + \ldots + \beta_{35} Video16xPHYS + \beta_{36} C1xPHYS + \beta_{37} C2xPHYS + \beta 0 + \epsilon. \quad (2)$$

These contrasts of parameter estimates were then carried forward into two fixed-effects second-level analyses (one analysis for each run). At the second-level, indicator variables encoded subject drug-use risk (high/low). Planned contrasts evaluated functional connectivity for each PSA by risk group. A 5 mm binary spherical mask was drawn for each a-priori defined target ROI. Finally, FSL's Featquery tool was used to extract median functional connectivity parameter estimates (PEs) between seed and target ROIs for each PSA for each of the second-level

contrasts. These parameter estimates represent the strength of functional connectivity between key structures in the persuasion network for each risk group, as modulated by each PSA.

We then computed a zero order correlation matrix to examine the relationship between PEs and PME ratings in our two independent samples for each risk group. We considered PEs that were significantly correlated with out-of-sample PME ratings as promising candidates in our network-as-predictor regression models. The actual prediction regression models followed the analytical logic established in (Falk *et al.*, 2012; Weber *et al.*, 2014) with one notable change: instead of transforming all PEs and PME evaluations into ranks for each PSA and risk group, we used the original unranked continuous data in our prediction models (making sure that no outliers were present). Unranked data did not lead to any substantial differences (compared to ranked data) and in a few cases, our modified procedure was slightly more conservative (that is, it led to lower prediction accuracies). As in (Falk *et al.*, 2012; Weber *et al.*, 2014) our network-as-predictor analysis employed a step-wise regression procedure. In a first step we estimated baseline regression models which included the self-reported PSA evaluations in our MRIS sample that are known to be predictive from traditional persuasion theory (pMSV, pAS, and the pMSV × pAS interaction). All self-report predictors were group mean centered to reduce bias in predictor coefficients due to the inherent multicollinearity between main- and interaction-effect predictors. We then included the brain network PEs as predictors in a second step. This procedure allowed us to identify the network connectivity PEs that increased PME prediction accuracy above and beyond self-reported PSA evaluations.

Results

The primary aims of this study were to evaluate (i) differences in functional connectivity patterns between risk groups when responding to anti-drug PSAs and (ii) whether these patterns provide support the affective-executive and encoding-disruption hypotheses, and the degree to which network connectivity patterns unique to each PSA are predictive of PME ratings in two independent samples. We report on the whole-brain PPI results as they relate to the affective-executive and encoding-disruption hypotheses before turning to the network-as-predictor results.

PPI results

In this study, we constructed contrasts that encoded between-group comparisons as well as mean within-group connectivity patterns. With the exception of just one case, between-group comparisons did not survive statistical thresholding. Therefore, primarily qualitatively different connectivity patterns emerged by risk-group. Said differently, unless otherwise specified, differences reported are within risk-group and do not represent the more stringent between risk-groups contrast. However, it is worth noting that our analytic decision to contrast connectivity patterns in each risk group by an active control, and then contrast the resulting group-level PEs against each other is conservative and makes it more difficult to detect group-level differences. A less-conservative approach would have been to contrast group-level connectivity patterns without also accounting for an active baseline.

However, this would make it more difficult to distinguish between functional connections simply associated with watching a PSA (and are therefore driven by low-level audiovisual features), and those that account for the semantic content of a PSA. For clarity, we split the reporting of our PPI results according to the different hypotheses they address.

Affective-executive hypothesis. This hypothesis argued that differences in AS should modulate functional connectivity between structures implicated in affective and executive processing, and that these connectivity patterns should differ by issue-involvement. Accordingly, and consistent with existing literature, the results presented in this subsection correspond to the AS × PHYS > C1 × PHYS contrast. When seeding from the MPFC (Table 2, Figure 1) among high-drug risk subjects, the MPFC showed functional connections with the dorsoanterior insula, central operculum, temporal pole, putamen (dorsal striatum) and orbitofrontal cortex (OFC). For subjects in the low-drug risk group, the MPFC showed connections with the inferior parietal lobe (IPL), primary somatosensory cortex, supramarginal gyrus, superior frontal gyrus, and pre-motor cortex.

For low-risk subjects, the MFG (Table 3, Figure 2) showed bilateral connectivity with the superior lateral occipital cortex (sLOC), IPL, and anterior intra-parietal sulcus. Results did not survive thresholding among high-risk subjects. Similarly, significant results were not observed for the STG seed for either risk-group. Lastly, we were unable to replicate previous findings (Ramsay *et al.*, 2013; Dinh-Williams *et al.*, 2014) as significant results were not observed when seeding from the IFG seed for either the high- or low-drug risk groups.

Encoding-disruption hypothesis. When seeding from the iLOC (Table 4, Figure 3) for the MSV × PHYS > C1 × PHYS contrast, both high- and low-drug risk subjects showed connections with MTG, superior lateral occipital cortex, OFC, and superior frontal gyrus. In general, these iLOC connections with temporal cortices do not conform to the encoding-disruption hypothesis (Seelig *et al.*, 2014) which predicts that high MSV should lead to lower connectivity between occipital and temporal cortices.

Furthermore, significant group differences were observed for the high- > low-drug risk contrast. High-drug risk subjects showed iLOC connectivity with the central operculum, amygdala, dorsal striatum (both putamen and caudate nucleus), and MFG.

Network-as-predictor results

Of the video-wise seed-target connectivities (PEs) evaluated in this study, four were significantly correlated with PME in our independent samples (Table 5). In the stepwise regressions, which adjusted each PE for both the effect of the self-report baseline predictors (pMSV, pAS, and pMSV x pAS) and each of the other PEs in the model, only MFG-SPL PEs significantly improved out-of-sample prediction accuracy. Therefore, we only report these results below.

The baseline models using self-reported PSA evaluations in our MRIS sample to predict PSA PME in two independent large samples (IS1 and IS2) were significant. For IS1 the prediction accuracies were $R2 = 23.9\%$ ($F = 4.24$, $P = 0.014$; all accuracies are reported as adjusted $R2$) for the high-drug risk group and $R2 = 49.9\%$ ($F = 11.31$, $P < 0.001$) for the low-drug risk group. For IS2, the accuracies were $R2 = 36.6\%$ ($F = 6.95$, $P < 0.001$) for the high-drug risk group and $R2 = 62.7\%$ ($F = 18.35$, $P < 0.001$) for the low-drug risk group. None of the models violated any central assumptions of the general linear model.

As described above, adding the MFG-SPL PEs to the baseline model in a stepwise regression increased out-of-sample

Table 2. Psychophysiological interaction results when seeding from the medial prefrontal cortex

Structure	Laterality	Cluster size	Max Z-score	Coordinates
ASxPhys > CtrlxPhys				
High-risk subjects				
Dorsoanterior insula	Right	1930	3.73	38, 8, 4
Temporal pole	Right		3.65	56, 8, 0
Putamen (dorsal striatum)	Right		3.46	28, −6, 4
Frontal operculum	Right		3.26	42, 18, 0
Frontal operculum	Right		3.26	42, 24, 0
Dorsoanterior insula	Right		3.25	44, 8, −4
Putamen	Left	1781	3.46	−32, 2, 2
Orbitofrontal cortex	Left		3.45	−44, 20, −8
Central operculum	Left		3.32	−46, 6, 0
Central operculum	Left		3.17	−48, 0, 4
Low-risk subjects				
Inferior parietal lobe	Right	2451	3.68	58, −34, 44
Primary somatosensory cortex	Right		3.43	36, −34, 42
Supramarginal gyrus	Right		3.38	38, −38, 42
Supramarginal gyrus	Right		3.37	62, −24, 30
Superior frontal gyrus	Right		3.17	20, 6, 68
Premotor cortex	Right		3.02	30, −20, 66
Inferior parietal lobe	Left	1096	3.87	−60, −40, 36
Inferior parietal lobe	Left		3.86	−62, −30, 32
Inferior parietal lobe	Left		3.86	−62, −36, 34
Supramarginal gyrus	Left		3.58	−68, −28, 26
Supramarginal gyrus	Left		3.43	−66, −30, 40
MSVxPhys > CtrlxPhys				
High-risk subjects				
Central operculum	Left	1017	3.67	−50, −2, 8
Precentral gyrus	Left		3.61	−52, 4, 10
Dorsoanterior insula	Left		3.39	−32, 20, 4
Precentral gyrus	Left		3.05	−60, 4, 4
Central operculum	Left		2.99	−48, −6, −4
Putamen (dorsal striatum)	Left		2.96	−32, 0, 0
Low-risk subjects				
Inferior parietal lobe	Right	3063	4.61	58, −36, 42
Supramarginal gyrus	Right		4.01	62, −24, 32
Inferior parietal lobe	Right		3.86	54, −50, 44
Inferior parietal lobe	Right		3.84	48, −46, 50
Primary somatosensory cortex	Right		3.43	46, −32, 48
Anterior intra-parietal sulcus	Right		3.35	40, −42, 42
Ventrolateral prefrontal cortex	Right	1579	3.99	40, 58, 2
Dorsolateral prefrontal cortex	Right		3.84	38, 58, 12
Dorsolateral prefrontal cortex	Right		3.66	40, 52, 18
Ventrolateral prefrontal cortex	Right		3.38	46, 54, −4
Orbitofrontal cortex	Right		3.37	40, 48, −10
Orbitofrontal cortex	Right		3.34	22, 36, −18
Inferior parietal lobe	Left	1480	4.06	−62, −36, 32
Supramarginal gyrus	Left		4.02	−62, −30, 40
Anterior intra-parietal sulcus	Left		2.94	−42, −44, 44

Notes: Results are cluster corrected for multiple comparisons with the cluster defining threshold set at $Z = 2.3$ and a cluster extent threshold of $P < 0.05$ and coordinates are shown in MNI152 space. Structures with a value in the cluster size column represent the peak Z-score while all others correspond to local maxima within the cluster.

prediction accuracy. Specifically, for the high risk-drug group, functional connectivity between MFG and the SPL increased the prediction accuracy in IS1 from $R2 = 23.9\%$ to $R2 = 35.1\%$ ($F = 5.18$, $P = 0.003$), an increase of 11.2% points in this traditionally difficult-to-predict target group. In IS2, the prediction accuracy increased from $R2 = 36.6\%$ to $R2 = 43.3\%$ ($F = 6.93$, $P < 0.001$) which is an increase of 6.7% points. In contrast, in the low-drug risk group, none of the functional connectivity PEs was able to increase the out-of-sample prediction accuracy significantly beyond self-report data, for either IS1 or for IS2. These results are summarized in Figure 4.

Discussion

In order to advance the persuasion neuroscience literature, the analyses reported in the present study were designed to identify

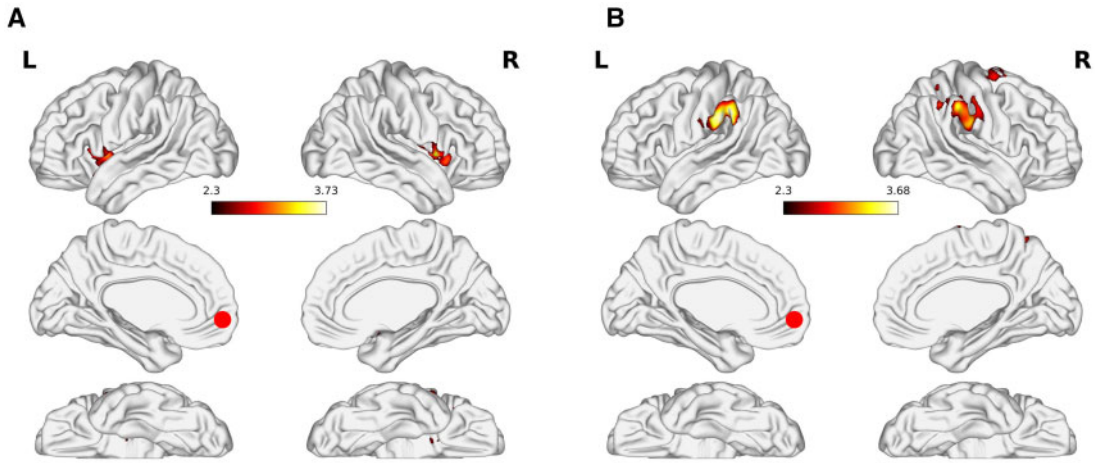

Fig. 1. Psychophysiological interaction results when seeding from the medial prefrontal cortex. The figure shows the AS x PHYS > Scrambled Control x PHYS contrast for (A) high-risk and (B) low-risk subjects. Results are cluster corrected for multiple comparisons with the cluster defining threshold set at $Z = 2.3$ and a cluster extent threshold of $P < 0.05$. The red circle represents the seed ROI.

Table 3. Psychophysiological interaction results when seeding from the middle frontal gyrus

Structure	Laterality	Cluster size	Max Z-score	Coordinates
ASxPhys > CtrlxPhys				
Low-risk subjects				
Superior lateral occipital cortex	Left	1678	3.61	−26, −64, 38
Inferior parietal lobe	Left		3.53	−54, −28, 38
Superior lateral occipital cortex	Left		3.34	−26, −72, 60
Primary somatosensory cortex	Left		3.32	−44, −42, 60
Anterior intra-parietal sulcus	Left		3.30	−32, −44, 44
Anterior intra-parietal sulcus	Left		3.29	−40, −46, 46
Anterior intra-parietal sulcus	Right	999	3.53	44, −46, 36
Inferior parietal lobe	Right		3.44	64, −46, 54
Superior lateral occipital cortex	Right		3.12	32, −62, 40
Superior lateral occipital cortex	Right		3.09	40, −62, 60
Superior lateral occipital cortex	Right		3.08	32, −64, 54
Inferior parietal lobe	Right		3.08	42, −56, 58
MSVxPhys > CtrlxPhys				
Low-risk subjects				
Superior lateral occipital cortex	Left	2142	3.77	−38, −62, 42
Superior lateral occipital cortex	Left		3.73	−26, −64, 38
Inferior parietal lobe	Left		3.60	−44, −48, 52
Anterior intra-parietal sulcus	Left		3.60	−38, −50, 48
Inferior parietal lobe	Left		3.47	−44, −54, 52
Inferior parietal lobe	Left		3.40	−56, −28, 40
Anterior intra-parietal sulcus	Right	1803	3.85	46, −44, 38
Anterior intra-parietal sulcus	Right		3.70	40, −50, 34
Inferior parietal lobe	Right		3.51	46, −46, 54
Central precuneous	Right		3.26	8, −78, 58
Superior lateral occipital cortex	Right		3.23	28, −74, 54

Notes: Results are cluster corrected for multiple comparisons with the cluster defining threshold set at $Z = 2.3$ and a cluster extent threshold of $P < 0.05$ and coordinates are shown in MNI152 space. Structures with a value in the cluster size column represent the peak Z-score while all others correspond to local maxima within the cluster.

the message-feature by issue-involvement interactions that contribute to persuasion-related neural processes, particularly among high-drug subjects. This is the critical advance necessary if we wish to design effective messages that motivate attitude change among this difficult-to-reach audience. To that end, this discussion section primarily focuses on network connectivity patterns among our high-drug risk subjects. However, important differences among low-drug risk subjects, particularly as they relate to earlier findings, are also discussed.

Two functional perspectives currently exist on the patterns of network connectivity that underpin persuasive message processing. The first argues that persuasive messages rely on affective-executive processing where network connectivity may

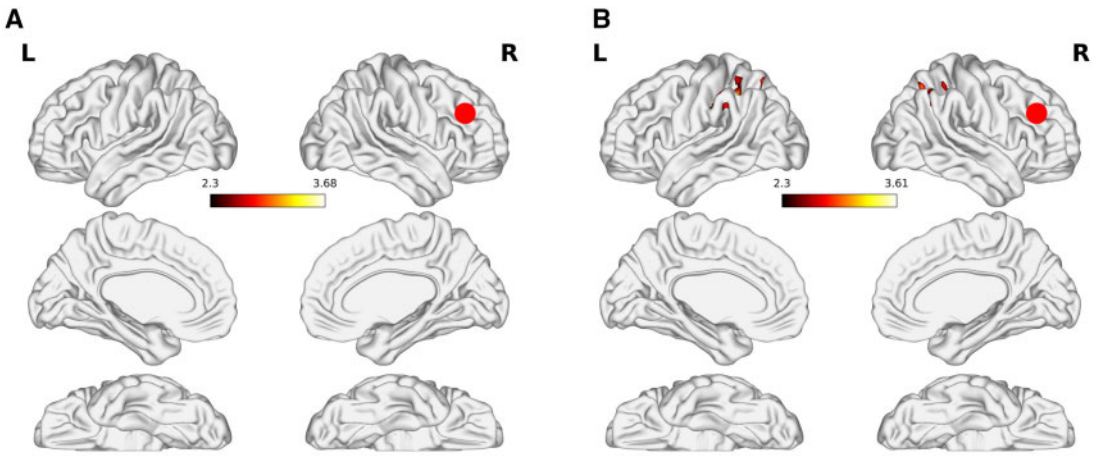

Fig. 2. Psychophysiological interaction results when seeding from the middle frontal gyrus. The figure shows the AS x PHYS > Scrambled Control x PHYS contrast for (A) high-risk and (B) low-risk subjects. Results are cluster corrected for multiple comparisons with the cluster defining threshold set at $Z = 2.3$ and a cluster extent threshold of $P < 0.05$. Note: No activations survived thresholding for high-risk subjects. The red circle represents the seed ROI.

Table 4. Psychophysiological interaction results when seeding from the inferior lateral occipital cortex

Structure	Laterality	Cluster size	Max Z-score	Coordinates
ASxPhys > CtrlxPhys				
High-risk subjects				
Inferior temporal gyrus	Left	8636	4.16	−46, −38, −10
Caudate Nucleus (Dosral Striatum)	Left		4.10	−8, 4, 10
Superior lateral occipital cortex	Left		3.98	−36, −64, 18
Inferior temporal gyrus	Left		3.78	−44, 0, −38
Dorsomedial prefrontal cortex	Left		3.59	−2, 46, 40
Superior lateral occipital cortex	Left		3.59	−50, −66, 16
Subcallosal cortex	Left	2012	3.86	−10, 16, −20
Subcallosal cortex	Right		3.86	6, 14, −18
Orbitofrontal cortex	Left		3.64	−8, 60, −14
Orbitofrontal cortex	Left		3.59	−32, 30, −8
Ventromedial prefrontal cortex	Left		3.45	−8, 70, 6
Cerebellum	Right	1203	4.06	24, −72, −34
Cerebellum	Right		3.64	42, −68, −40
Cerebellum	Right		3.17	40, −70, −32
Low-risk subjects				
Angular gyrus	Left	3275	4.07	−44, -52, 26
Superior lateral occipital cortex	Left		4.02	−46, −66, 36
Middle temporal gyrus	Left		3.97	−62, −34, −10
Middle temporal gyrus	Left		3.91	−60, −52, −4
Middle temporal gyrus	Left		3.75	−56, −38, −4
Superior lateral occipital cortex	Left		66	−58, −62, 28
Orbitofrontal cortex	Left	2596	3.64	−40, 42, −14
Ventrolateral prefrontal cortex	Left		4.09	−52, 34, −10
Orbitofrontal cortex	Left		4.03	−42, 32, −16
Ventrolateral prefrontal cortex	Left		3.77	−28, 64, 0
Middle frontal gyrus	Left	2427	4.08	−32, 12, 52
Dorsomedial prefrontal cortex	Left		3.39	−6, 42, 36
Superior frontal gyrus	Left		3.31	−14, 32, 50
Middle frontal gyrus	Left	2427	3.31	−38, 24, 38
Superior frontal gyrus	Left		3.29	−10, 30, 52
Superior frontal gyrus	Left		3.23	−20, 22, 56
Cerebellum	Right	1762	4.46	24, −74, −42
MSVxPhys > CtrlxPhys				
High-risk > low-risk subjects				
Central operculum	Left	1178	3.15	−34, 8, 16
Amygdala	Left		3.09	−22, −8, −8
Putamen (dorsal striatum)	Left		2.89	−24, 10, 8
Central operculum	Left		2.73	−48, 0, 8
Caudate nucleus (dorsal striatum)	Right	1097	3.20	12, 12, 20
Middle frontal gyrus	Right		2.95	36, 16, 28

(Continued)

Table 4. (Continued)

Structure	Laterality	Cluster size	Max Z-score	Coordinates
High-risk subjects				
Inferior temporal gyrus	Left	3905	4.05	−46, −38, −10
Middle temporal gyrus	Left		4.01	−50, −44, −6
Middle temporal gyrus	Left		3.91	−58, −6, −28
Superior lateral occipital cortex	Left		3.74	−38, −62, 18
Inferior lateral occipital cortex	Left		3.74	−40, −62, 4
Superior lateral occipital cortex	Left		3.62	−52, −68, 18
Cerebellum	Right	3035	4.33	20, −76, −36
Cerebellum	Right		4.30	38, −72, −42
Superior lateral occipital cortex	Right		4.08	44, −60, 22
Cerebellum	Right		3.86	42, −66, −24
Temporal occipital fusiform cortex	Right		3.72	42, −54, −12
Caudate nucleus (dorsal striatum)	Left	2926	4.10	−6, 2, 10
Orbitofrontal cortex	Left		3.89	−32, 30, −8
Orbitofrontal cortex	Left		3.45	−44, 48, −16
Orbitofrontal cortex	Left		3.36	−26, 52, −6
Temporal pole	Left		3.32	−20, 12, −32
Superior frontal gyrus	Left	1108	3.17	−4, 24, 54
Superior frontal gyrus	Left		3.03	−6, 46, 36
Paracingulate cortex	Left		3.01	−6, 36, 34
Paracingulate cortex	Left		3.01	−4, 40, 32
Dorsomedial prefrontal cortex	Left		2.87	−12, 52, 44
Low-risk subjects				
Superior frontal gyrus	Left	3388	4.11	−26, 20, 54
Superior frontal gyrus	Left		4.08	−12, 32, 52
Orbitofrontal cortex	Left		3.97	−32, 64, −8
Orbitofrontal cortex	Left		3.94	−24, 66, −0
Middle frontal gyrus	Left		3.83	−32, 20, 50
Middle frontal gyrus	Left		3.77	−34, 8, 54
Middle temporal gyrus	Left	2928	4.08	−60, −32, −8
Angular gyrus	Left		4.03	−44, −54, 26
Middle temporal gyrus	Left		3.90	−66, −32, −6
Middle temporal gyrus	Left		3.82	−50, −38, −4
Middle temporal gyrus	Left		3.81	−46, −46, 2
Superior lateral occipital cortex	Left		3.72	−48, −70, 36
Cerebellum	Right	1782	4.34	38, −78, −42
Orbitofrontal cortex	Left	1310	4.78	−42, 30, −20
Orbitofrontal cortex	Left		4.62	−40, 34, −16
Orbitofrontal cortex	Left		4.48	−50, 30, −12
Ventrolateral prefrontal cortex	Left		4.06	−48, 44, −12
Orbitofrontal cortex	Left		3.87	−36, 46, −18
Ventrolateral prefrontal cortex	Left		3.85	−48, 50, −2

Notes: Results are cluster corrected for multiple comparisons with the cluster defining threshold set at $Z = 2.3$ and a cluster extent threshold of $P < 0.05$ and coordinates are shown in MNI152 space. Structures with a value in the cluster size column represent the peak Z-score while all others correspond to local maxima within the cluster.

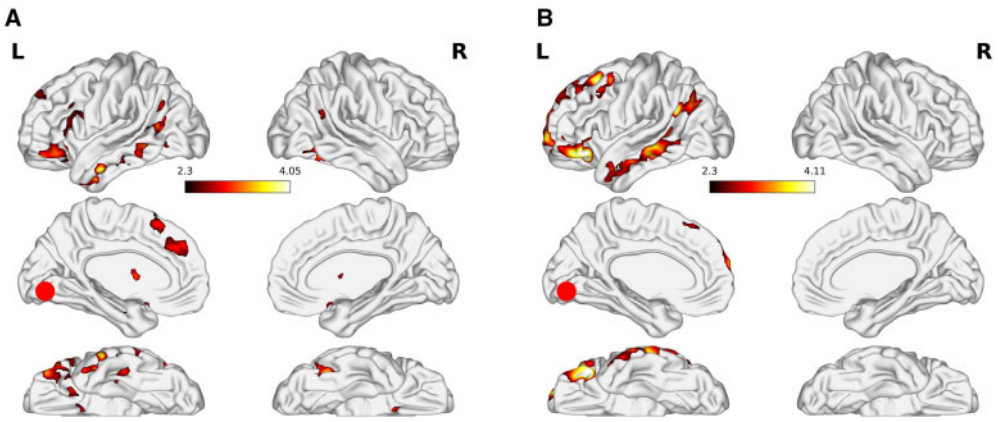

Fig. 3. Psychophysiological interaction results when seeding from the inferior lateral occipital cortex. The figure shows the MSV x PHYS > Scrambled Control x PHYS contrast for (A) high-risk and (B) low-risk subjects. Results are cluster corrected for multiple comparisons with the cluster defining threshold set at $Z = 2.3$ and a cluster extent threshold of $P < 0.05$. The red circle represents the seed ROI.

Table 5. Pearson correlations between PEs and PME in IS1 (college students) and IS2 (nationally representative adolescents)

	1	2	3	4	5	6
High-risk subjects						
1 IS1 PME	1					
2 IS2 PME	0.935**	1				
3 MPFC/TP	0.257	0.323	1			
4 MFG/AG	0.365*	0.286	0.204	1		
5 MFG/SPL	0.385*	0.337	0.270	0.695**	1	
6 STG/PSC	−0.146	−0.184	−0.062	0.254	0.376*	1
Low-risk subjects						
1 IS1 PME	1					
2 IS2 PME	0.896**	1				
3 MPFC/TP	0.170	0.024	1			
4 MFG/AG	0.009	0.036	0.043	1		
5 MFG/SPL	−0.049	−0.042	0.083	0.922**	1	
6 STG/PSC	−0.230	−0.374*	0.207	−0.176	−0.184	1

Correlation is significant at the *$P = 0.05$ level (two-tailed) and **$P = 0.01$ level (two-tailed).

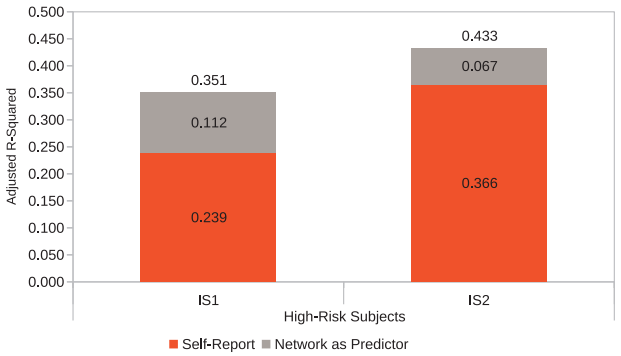

Fig. 4. Network-as-predictor results for high-risk subjects. A stepwise regression model was constructed where self-reported pMSV, pAS, and their interaction was entered in the first step and MFG-SPL connectivity parameter estimates (PEs) were entered in the second step. This analysis shows the unique contribution of self-report and network-connectivity PEs to overall prediction accuracy of perceived message effectiveness in two independent samples (IS1 and IS2). Results are not shown for the low-risk group as the inclusion of network-connectivity PEs did not significantly improve prediction accuracies.

be up- or down-regulated depending on whether audiences find the message convincing (Ramsay *et al.*, 2013; Cooper *et al.*, 2017) or choose to engage in defensive processing (Dinh-Williams *et al.*, 2014). Another (but not necessarily competing) view argues that persuasive messages that are too arousing disrupt sensory-encoding network connectivity, thereby inhibiting message processing (Seelig *et al.*, 2014). While informative, these studies manipulate just one of the many factors known to play a role in message processing (i.e. issue involvement, MSV, AS). In the present study, we draw on the ELM which characterizes persuasion as a multi-determined process where features of a message (i.e. MSV, AS) are modulated by issue involvement. By better specifying the interactions between these factors, we find continued support for the affective-executive hypothesis, particularly among high-risk subjects. By comparison, we fail to replicate earlier findings related to the encoding-disruption hypotheses. Furthermore, we show that the strength of these network connections is predictive of PME in two independent samples. We turn our attention to these findings, and their implications, below.

The affective-executive hypothesis

For nearly all seed ROIs we see qualitative differences in neural processing by risk-group. Previous work has observed that AS and issue involvement together modulate functional connections between regions commonly activated in affective and executive processing (Ramsay *et al.*, 2013; Dinh-Williams *et al.*, 2014). While we are unable to replicate earlier findings using an IFG seed, other results observed in the AS × PHYS > C1 × PHYS contrast provide support for this hypothesis. For instance, when seeding from the MPFC among high-drug risk subjects, we see connectivity with structures implicated in salience (dorsoanterior insula) and consummatory reward (putamen, orbitofrontal cortex) processing. It is also interesting to note that, while the iLOC seed was evaluated to test the encoding-disruption hypothesis, connectivity patterns with the amygdala and dorsal striatum (unique to high-drug risk subjects) provide additional support for the role of affect in message evaluation (we return to this point below).

One question that remains is why our results in this contrast do not perfectly map onto findings from earlier studies. Two possible explanations exist. First, our task was more of a conceptual rather than a direct replication. One difference is that our subjects continuously rated each PSA using a CRM, whereas participants passively watched messages in earlier studies. This difference may have subtly influenced the way audiences engaged with the message. This is particularly true for the Dinh-Williams *et al.* study (2014), which provided evidence that at-risk subjects cognitively disengaged from the messages. By comparison, our task required slightly more message engagement compared to passive watching, which according to the ELM, should result in counterarguing (a form of defensive processing) among issue-involved subjects.

A second difference may be attributed to neural development. Ramsay *et al.* conducted their study among adolescents who have important neuropsychological differences from adults in the amygdala, prefrontal cortex, and striatum (Tottenham and Galván, 2016). Evidence shows that the strength of functional connections between affective and executive regions while watching videos featuring smokers (Do and Galván, 2016) or when viewing graphic warning labels on cigarette packages (Do and Galván, 2015) influences short-term craving impulses differently among adolescents compared to adults.

At the same time, we also know that many of these developmental changes persist into early adulthood. While our participant group was older than those reported by Do and Galván (2015, 2016) and Ramsay *et al.* (2013), the brains (and underlying network connections) of college students might still be undergoing developmental changes. That our network-as-predictor results were more accurate among the adolescent sample (compared to the college sample) provides some evidence for this view. Teasing apart these developmental differences in processing, and exactly when they occur, is an important future direction as persuasion strategies that are effective among adults may have different impacts among adolescents.

More broadly, a number of studies implicate MPFC activity in successful persuasion (for a review, see Falk and Scholz, 2018). The MPFC seed ROI we used was centered on a sub-region of the MPFC implicated in positive subjective valuation during persuasive message processing (Cooper *et al.*, 2017). When seeding from this region among high-risk subjects in the AS × PHYS > C1 × PHYS contrast, we see connectivity with

dorsal striatum structures, particularly the putamen. This connectivity pattern is worth further consideration as it might, on first glance, seem to contradict findings from Cooper *et al.* who characterized MPFC connectivity with the ventral striatum as being associated with positive subjective valuation. Here, we return to meta-analytic results to help clarify this pattern. The ventral striatum region used by Cooper *et al.* (2017) corresponded to the peak meta-analytic voxel by Bartra *et al.* (2013). However, these meta-analytic results also indicate that subjective valuation extends more broadly within the striatum, including into more dorsal regions such as the putamen. Accordingly, our MPFC connectivity patterns with the putamen are not inconsistent with the view that persuasion requires message receivers to evaluate the positive outcomes associated with behavior change. These findings also extend this view to implicate the striatum more broadly in the persuasion process.

Still, we do not directly replicate MPFC connectivity with ventral striatum among either our high- or low-risk subjects as observed in Cooper *et al.* (2017). This may be driven, at least in part, by our choice of stimuli and experimental task. Cooper *et al.* employed a positive intervention where subjects were encouraged to mentalize ways in which they might enact lifestyle changes that would lead to positive health outcomes. Accordingly, it is not unexpected that their task engaged the ventral striatum as this structure is heavily implicated in reward anticipation (O'Doherty *et al.*, 2004), particularly as related to the calculation or reward/effort tradeoffs associated with goal-directed behavior (Botvinick *et al.*, 2009; Kool *et al.*, 2013). By comparison, our stimuli were consistent with more typical anti-drug PSAs which underscore the negative consequences associated with drug use.

The encoding-disruption hypothesis

One surprising area where we see similarities between risk-groups is in network connectivity patterns between occipital and temporal cortices. Contrary to previous findings (Seelig *et al.*, 2014), when seeding from the iLOC in the $MSV \times PHYS > C1 \times PHYS$ contrast, we see connectivity with MTG for both risk groups, with high-risk subjects also showing connectivity with inferior temporal gyrus (ITG). As has been argued elsewhere (Weber *et al.*, 2013), even though MSV may be quite useful for capturing the attention of high sensation-seeking, high drug-risk audiences, this can backfire if these audiences are then required to process a counter-attitudinal message.

When seeding from the iLOC, we see an important group difference where high-risk but *not* low-risk subjects show functional connectivity with regions implicated in affective and salience processing including AMY, dorsal striatum (both putamen and caudate), and central operculum. We interpret these results as further evidence that message strategies that rely solely on MSV may not be effective among counter-attitudinal target audiences. This interpretation is further supported by our self-reported pMSV evaluations. In prediction models that use our self-reported PSA evaluations in our MRI sample to predict PME in two independent, large samples of adolescents and freshman college students, pMSV was consistently the weakest (and most non-significant) predictor.

The video-specific network connections that underpin message effectiveness

To recapitulate our network-as-predictor results, we found that MFG-SPL connectivity significantly improved predictions of per-video ratings in two independent samples, exclusively within our high-drug risk group. Next, we consider how this result relates to the results of our ELM-based whole-brain PPI analyses, in which we did not observe any significant modulation in MFG-SPL connectivity by either MSV or AS for either risk group.

Practical implications. The degree to which MSV and AS interact with subject specific drug use risk to modulate brain network connectivity patterns in response to anti-drug messages helps advance our mechanistic understanding of the persuasion process, but practitioners, public-health officials, and message designers have a different need. They are looking for principled yet data-driven ways to identify which anti-drug PSAs are most effective among difficult to reach high-risk target audiences. Unfortunately, traditional self-report measures either misidentify which messages are most effective (e.g. Falk *et al.*, 2012), or generate so little variability as to make identifying effective PSAs all but impossible (e.g. Weber *et al.*, 2013). The brain-as-predictor analysis and our network-as-predictor extension, was designed to help solve this problem by asking if there are connectivity patterns in the brain that explain additional variance in message-effectiveness evaluations above and beyond self-report data. Such an analysis relies on the following logic: (i) identify network connectivity patterns that systematically vary by risk group, (ii) identify what message features drive this variability, (iii) use this information to improve out-of-sample predictions in high-risk target audiences, and (iv) use this information for message selection.

While our study and results conform to this general logic, points #2 and #3 are worth considering further. This study hypothesized (and found) that MSV and AS interact with issue involvement to modulate persuasion network connectivity patterns. However, we did not observe MFG-SPL connectivity in either the $MSV \times PHYS > C1 \times PHYS$ or $AS \times PHYS > C1 \times PHYS$ contrasts. This suggests that, even though the MFG and SPL are commonly implicated in the persuasion neuroscience literature (for a synthesis, see Kaye *et al.*, 2016), regardless of risk-group, connectivity patterns between these structures do not seem to be systematically modulated by MSV, or AS. It seems that some other feature of the PSAs used in this study influences MFG-SPL connectivity, and this feature appears to matter. We return to this idea again in the ELM section below.

Point #3 is focused on improving out-of-sample prediction accuracy. While MFG-SPL PEs allow for better PME prediction among high-risk subjects for both IS1 and IS2, they fall short of what has been achieved in other brain-as-predictor analyses. Weber *et al.* (2014) show that measures of within-ROI signal change in the MFG and SPL contribute to prediction accuracies as high as 59%. Falk *et al.* have shown similar accuracies for the MPFC (e.g. 2012; 2016). If we believe that persuasion is a network-level process, then we should expect that capturing network information should improve prediction accuracies above and beyond self-report and signal change (as has been shown in Cooper *et al.*, 2017). Indeed, exploratory analyses on our dataset show that, in models including both signal changes and connectivity PEs, the inclusion of connectivity PEs no longer significantly improved adjusted R2. One possible methodological reason for this finding could be that our functional connectivity parameter estimates were drawn from relatively small ROIs that were based on previous studies in the literature and not from meta-analyses such as Neurosynth (Yarkoni *et al.*, 2011) or even an independent localizer task, as is also common in the persuasion neuroscience literature (Falk *et al.*, 2015). Combined with the facts that PPI analyses are quite sensitive to

the accurate selection of seed regions and that the increased number of parameters in a model reduces statistical power (Friston *et al.*, 1997; O'Reilly *et al.*, 2012), it is possible that our PEs simply contain higher levels of noise. This coupled with the additional analytical effort required to conduct a PPI analysis suggests that, for now, message designers may still be better suited by using the traditional brain-as-predictor approach which is based on simple BOLD signal change parameters.

Mechanistic implications. We know from previous research that the MFG is an important structure implicated in counterarguing processes, particularly among high-drug risk subjects (Weber *et al.*, 2014). Further, the fact that activity in this structure is commonly observed across a number of studies suggests that it is a core component of the persuasion network (e.g. Chua *et al.*, 2009; Falk *et al.*, 2010; Ramsay *et al.*, 2013). We also know that the MFG is implicated in processing audiovisual narratives (Wilson *et al.*, 2008) like the PSAs used in this study. However, a recent study investigating intersubject correlations (ISC) among both high- and low-risk subjects viewing health-message PSAs, did not observe ISCs in this region (Imhof *et al.*, 2017). To be clear, we are not arguing that the absence of a statistically significant result implies no effect. But it does suggest that there is something important about the MFG that emerges in group-level GLMs and at the individual video level, but does not systematically vary for ISCs or group-level PPIs. Identifying what message features contribute to MFG activity in future studies is important both for our mechanistic understanding of the persuasion process as well as for message designers looking for practical ways of improving anti-drug campaigns.

More broadly, while this study replicates and extends our understanding of connectivity patterns between structures within the persuasion network, the PPI analyses employed herein and in previous studies are not without limitation. Most notably, PPI only considers the way in which a task modulates connectivity between one seed ROI and all other regions in the brain. Such a one-to-many connectivity analysis necessarily oversimplifies the true network architecture. In fact, it is entirely likely that dynamic connections between multiple ROIs within the network better characterize the persuasion process. Recent advances in the application of graph theory to neuroimaging data (e.g. Bassett and Sporns, 2017) provide an analytical approach for interrogating these complex dynamics. Two immediate next-steps can help resolve ambiguity about network structure and dynamics. First, hypothesis driven analyses should construct sparse graphs where nodes in the graph represent recurrent structures within the persuasion network. At the same time, more data-driven approaches based on a whole-brain parcellation (e.g. Glasser *et al.*, 2016) may illuminate as of yet undiscovered network typologies. New toolboxes, particularly as implemented in both Python Abraham *et al.* (2014) and Matlab (Rubinov and Sporns, 2010; Sizemore and Bassett, 2017), should help advance research in this area.

Returning to the ELM

The ELM (Petty and Cacioppo, 1986) makes the case that message features interact with individual differences to shape motivation and ability to process a message. Different message-by-audience interactions determine if audiences (i) choose to process a message deeply or superficially and (ii) if this processing is biased (e.g. counterarguing, message disengagement) or not. With two exceptions (Weber *et al.*, 2014; Imhof *et al.*, 2017), tests using an ELM framework are largely absent from the persuasion neuroscience literature.

Accordingly, much of the language used to describe the neural processes that underpin persuasion are not framed using ELM-consistent language; in-fact many test single- and *not* dual-process models (such as the ELM) of persuasion (for more detail, see Vezich *et al.*, 2016). In an effort to draw linkages between these literatures, we might re-frame the cognitive processes identified with the ELM to be more consonant with the persuasion neuroscience literature. Explained in these terms, message features and issue-involvement modulate affective responses that motivate executive processing strategies. The results reported within this manuscript, as well as across a number of studies (Kaye *et al.*, 2016) largely support this view.

For instance, affective processing associated with subjective value is well-known to motivate executive processing and behavior (for a review, see Braver *et al.*, 2014) and studies implicating subjective value in persuasion (e.g. Cooper *et al.*, 2017; Falk and Scholz, 2018) provide additional clarity as to what might be driving this motivational response. At the same time, our network-as-predictor analysis identified variation in individual videos that was not captured by ELM-relevant variables (i.e. MSV, AS) but still predicted message perceptions in independent samples. This suggests possible extensions to the ELM.

Falk *et al.* (2009) implicated social processing in persuasion, and the literature to-date extends this into additional areas including reward processing, self-reflection, social pain, and salience detection (Kaye *et al.*, 2016). Another possible explanation lies in the ways audiences process narratives (Slater and Rouner, 2002). Recent work shows that audiences with different initial beliefs interpret narratives in rather different ways, and that these differences are driven by similar or dissimilar activation in a variety of structures, including the right MFG (the seed ROI we used in our network-as-predictor analysis; for further details, see Yeshurun *et al.*, 2017). These studies hint at factors that underpin individual differences in processing and suggest future directions for ELM-based research.

On a more critical note, one limitation on much of the persuasion neuroscience research, including the present study, is that it largely relies on a 'magic bullet' logic where a single exposure to a single message should result in persuasion. In today's media landscape, individuals are often exposed to the same (or even competing) messages multiple times across different channels. And, even with these multiple exposures, we know that changing attitudes and behaviors is nontrivial. Addressing these challenges requires longitudinal work focused on how plasticity in attitudes and behaviors corresponds with neurobiological changes. There is some work on this, particularly among smoker populations (Berkman *et al.*, 2011; Falk *et al.*, 2011), but these studies still utilize a one-shot fMRI session. Admittedly, longitudinal fMRI work is both costly and slow, but repeated measures are necessary if we wish to understand how message dynamics modulate neural dynamics and behavior over time.

Conclusion

Although no meta-analysis currently exists, the persuasion neuroscience literature has grown to a point that the same regions consistently emerge across a number of studies (Kaye *et al.*, 2016). However, our study shows important qualitative differences in connectivity patterns by risk group. These results are bolstered by their predictive-validity. Specifically, these

network connections are predictive of how independent audiences process the message. This suggests that careful work is needed to identify and tailor (Noar *et al.*, 2007) messages for specific and narrowly defined target audiences. This point is particularly important given that recent evidence demonstrates that individual differences in message encoding and social network structure predict persuasive message outcomes (Pegors *et al.*, 2017). In the past, traditional message channels (e.g. broadcast television) made such fine-grained tailoring difficult. However, with the advent of social media and other Internet channels, micro-targeting strategies have become increasingly feasible. Future work should seek to identify how specific message components interact with theoretically relevant individual differences in order to increase message effectiveness.

Acknowledgement

We thank our colleagues at the Annenberg School for Communication at The University of Pennsylvania for generously sharing the stimuli and one of the data sets used in this study.

Funding

This work was supported by the University of California Santa Barbara Brain Imaging Center. R.H. is an awardee of the George D. McCune Dissertation Fellowship.

Conflict of interest: None declared.

References

Abraham, A., Pedregosa, F., Eickenberg, M., *et al.* (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, **8**, 14.

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**, 412–27.

Bassett, D.S., Gazzaniga, M.S. (2011). Understanding complexity in the human brain. *Trends in Cognitive Sciences*, **15** (5),200–9.

Bassett, D.S., Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, **20** (3), 353–64.

Beckmann, C.F., Smith, S.M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, **23** (2), 137–52.

Berkman, E.T., Falk, E.B. (2013). Beyond brain mapping: using neural measures to predict real-world outcomes. *Current Directions in Psychological Science*, **22**(1), 45–50.

Berkman, E.T., Falk, E.B., Lieberman, M.D. (2011). In the trenches of real-world self-control: neural correlates of breaking the link between craving and smoking. *Psychological Science*, **22**(4),498–506.

Bigsby, E., Cappella, J.N., Seitz, H.H. (2013). Efficiently and effectively evaluating public service announcements: additional evidence for the utility of perceived effectiveness. *Communication Monographs*, **80**(1), 1–23.

Botvinick, M.M., Huffstetler, S., McGuire, J.T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience*, **9**(1), 16–27.

Braver, T.S., Krug, M.K., Chiew, K.S., *et al* (2014). Mechanisms of motivation-cognition interaction: Challenges and opportunities. *Cognitive, Affective & Behavioral Neuroscience*, **14**(2), 443–72.

Bullmore, E., Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, **13**(5), 336–49.

Cappella, J.N., Yzer, M.C., Fishbein, M. (2003). Using beliefs about positive and negative consequences as the basis for designing message interventions for lowering risky behavior. In: Romer D. editors. *Reducing Adolescent Risk: Toward an Integrated Approach*, pp. 210–220. Thousand Oaks, CA: Sage Publications.

Chua, H.F., Liberzon, I., Welsh, R.C., Strecher, V.J. (2009). Neural correlates of message tailoring and self-relatedness in smoking cessation programming. *Biological Psychiatry*, **65**(2), 165–8.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Cooper, N., Bassett, D.S., Falk, E.B. (2017). Coherent activity between brain regions that code for value is linked to the malleability of human behavior. *Scientific Reports*, **7**(43250), 4325.

Dillard, J.P., Shen, L., Vail, R.G. (2007a). Does perceived message effectiveness cause persuasion or vice versa? 17 consistent answers. *Human Communication Research*, **33**(4), 467–88.

Dillard, J.P., Weber, K.M., Vail, R.G. (2007b). The relationship between the perceived and actual effectiveness of persuasive messages: a meta-analysis with implications for formative campaign research. *Journal of Communication*, **57**(4), 613–31.

Dinh-Williams, L., Mendrek, A., Dumais, A., Bourque, J., Potvin, S. (2014). Executive-affective connectivity in smokers viewing anti-smoking images: an fMRI study. *Psychiatry Research: Neuroimaging*, **224**(3), 262–8.

Do, K.T., Galván, A. (2015). FDA cigarette warning labels lower craving and elicit frontoinsular activation in adolescent smokers. *Social Cognitive and Affective Neuroscience* **10** (11),1484–96.

Do, K.T., Galván, A. (2016). Neural sensitivity to smoking stimuli is associated with cigarette craving in adolescent smokers. *Journal of Adolescent Health*, **58**(2), 186–94.

Falk, E.B., Berkman, E.T., Lieberman, M.D. (2012). From neural responses to population behavior: neural focus group predicts population-level media effects. *Psychological Science*, **23**(5), 439–45.

Falk, E.B., Berkman, E.T., Mann, T., Harrison, B., Lieberman, M.D. (2010). Predicting persuasion-induced behavior change from the brain. *Journal of Neuroscience*, **30**(25), 8421–4.

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 Psychology*, **30**(2), 177–85.

Falk, E.B., Cascio, C.N., Coronel, J.C. (2015). Neural prediction of communication-relevant outcomes. *Communication Methods and Measures*, **9**(1–2), 30–54.

Falk, E.B., O'Donnell, M.B., Tompson, S., *et al.* (2016). Functional brain imaging predicts public health campaign success. *Social Cognitive and Affective Neuroscience*, **11**(2), 204–14.

Falk, E. B., Rameson, L., Berkman, E. T. *et al.* (2009). The neural correlates of persuasion: A common network across cultures and media. *Journal of Cognitive Neuroscience*, **22**(11), 2447–2459.

Falk, E.B., Rameson, L., Berkman, E.T., *et al.* (2010). The neural correlates of persuasion: a common network across cultures and media. *Journal of Cognitive Neuroscience*, **22**(11), 2447–59.

Falk, E.B., Scholz, C. (2018). Persuasion, influence and value: perspectives from communication and social neuroscience. *Annual Review of Psychology*, **69**(1), doi: 10.1146/annurev-psych-122216-011821.

Friston, K.J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping*, **2**(1–2), 56–78.

Friston, K.J. (2011). Functional and effective connectivity: a review. *Brain Connectivity*, **1**(1), 13–36.

Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, **6**(3), 218–29.

Glasser, M.F., Coalson, T.S., Robinson, E.C., *et al.* (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, **7615**(563), 171–8.

Imhof, M.A., Schmalzle, R., Renner, B., Schupp, H.T. (2017). How real-life health messages engage our brains: shared processing of effective anti-alcohol videos. *Social Cognitive and Affective Neuroscience*, **12**(7), 1188–96.

Kang, Y., Cappella, J.N., Fishbein, M. (2006). The attentional mechanism of message sensation value: interaction between message sensation value and argument quality on message effectiveness. *Communication Monographs*, **73**(4), 351–78.

Kaye, S.-A., White, M.J., Lewis, I. (2016). The use of neurocognitive methods in assessing health communication messages: a systematic review. *Journal of Health Psychology*, **22**(12), 1534–55.

Kool, W., McGuire, J.T., Wang, G.J., Botvinick, M.M., Brosnan, S.F. (2013). Neural and behavioral evidence for an intrinsic cost of self-control. *PLoS ONE*, **8**(8), 72626.

Lang, A. (2006). Using the limited capacity model of motivated mediated message processing to design effective cancer communication messages. *Journal of Communication*, **56**(1), S57–80.

McLaren, D.G., Ries, M.L., Xu, G., Johnson, S.C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *NeuroImage*, **61**(4), 1277–86.

Noar, S.M., Benac, C.N., Harris, M.S. (2007). Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychological Bulletin*, **133**(4), 673–93.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, **304**(5669), 452–4.

O'Reilly, J.X., Woolrich, M.W., Behrens, T.E.J., Smith, S.M., Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, **7**(5), 604–9.

Pegors, T.K., Tompson, S., O'Donnell, M.B., Falk, E.B. (2017). Predicting behavior change from persuasive messages using neural representational similarity and social network analyses. *NeuroImage*, **157**, 118–28.

Petersen, S.E., Sporns, O. (2015). Brain networks and cognitive architectures. *Neuron*, **88**(1), 207–19.

Petty, R.E., Cacioppo, J.T. (1986). The elaboration likelihood model of persuasion. In: Berkowitz L., editor, *Advances in Experimental Social Psychology* (v19 ed., pp. 123–205). New York, NY: Academic Press.

Poldrack, R.A., Farah, M.J. (2015). Progress and challenges in probing the human brain. *Nature* **526**(7573), 371–9.

Ramsay, I.S., Yzer, M.C., Luciana, M., Vohs, K.D., Macdonald, A.W.I. (2013). Affective and executive network processing associated with persuasive antidrug messages. *Journal of Cognitive Neuroscience*, **25**(7), 1136–47.

Rogers, E.M. (1994). *A History of Communication Study: A Biographical Approach*. New York, NY: The Free Press.

Rubinov, M., Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, **52**(3), 1059–69.

Seelig, D., Wang, A.-L., Jaganathan, K., Loughead, J.W., Blady, S.J., Childress, A.R., Romer, D. (2014). Low message sensation health promotion videos are better remembered and activate areas of the brain associated with memory encoding. *PLoS One*, **9**(11), e113256.

Sizemore, A.E., Bassett, D.S. (inpress). Dynamic graph metrics: Tutorial, toolbox, and tale. *NeuroImage*, doi: 10.1016/j.neuroimage.2017.06.081.

Slater, M.D., Rouner, D. (2002). Entertainment-education and elaboration likelihood: understanding the processing of narrative persuasion. *Communication Theory*, **12**(2), 173–91.

Tottenham, N., Galván, A. (2016). Stress and the adolescent brain: amygdala-prefrontal cortex circuitry and ventral striatum as developmental targets. *Neuroscience and Biobehavioral Reviews*, **70**, 217–27.

Vezich, I.S., Falk, E.B., Lieberman, M.D. (2016). Persuasion neuroscience: new potential to test dual-process theories. In: Harmon-Jones E., Inzlicht M., editors. *Social Neuroscience: Biological Approaches to Social Psychology*, 1st edn., pp. 34–58, New York, NY: Routledge.

Vezich, I.S., Katzman, P.L., Ames, D.L., Falk, E.B., Lieberman, M.D. (2016). Modulating the neural bases of persuasion: why/how, gain/loss, and users/non-users. *Social Cognitive and Affective Neuroscience*, **12**(2), 283–97.

Wang, Z., Vang, M., Lookadoo, K., Tchernev, J.M., Cooper, C. (2014). Engaging high-sensation seekers: the dynamic interplay of sensation seeking, message visual-auditory complexity and arousing content. *Journal of Communication*, **5**(1), 101–24.

Weber, R., Eden, A., Huskey, R., Mangus, J.M., Falk, E. (2015). Bridging media psychology and cognitive neuroscience: challenges and opportunities. *Journal of Media Psychology*, **27**(3), 146–56.

Weber, R., Huskey, R., Mangus, J.M. *et al.* (2014). Neural predictors of message effectiveness during counterarguing in anti-drug campaigns. *Communication Monographs*, **82**(1), 4–30.

Weber, R., Huskey, R., Mangus, J.M., Westcott-Baker, A., Turner, B.O. (2015). Neural predictors of message effectiveness during counterarguing in anti-drug campaigns. *Communication Monographs*, **82**(1), 4–30.

Weber, R., Mangus, J.M., Huskey, R. (2015). Brain imaging in communication research: a practical guide to understanding and evaluating fMRI studies. *Communication Methods and Measures*, **9**(1-2), 5–29.

Weber, R., Westcott-Baker, A., Anderson, G.L. (2013). A multilevel analysis of antimarijuana public service announcement effectiveness. *Communication Monographs*, **80**(3), 302–30.

Wilson, S.M., Molnar-Szakacs, I., Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, **18**(1), 230–42.

Woolrich, M.W., Behrens, T.E.J., Beckmann, C.F., Jenkinson, M., Smith, S.M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, **21**(4), 1732–47.

Worsley, K.J. (2001). Statistical analysis of activation images. In: Jezzard, P., Matthews, P.M., Smith, S.M., editors. *Functional Mri: An Introduction to Methods* (pp. 251–270). Oxford, United Kingdom: Oxford University Press.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, **8**(8), 665–70.

Yeshurun, Y., Swanson, S., Simony, E., *et al.* (2017). Same story, different story: the neural representation of interpretive frameworks. *Psychological Science*, **28**(3), 307–19.

Zhao, X., Strasser, A., Cappella, J.N., Lerman, C., Fishbein, M. (2011). A measure of perceived argument strength: reliability and validity. *Communication Methods and Measures*, **5**(1), 48–75.