Health Communication

ISSN: 1041-0236 (Print) 1532-7027 (Online) Journal homepage: https://www.tandfonline.com/loi/hhth20

Deliberation and Valence as Dissociable Components of Counterarguing among Smokers: Evidence from Neuroimaging and Quantitative Linguistic Analysis

Jiaying Liu, Matthew B. O'Donnell & Emily B. Falk

HEALTH COMMUNICATION
https://doi.org/10.1080/10410236.2020.1712521

Routledge
Taylor & Francis Group

Check for updates

Deliberation and Valence as Dissociable Components of Counterarguing among Smokers: Evidence from Neuroimaging and Quantitative Linguistic Analysis

Jiaying Liu [a], Matthew B. O'Donnell[b], and Emily B. Falk[b,c,d]

[a]Department of Communication Studies, University of Georgia; [b]Annenberg School for Communication, University of Pennsylvania; [c]Department of Psychology, University of Pennsylvania; [d]Marketing Department, The Wharton School, University of Pennsylvania

ABSTRACT

Counterarguing is a key obstacle to successful persuasion. However, the difficulty of directly measuring counterarguing during message exposure limits knowledge of the underlying mechanisms. The current study combines neuroimaging and linguistic measures to unpack neurocognitive and psychological mechanisms associated with counterarguing among a sample of established smokers in response to anti-smoking messaging. We capture participants' neural activity in brain regions associated with effortful deliberation and negative argumentation during message exposure, and link it with their subsequent language patterns to further understanding of counterarguing in the brain. Greater brain activity within key regions of interest associated with deliberation and negative argumentation is associated with greater cognitive depth and less positivity in the post-scan message descriptions, respectively, among those who have lower intention to change their smoking behavior. We connect these neural representations of counterarguing with psychological theories and discuss implications that may increase the impact of persuasive communications.

The ultimate effectiveness of persuasive communication often depends on the amount, valence, and nature of thoughts audiences generate in response to the persuasive messages (Petty & Cacioppo, 1986). When individuals are confronted with counter-attitudinal information, and when their issue involvement is high, they are likely to counterargue (Miller & Baron, 1973). Counterarguing, the process of negatively elaborating and actively generating rebuttals to persuasive statements, is one of the key obstacles to successful persuasion (Fransen, Smit, & Verlegh, 2015; Slater & Rouner, 2002). Due to the hidden nature of mental processes that take place during exposure to persuasive messages, many intriguing questions about the processes underlying counterarguing remain.

In the current manuscript, we provide convergent evidence for two distinct components of counterarguing, cognitive deliberation and negative valence, combining neuroimaging during message exposure and quantitative linguistic analysis of thoughts that follow. Building on prior biological paradigms for measuring responses to arguments (Cacioppo, Tassinary, & Berntson, 2007; Potter & Bolls, 2012), this paradigm allows us to examine the processes that unfold in real time during message exposure. The neural measure also does not contaminate the process by eliciting self-reports during exposure. Next, we link neural activity during message exposure with subsequent naturalistic language use patterns in the recipients' free-form description of the messages (Davison, Navarre, & Vogel, 1995). Finally, given that neural activity associated with counterarguing has been observed to be more predictive of message effectiveness in high-risk individuals (Weber, Huskey, Mangus, Westcott-Baker, & Turner, 2015), we also investigate how individual differences in intention to change the behavior in question may affect this process.

Deliberative, negative processing among high-risk populations

Both dual-process models of human cognition and reactance theory shed important light on this line of inquiry. Dual process models suggest that under conditions of high deliberation, the outcomes of persuasive efforts largely depend on the valence of issue-relevant thinking (Chaiken, Liberman, & Eagly, 1989; Petty & Cacioppo, 1986). If messages lead people to actively refute the core arguments conveyed by the (often uncongenial) messages (Fransen et al., 2015; Ringold, 2002), the persuasive attempts are most likely to fail. In some situations, message-induced counterarguments might even entrench people's prior beliefs and produce counterproductive results (e.g., Freeman, Hennessy, & Marzullo, 2001). Counterarguing consists of negative arguments to persuasive statements. Based on the two crucial elements that determine how individuals approach information, route and valence of processing (Chaiken & Trope, 1999; Eagly & Chaiken, 2005; Petty & Cacioppo, 1986), there are two major components of

CONTACT Jiaying Liu jiaying.liu@uga.edu Department of Communication Studies, University of Georgia, 302 Herty Drive, 613 Caldwell Hall, Athens, GA 30602, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hhth.
Supplemental data for this article can be accessed at https://doi.org/10.1080/10410236.2020.1712521.

counterarguing. First is cognitive deliberation. In order to actively generate counterarguments, high cognitive engagement, that is, effortful central processing is required. One key component of this type of deliberation is cognitive depth, which involves careful scrutiny and critical, inferential judgment (Petty & Cacioppo, 1986). It is also possible that greater elaboration could lead to more complex thinking (Burleson, 1987) or greater consideration of how the issue connects or relates to other issues (Mandl & Ballstaedt, 1982-). Second, when people engage in counterarguing, the valence of issue-relevant thoughts will be mostly unfavorable (e.g., prevailing overall negativity in thoughts, and fewer positive thoughts). Therefore, counterarguing contains deliberative and valence components and is defined as deliberative and negative processing that defends prior beliefs and resists persuasion (Dillard & Shen, 2005; Slater & Rouner, 2002). Offering a complementary perspective, reactance theory posited that when a message inadvertently threatens one's freedom, an unpleasant motivational arousal emerges and triggers reactance in the form of negative cognitions; individuals are motivated to come up with arguments against the persuasive attempt, i.e., counterarguing, by downgrading the message and/or derogating the source (Brehm & Brehm, 1981; Dillard & Shen, 2005; Rains, 2013). Due to its important implications to persuasion outcomes, counterarguing and reactance more broadly, have been extensively studied in areas of communication and cognate disciplines (Brehm & Brehm, 1981; Dillard & Shen, 2005; Petty, Tormala, & Rucker, 2004; Rains & Turner, 2007; Tormala & Petty, 2002, 2004; Zuwerink & Devine, 1996). More recently, neuroscientists have also considered neural processes relevant to counterarguing; Weber and colleagues found that middle frontal gyrus and superior temporal gyrus were associated with perceived message effectiveness in the context of drug use, and predicted outcomes for high-risk users where self-reports did not (Weber, Huskey et al., 2015). Huskey, Mangus, Turner, and Weber (2017) also observed network connectivity patterns relevant to counterarguing among high-risk subjects.

Consistent with this work, counterarguing is especially prominent in higher risk groups who have less intention to change their behavior (Brehm & Brehm, 1981). For example, heavy smokers engage graphic warning labels more negatively and develop defensive and maladaptive psychological responses toward them (Erceg-Hurn & Steed, 2011; Freeman et al., 2001). High-risk individuals are likely to have lower internal drive to change and higher motivation to actively resist persuasion, and thus are more likely to argue against messages encouraging behavior change. Building on these insights, we extend prior research on the neural bases of counterarguing in several ways. First, we employ a functional localization approach to orthogonally map two mental functions of interest in the brain that simultaneously take place during counterarguing using an independent sample; this approach provides another window into considering some of the key processes implicated in counterarguing. Second, the current study also focuses on different outcome measures, i.e., deliberative and valence components of counterarguing as quantified by language measures, which have the potential to examine naturalistic thoughts in a highly scalable fashion. Finally, we also examine the potential moderation effects of participants' baseline individual differences in intention to change.

Thought-listing techniques

Thought listing has been commonly used as a gold-standard technique to gauge both the degree and valence of counterarguing in the context of message testing (Cacioppo, von Hippel, & Ernst, 1997). Following exposure to persuasive messages, individuals are asked to list the thoughts that went through their minds during message presentation. The thoughts listed are then content analyzed and sorted into different categories; most previous studies have focused on grouping thoughts with different valence positions by trained judges or by participants themselves (Brown & Gold, 2014; Cialdini, Levy, Herman, Kozlowski, & Petty, 1976; Petty & Cacioppo, 1979). Thought listing can help researchers capture thoughts and internal dialogue that reflect message recipients' subvocal psychological processing (Cacioppo et al., 1997; Heimberg, Nyman, & O'Brien, 1987).

In the context of message testing, where thought-listing is often employed *after* the fact, it remains a question whether what is captured by the retrospective measure equates to the psychological processing as it naturally unfolds *during* exposure to messages. As early as in the late 1970s, persuasion scholars used real-time psychophysiological methods (e.g., electroencephalogram [EEG] and electromyogram [EMG] data) in the development of thought-listing methods, which in general suggested that the thoughts listed afterward could well reflect their online processing during initial exposure to stimuli (Cacioppo & Sandman, 1981). Since then, studies including magnetoencephalogram (MEG), electrodermal activity measures (EDA), galvanic skin response (GSR), heart rate (HR), eye movements, event-related potentials (ERP), positron emission tomography (PET), and functional Magnetic Resonance Imaging (fMRI) have also explored psychologically relevant physiological responses to persuasive messages and related these measures to subsequent elaboration, affective reactions, and behavioral decision-making (for a review, see Cacioppo et al., 2007; Potter & Bolls, 2012). This type of real-time evidence, combined with subsequent descriptions, also highlighted that psychological responses can evolve following initial exposure to messages. For example, in some research, responses reported after initial exposure were observed to be more unfavorable compared to that gauged by concurrent measures (Roberts & Maccoby, 1973).

The present investigation builds on these studies by combining functional magnetic resonance imaging (fMRI) and linguistic analysis on transcriptions of post-exposure descriptions the participants generated to quantify neural and linguistic responses to messages during and after message exposure. Neuroimaging methods can examine a wide variety of neurocognitive processes simultaneously in real time during message processing and have been shown to enrich understanding of the underlying mechanisms that motivate behavior changes, and inform theories of persuasion (Falk, 2012; Lieberman, 2010; Weber, Huskey et al., 2015; see Berkman & Falk, 2013; Weber, Eden, Huskey, Mangus, & Falk, 2015 for discussion). In addition, quantitative linguistic measures such as a dictionary-based approach can capture both controlled and automatic reactions to persuasive messages and pick up nuanced language use patterns that may map systematically onto counterargument but are not easily identified by human coders. Different instructions and scoring procedures have been used in previous thought-listing studies to represent individuals' covert psychological processing (Cacioppo & Petty, 1981; Davison et al., 1995). Conducting linguistic analysis on free-form description transcripts can be broadly considered as belonging to the thought-listing approach (Cacioppo et al., 1997; Davison et al., 1995), with particular focus on instructions that encourage naturalistic thought listing and an analysis approach that is highly scalable. This can facilitate automatic linguistic analysis on free-from descriptions of thoughts for future investigations that may need to process a large volume of descriptive textual data. In addition, from a basic science perspective, integrating brain and linguistic analyses helps advance understanding of the mechanisms that underpin counterarguing. This methodological combination allows us to provide a complementary account that triangulates mechanisms involved in counterarguing, including processes associated with people's covert processing during message exposure, and examining how brain activity is reflected in the language used afterward to describe their thoughts.

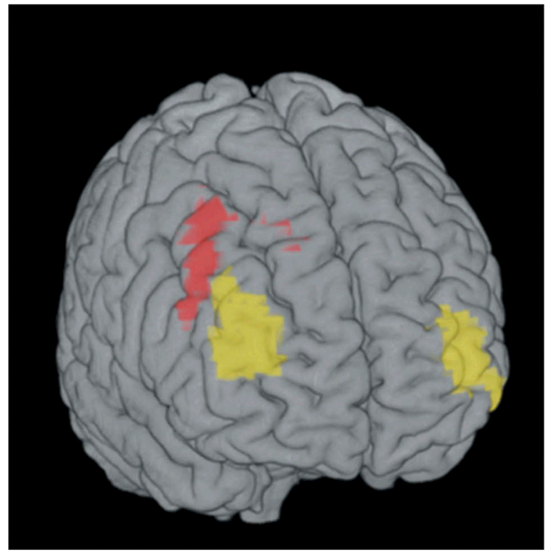


Figure 1. Functional regions of interest (ROIs).
Right and left dorsal lateral prefrontal cortex (DLPFC) cluster associated with deliberative argumentation as highlighted in yellow; Right DLPFC cluster associated with negative position in argumentation as highlighted in red. The two ROIs are distinct sub-regions of DLPFC.

Combining brain and language data in the present study

In the current study, we first extend prior work that began to interrogate neural mechanisms related to counterarguing (Huskey et al., 2017; Weber, Huskey et al., 2015) as smokers are exposed to anti-smoking messages. We employ a *functional localization* approach, focusing on neural activity within key regions of interest (ROIs, hereafter) in the brain associated with core theorized dimensions of counterarguing. This approach uses an independent and targeted fMRI "localizer" task to stimulate specific cognitive processes of interest (in this case counterarguing), and the resulting areas of the brain involved during this process can then be employed to help test theoretical predictions in another task with a new, independent sample (Poldrack, 2007; Weber, Eden et al., 2015), such as exposure to persuasive messages. Our study builds on a prior *counterarguing localizer* study that localized brain ROIs that are most strongly engaged during counterarguing, i.e., two non-overlapping ROIs in sub-regions of dorsal lateral prefrontal cortex (DLPFC hereafter; O'Donnell, Coronel, Cascio, Lieberman, & Falk, 2018). As illustrated in Figure 1, the anterior and bilateral DLPFC (474 voxels) were associated with cognitive deliberation as subjects were explicitly instructed to make deliberative versus quick responses to generic behavior statements (e.g., "People should give to charity", "People should text while driving"). Second, the study identified a cluster of voxels in the right posterior DLPFC (79 voxels) associated with the negative position in argumentation, as subjects were prompted to argue against versus in favor of the same types of statements (i.e., negative deliberation). The current study makes use of these independent, functionally localized ROIs to investigate smokers' neural activity during exposure to naturalistic anti-smoking messages. Details about the localizer task can be found in Appendix A in the Online gmentary Materials.

We then link the proposed processes captured during message exposure to an automated linguistic analysis of free-form text captured following exposure to the messages. Specifically, we used the LIWC dictionary-based approach (Linguistic Inquiry and Word Count; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007), which is one of the most frequently applied and well-validated approaches of computational analysis for measuring psychological processes reflected in language use, to quantify the two components of counterarguing (i.e., deliberation and valence) in each piece of text. There are two linguistic measures in LIWC that may reflect the *degree of deliberation in the thoughts described*: the word count of each text, which captures the elaborateness in the thoughts, and the "cognitive mechanism" word category, which captures the level of cognitive effort involved in the thoughts (Pennebaker et al., 2007). Simultaneously examining the two language measures allows us to explore how the two dimensions of deliberative argumentation, i.e., breadth and depth, may be related to neural activity within brain regions of interest chosen for their prior roles in cognitive deliberation as identified by the localizer study. Secondly, to quantify the *valence of thoughts described*, we use the emotion word categories in LIWC. The categories contain positive and negative emotion words, based on which we derive the measure of "positivity" score to serve as a proxy for assessing the valence dimension of counterarguing (See Tausczik & Pennebaker, 2010 for a full discussion).

In line with prior work on counterarguing (Huskey et al., 2017; Weber, Huskey et al., 2015), we assume that those with least intention to change smoking behavior will show the strongest effects. We hypothesized that activity in the "deliberative argumentation" ROI during message exposure, would be associated with greater word counts (H1a) and "cognitive mechanism" words (H2a) in post-scan descriptions, and activity in the "negative position" ROI during message exposure, would be associated with lower positivity in post-scan descriptions (H3a). We also examined whether smokers' self-reported intention to change would moderate the above associations. Specifically, we expected that the associations between brain activity and word counts (H1b), cognitive mechanism use (H2b), and negative position (H3b) would be stronger for those who have lower intention to change their smoking behavior.

Method

Participants

Fifty adult smokers participated in this fMRI study. Six participants were excluded due to excessive head motion (n = 3), data corruption (n = 1), or missing audio recordings (n = 2), resulting in an eligible sample of 44 participants (13 female; mean age = 32.43, age range = 19–64, SD = 12.81). Eight participants had a bachelor's degree, 4 had an associate degree from a 2-year college, 21 were currently attending a 4-year college, 2 had some training from professional schools after finishing high school, and 4 had a high school or equivalent education.

Participants were recruited from the general population using Craigslist and the online platform of the University's Institute for Clinical and Health Research. To be eligible to participate in the study, participants had to complete an eligibility screening phone call. To increase our likelihood of observing the counterarguing process provoked by the anti-smoking messages, we recruited smokers who reported smoking at least 5 cigarettes per day for the past month, having been a smoker for at least 12 months, and having no intention of quitting within next 30 days. They also had to: meet the standard fMRI eligibility criteria (including having no history of neurological or psychiatric disorders, no metal in their body, no current pregnancy or breast feeding, and not currently taking psychiatric medication or illicit drugs), speak fluent English, be right-handed, and be between 18 and 65. On the day of the scan, eligible participants reported smoking an average of 13 cigarettes per day (M = 13.21, SD = 5.51; ranging from 4 to 25 cigarettes per day) and had smoked for an average of 15 years (SD = 12.23).

Stimuli

The stimuli used in the task were 23 animated anti-smoking banner ads designed to promote cessation among adult smokers who were interested in quitting, created as part of the American Legacy Foundation's *EX* campaign. The ads acknowledged the struggles that smokers face in the cessation process and provide suggestions on quitting resources (McCausland et al., 2009; Vallone et al., 2010). Given the match between the goal and target audience of the campaign, i.e., encouraging cessation among smokers, and the sample of the current study, i.e., adult heavy smokers who have no intention to quit in the next month, we consider the *EX* campaign ads as appropriate and ecologically valid stimuli which have the potential to trigger counterarguments among the participants. The 23 ads were about 17.7 s long on average (SD = 3.9 s), ranging from 13.9 to 30 s (see screenshots of the 23 ads in Appendix B of the Online Supplementary Materials).

Procedures

At baseline, which took place an average of 5 days before the fMRI scanning session, the participants completed self-reported measures including intention to change and nicotine dependence levels. During a one-hour fMRI session, participants completed our main task of interest (Cooper, Tompson, O'Donnell, & Falk, 2015) after three other tasks, which were not the focus of the current study (including localizer tasks that identify regions associated with self-relevance, and exposure to and evaluation of smoking-relevant images). During our task of interest, each participant was presented with all 23 banner ads in random order. Immediately following each ad, participants were asked to rate the effectiveness of the ad based on how much the ad makes them want to quit. They had 4 s to make a response using a five-button response box. An inter-trial fixation period (M = 4.1 s, range = 3.1–7.5 s, SD = 1.1 s) occurred between each ad. Figure 2 summarizes the procedures described above. After the scanning session, in order to capture free-form language use in a naturalistic way, we audiotaped our participants responding to the instruction: "You will be shown each banner ad again but try to remember what your initial opinion of it was. Think about how you might describe it to your friend." This method builds on prior thought-listing procedures that require participants to list thoughts in response to messages by calling to mind a naturalistic context in which such an activity often takes place (Davison et al., 1995). All audiotaped descriptions of the banner ads collected following fMRI scanning were transcribed by trained research assistants.

Linguistic measures

Participants' verbal descriptions of each banner ad were scored using the English 2007 LIWC dictionary (Pennebaker et al., 2007). We first obtained the word count or length of utterance produced by LIWC for each text, which gauges the degree of engagement; it is likely that the longer the description, the more engaged the speaker is with the message and more details are provided (Tausczik & Pennebaker, 2010). Therefore, we used this measure as a proxy for the breadth dimension of individuals' deliberative argumentation. We also selected the cognitive mechanisms and emotion word categories from LIWC for use in the analysis.[1] The cognitive mechanism category consists of 719 words and reflects various ways of people processing and interpreting information in the environment, such as offering insights (e.g., "think", "know", "consider"), establishing causality or reconstructing past events (e.g., "because", "effect", "hence"). This category has often been used to gauge the extent of cognitive efforts as

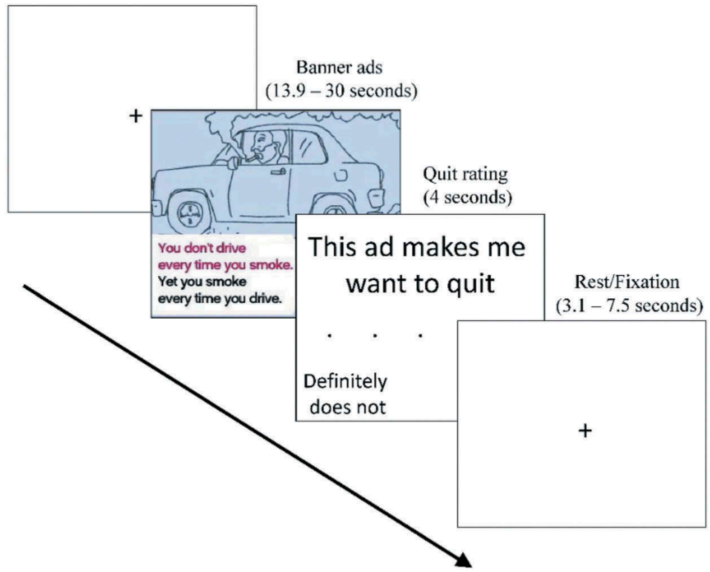


Figure 2. Banner ads fMRI task design.

people connect thoughts, reevaluate past events, and integrate among solutions when processing information, and has been found to be most indicative of subsequent behavior change (Chung & Pennebaker, 2012; Clinton, Carlson, & Seipel, 2016; Pennebaker, Mayne, & Francis, 1997; Tausczik & Pennebaker, 2010). It thus serves as a proxy for the depth dimension of individuals' deliberative argumentation in the current study. Next, we operationalized valence stance in the description using the emotion word categories in LIWC. Specifically, we focused on the positive emotion words (e.g., "good", "love", "nice", "sweet") and negative emotion words (e.g., "bad", "hurt", "ugly", "nasty") subcategories. The two categories consist of 905 words (405 positive, 500 negative). We then created a combined "positivity" score by subtracting proportions of negative emotion words from that of positive emotion words in the description. This means that texts containing more negative words than positive words produce scores on the negative side of the scale, while texts with more positive than negative words score on the positive side of the positivity scale. Scores of 0 occur either when the text contains no or a balanced number of words from the two categories. In this way, we focused on extracting the valence stance or polarity of each text. The average length of each participant's descriptions across ads was calculated to serve as a proxy for their individual tendency to generate more or fewer thoughts or describe with more or less detail in general.

Self-report measures

Intention to change (proposed moderator)

Although all eligible participants indicated no intention of quitting smoking within next 30 days at the screening stage (to increase overall likelihood of counterarguing during the study), they did differ considerably in their intentions to make changes to their smoking behavior over a longer time horizon, using a more nuanced measure of openness to change. Specifically, participants were asked "In the next three months, how likely is it that you will" (1) quit smoking completely, (2) reduce the number of cigarettes you smoke, and (3) refrain from smoking in the near future. The items were scored on a 4-point scale (1 = *definitely will not*, 4 = *definitely will*), and yielded a good reliability (Cronbach's $\alpha = .88$). They were averaged into an intention to change variable. This measure distinguishes those who were most determined to avoid any action or change related to their smoking behavior (and thus increases the likelihood of generating counterarguments to anti-smoking messages).[2]

Nicotine dependence (covariate)

Nicotine dependence levels reflect the extent to which smoking abstinence is challenged physiologically. Previous empirical evidence suggests that low-dependent and less-committed smokers are more likely to show acceptance and yielding to anti-smoking messages (Layoun et al., 2017; Loeber et al., 2011). Therefore, we controlled for individual-level nicotine addiction in all analyses. Participants' level of physiological addiction to nicotine was assessed using the standard Fagerström Test for Nicotine Dependence (FTND) with 6 items. The sum of the answers to the 6 items corresponded to the participant's nicotine dependence level: very low (0–2); low (3–4); moderate (5); high (6–7); and very high dependence (8–10) (Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991).

Evaluation of message effectiveness (covariate)

Messages perceived as convincing and effective are more likely to be processed attentively and favorably (Kang, Cappella, & Fishbein, 2006; Noar, Palmgreen, Zimmerman, Lustria, & Lu, 2010). Therefore, individuals' evaluations of message effectiveness could be correlated with the linguistic measures at the message level. Therefore, we controlled for the influence of message-level evaluations in all the analyses. Specifically, during the scanning session, immediately following exposure to each message (banner ad), participants were asked to indicate the effectiveness of how much the ad makes them want to quit with a 5-point rating scale (1 = *definitely does not*, 2 = *does not*, 3 = *neutral*, 4 = *does*, 5 = *definitely does*).[3]

fMRI acquisition and analysis

fMRI data acquisition and preprocessing

Neuroimaging data were acquired using a 3 Tesla GE Signa MRI scanner and were pre-processed according to a standard preprocessing stream including despiking, slice-timing correction, spatial realignment, smoothing, etc. (Detailed procedures are described in Appendix C of the Online Supplementary Materials).

fMRI data analysis

We examined neural activity during message exposure in relation to the subsequent language use patterns separately for each participant. We created design matrices for each participant in SPM8, modeling activity that was greater during exposure to the banner ads in the scanner, than during rest/fixation periods, with a single boxcar regressor for each ad of varying durations (13.9 s–30 s). Response periods were all modeled using one regressor of no interest. The six rigid-body translation and rotation parameters derived from spatial realignment were also included as nuisance regressors. Data were high-pass filtered with a cutoff of 128 s. The two ROIs identified by the counterarguing localizer task for deliberative argumentation and negative position were used to extract neural activity during the exposure to each banner for each participant using MarsBaR (Brett, Anton, Valabregue, & Poline, 2002). This resulted in an estimate of activity within each brain ROI for each participant during each ad, which was then combined with linguistic data, as described below. See Online Supplementary Materials Appendix D for more methodological details.

Analysis combining fMRI and linguistic data

We combined neuroimaging data of participants being exposed to the banner ads during scanning, with the linguistic data of word count, cognitive mechanism, and level of positivity in the participants' post-scan descriptions of the banner ads as quantified by LIWC. We also examined the interaction effects between neural activity and the smokers' intention to change in each of the ROI analysis models. Statistical analysis combining the two sets of data in the current study was conducted in R with the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2014) to perform linear mixed-effects modeling. We specified fixed effects of neural activity in the hypothesized ROIs, intention to change, and the interaction term between the two, controlling for each participant's average length of description, nicotine dependence level, and their evaluation of ad effectiveness.[4] The outcome variables are word counts (Models 1 & 2), cognitive mechanism scores (Models 3 & 4), and the positivity scores (Models 5 & 6) for each ad. Participants and banner ads were treated as random effects with intercepts allowed to vary randomly to account for non-independence in the data from these two sources.

Results

Language features of the post-scan descriptions

We obtained 987 transcripts in total (10 banner descriptions were missing, and 5 audio recordings had no sound), with an average length of 50.18 words (*SD* = 26.84), ranging from 1 to 162 words across all banner descriptions. Ninety-nine percent of texts contain at least one cognitive mechanism word, with a mean score of 22% (*SD* = 8%), i.e., on average, 22% of the words used were cognitive words. This indicates that on average participants used a considerable number of words that were related to thinking, reasoning, comprehension, or justification, which makes up about a fifth of what they said when describing each ad. For valence words, 74% of texts had at least one positive emotion word, and 50% had at least one negative emotion word. These numbers indicate a reasonable fit between the LIWC dictionaries and our language data. More of the texts (51%) were on the positive side of the positivity scale than on the negative side (24%), indicating that on average participants tended to use more positive than negative words in their descriptions.

Table 1 provides examples of the language used by participants with high (≥3rd quantile) and low (≤1st quantile) LIWC scores on word count, cognitive mechanism, and positivity. No significant relationship was found between participants' quit intention and word count ($\gamma = 0.02$, $t(37) = 0.13$, $p = .89$), cognitive mechanism ($\gamma = -0.02$, $t(37) = -0.04$, $p = .97$) and positivity expressed in word use ($\gamma = -0.66$, $t(37) = -0.84$, $p = .41$). Cognitive mechanism word use was not significantly associated with the positivity score in people's responses ($\gamma = 0.04$, $t(952) = 1.49$, $p = .14$), indicating that these dimensions may operate relatively independently. To further explore and triangulate the language variables, in an exploratory analysis we showed that positivity was associated with self-reported message evaluation ($\gamma = 0.09$, $t(777) = 2.88$, $p < .01$).

Self-report measures

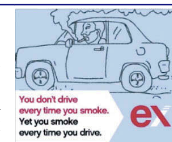
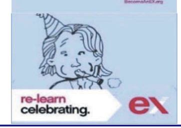
Intention to change

The mean score of the intention to change measures was 2.42 (range = 1–4; *SD* = 0.81), suggesting on average low to moderate intention to reduce or abstain from smoking in the next three months.

Covariates (nicotine dependence, message evaluation, and average word count)

Their average score on the FTND test score was 4.69 (range = 2–7; *SD* = 1.29) on the 0–10 point scale, indicating low to moderate addiction to nicotine. On average, the participants rated the ads as moderately effective in terms of how much it made them want to quit, with a mean score of 2.71 on a 5-point scale (range = 1–5; *SD* = 1.21). Each participant

Table 1. Examples of texts with high and low LIWC word count, cognitive mechanism, and positivity scores.

LIWC Category	Score	Text Examples	Banner Ads
Word Count	High[a] (135)	this one shows an animation of a car driving smoke coming out of the tail pipe and it says you dont drive every time you smoke but you smoke every time you drive pulls out and cars going down the road theres the guy sitting there with the beard smoking cigarette smoke pour out of the window from him while also pouring out the back i guess trying to show that you know you out there smoking is putting the same amount of chemicals as burning fuel in a car yeah it ends the whole relearn driving without cigarettes started with the text in red then went black then as relearn in red suppose to the driving websites big becomes an ex dot org and its in red same logo red ex with burning cigarette.	
	Low[b] (28)	again this ad seems like too much i like the whole story and everything just uh it doesnt make sense to me especially for like a banner ad.	
Cognitive Mechanism	High[a] (27.45)	i **THINK** this is a good ad **JUST BECAUSE** it has a good point um that you don't **ALWAYS** smoke **OR** that you don't **ALWAYS** drive when you smoke **AND-AND** which way **AROUND** i **THINK** it's an-interesting **THOUGHT** um **BECAUSE** i **KNOW MOST** people do smoke when they drive.	
	Low[a] (12.77)	this one has a guy driving in his car starts of the exhaust coming **OUT** of the tail pipe, **AND** then it zooms **OUT AND** he is smoking, there is **JUST** as much smoke coming **OUT** from this cigarette as there is coming from the tail pipe.	
Positivity	High[c] (7.55)	this ad was **PERFECT** this ad was exactly **PERFECT** cause a lot of people are trying to quit and that's what if first the first a few seconds of the ad the animation ad … was the **BEST** one i've seen i guess it's the **BEST** for product for whatever it is for any stop smoking cigarette ad whatever it was a **NICE** ad it was **PERFECTLY GREAT** it was the **BEST** one i've seen.	
	Low[d] (−6.52)	same thing with this ad this has like a dark aspect to it it's **WEIRD** animation is just like she's got this big thing in her mouth and no the banner ad's just **TERRIBLE** that's all i can say about it it's just a **TERRIBLE** ad.	

LIWC in-category words are bolded and in uppercase.

[a]Examples of word count (high) and cognitive mechanism (high & low) were descriptions of the "relearn driving" banner ad. The ad used cartoon animations and focused on disassociating smoking from driving, one of the common activities that would otherwise function as smoking cues, and encourage smokers to "relearn" driving without cigarettes;

[b]Example of word count (low) was a description of the "relearn annoying bosses" banner ad. The ad used cartoon animations and focused on disassociating smoking from stress, and encourage smokers to "relearn" stress reduction without cigarettes;

[c]Example of positivity in word use (high) was a description of the text-only "quit plan" banner ad which encouraged smokers to become an ex by offering help – "the free EX plan will help you track your progress". This description contained no negative emotion words but only had positive emotions words, thus the positivity score is high;

[d]Example of positivity in word use (low) was a description of the "relearn celebrating" banner ad. The ad used cartoon animations and focused on disassociating smoking from celebration and encourage smokers to "relearn" ways of celebration without cigarettes. This description contained no positive emotion words but only had negative emotions words; thus, the positivity score is low.

used 49.46 words (SD = 21.66) in their description on average, ranging from 5 to 118 words.

Hypothesis testing: Neural activity associated with language use patterns

To account for the nested nature of our data, a series of linear mixed-effects regression analyses were conducted to examine the associations between neural activity within the two ROIs during message exposure and the subsequent language patterns, with participants and ads as random effects (Ns = 44 and 23, respectively). For all three dependent variables, we examined the associations both without (Models 1, 3 and 5) and with a mean-centered interaction term between neural activity and intention to change (Models 2, 4 and 6). No multicollinearity was detected in the models (see Online Supplementary Materials Table S3 for details). As summarized in Table 2, we first examined whether participants' brain activity during message exposure in the "deliberative argumentation" ROIs in the bilateral DLPFC was associated with the word count and cognitive mechanism word use in the post-scan description, respectively, and whether intention to change moderated the associations. Our first hypothesis tests (H1a & b) focused on main effects of neural activity on post-scan language use. There was neither a significant main effect (Model 1) nor interaction effect (Model 2) of the neural activity within the "deliberative argumentation" ROI on word count. Thus, H1a and H1b were not supported. For cognitive mechanism word use, while the main effect of neural activity in the "deliberative argumentation" ROIs was not significant (Model 3; H2a), we observed a significant interaction effect between neural activity and intention to change (Model 4). H2b was supported. Further decomposition of the interaction (following Cohen, Cohen, West, & Aiken, 2003, p.564; details described in Online Supplementary Materials Appendix E) indicated that for people who had low levels of intention to change (M-1SD), more activity within the "deliberative argumentation" ROIs during banner ads exposure was positively associated with more cognitive mechanism word use while describing the ads post-scan (γ = 0.14, $t(754)$ = 2.63, p < .01); however, such

Table 2. Neural activity in the functionally localized deliberative argumentation and negative position rois interacting with intention to change on post-scan language use, controlling for evaluation of ads, nicotine dependence level and individuals' average word count.

DVs / IVs	Word Count		Cognitive Mechanism		Positivity	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Neural activity in deliberative argumentation ROIs	−.028 (.022)	.001 (.080)	0.042 (.036)	0.047 (.036)		
Neural activity in negative position ROI					−0.027 (.034)	−0.040 (.034)
Intention to change	.015 (.022)	.012 (.023)	−0.010 (.050)	−0.028 (.051)	−0.093 (.087)	−0.070 (.086)
Evaluation of ads	.008 (.022)	.008 (.022)	0.019 (.040)	0.017 (.040)	0.116** (.038)	0.118** (.038)
Nicotine dependence	−.007 (.021)	−.007 (.022)	−0.051 (.049)	−0.052 (.050)	−0.146 (.087)	−0.142 (.085)
Average word count	.790*** (.022)	.790*** (.022)	0.002 (.049)	0.000 (.050)	−0.112 (.086)	−0.107 (.085)
Deliberative argumentation ROIs × Intention to change		−.030 (.080)		−0.090* (.036)		
Negative position ROI× Intention to change						0.092** (.032)

Standardized multilevel regression coefficients γ are reported in this table; standard errors associated with each of the point estimates are in parentheses. Models 1, 3, and 5 present main effects models without the interaction term. Models 2, 4, and 6 present conditional main effects models where interaction between neural activity and intention to change are taken into consideration. Variables involved in interactions were all mean-centered before entering the regression models.
* $p < .05$, ** $p < .01$, *** $p < .001$.

association was not significant ($\gamma = -0.08$, $t(779) = -1.27$, $p = .21$) among those with high intention to change (M+ 1SD). Therefore, increasing neural activity within the "deliberative argumentation" ROIs during message exposure was associated with more cognitive mechanism word use in describing the messages only among individuals who had low intention to change (i.e., high motivation to argue against the ads). Analyses conducted to determine the specificity of the observed effects revealed that the effects were specific to the "deliberative argumentation" ROIs, since even among smokers with low intention to change (M-1SD), neural activity within the "negative position" ROI was not predictive of either word use ($\gamma = 0.02$, $t(86) = 0.35$, $p = .73$) or cognitive mechanism word use ($\gamma = -0.05$, $t(86) = -0.47$, $p = .64$).

We then conducted a similar set of analyses to examine whether neural activity within the "negative position" ROI is associated with valence stance expressed in the post-scan description, as quantified by the positivity score. We also examined the role of intention to change as a moderator. Similar patterns emerged, such that while the main effect was not significant (Model 5; H3a), a significant interaction between neural activity within the "negative position" ROI and intention to change was observed (Model 6). H3b was supported. Decomposing the interaction, we found that it was also the same group, those who had low intention to change (M-1SD), was driving the negative association between neural activity and the positive valence stance toward the messages ($\gamma = -0.13$, $t(776) = -2.63$, $p < .01$). Interestingly, this effect, albeit non-significant, was in the opposite direction for individuals with high intention to change (M+ 1SD; $\gamma = 0.08$, $t(759) = 1.58$, $p = .11$). We further discuss this pattern in the discussion section. Thus, greater neural activity within the "negative position" ROI during message exposure was associated with less positivity in post-scan descriptions only among smokers with low intention to change their smoking behavior. Similarly, we also confirmed that the level of positivity in language was selectively associated with the "negative position" ROI, as neural activity within the "deliberative argumentation" ROIs was not significantly associated with valence stance even among the high-risk smoker group ($\gamma = -0.02$, $t(86) = -0.19$, $p = .85$). In a nutshell, we observed consistent patterns from both sets of analyses, such that increasing neural activity in the respective ROIs was significantly associated with more cognitive mechanism word use and less positivity in their post-scan descriptions, only among heavy smokers who have lower intention to quit. The robustness of the findings observed above was confirmed with sets of additional sensitivity analyses (see Online Supplementary Materials Appendix F, Table S2, and Table S4).

Discussion

Counterarguing is critical to the effects of persuasive messages and campaigns. It has been conceptualized as negative, effortful/central processing of the arguments presented in a message (Dillard & Shen, 2005; Slater & Rouner, 2002). However, the difficulty of directly measuring counterarguing during message exposure limits attempts to uncover the underlying mechanisms involved. The present study combined neuroimaging and linguistic analysis, to unpack the underlying mechanism and components of counterarguing.

We show that among smokers with low intention to change their smoking behavior, brain activity during message exposure in a brain region previously engaged in "deliberative argumentation" in the bilateral DLPFC, as identified by an independent functional localizer, is associated with subsequent language patterns that indicate effortful cognitive deliberation. Furthermore, greater neural activity in a brain region associated with arguing against issues, functionally identified within the right posterior DLPFC, is significantly associated with less positivity in their post-scan descriptions of the persuasive anti-smoking messages. Although it was not entirely clear from previous studies whether the negative affect experienced during message exposure would also translate into more negativity in language use when reengaging with the messages afterward, our study suggests that more brain activity in regions chosen for their role in counterarguing is associated with greater negative language use among those who are less likely to change, but this same activity among those who are relatively more likely to change resulted in if anything slightly more positive post-scan language describing the ads.

Theoretical implications

Findings from the study demonstrate how predictions from the classic dual-process theories of counterarguing are mapped onto the neural functions in regions independently identified as being engaged in deliberative argumentation and negative position, but not length of elaboration, during deliberative processing. The results provide novel empirical evidence for two, dissociable, components involved in counterarguing, i.e., depth of thinking and negative valence of thoughts, which correspond to the neural activity in the two non-overlapping brain regions, and are selectively associated with distinct cognitions as measured by language use.

Moreover, although existing theories emphasize that during counterarguing, individuals engage in central processing or deliberative argumentation which requires careful consideration of persuasive arguments, the specific underlying processes involved in such deliberation were less known. Our results show that cognitive *depth* of thoughts, rather than sheer quantity or *breadth* of thoughts, may serve as a better indicator for the type of deliberation captured during counterarguing when people argue against vs. in favor of an argument as measured using our functional localizer task. In other words, when people argue against persuasive messages, they actively engage in analytical, evaluative, and inferential thinking process, which is well reflected in their language patterns at the time of message re-engagement after initial exposure and may result in more consequential and enduring effects on cognitions and behaviors (Eagly & Chaiken, 2005; Fransen et al., 2015). Quantity alone, however, does not capture the cognitive efforts invested in the logical quality of thoughts. Consistent with prior literature on the relationship between counterarguing and message effects (O'Keefe, 2016), we also observed that automated assessments of linguistic positivity were significantly correlated with participants' evaluations of message effectiveness. However, our results linking brain and language remained similar whether or not controlling for message evaluation (Appendix F in the Online Supplementary Materials), suggesting that neural responses are tracking related but not identical information to that captured by self-reports of message evaluation.

Our results are also consistent with work showing that individual differences in behavioral intention may bias information processing (Erceg-Hurn & Steed, 2011; Freeman et al., 2001). Here, intention to change, which determines the strength of motivation to resist persuasion influenced the type of processing that the brain engaged and how tightly coupled it was with subsequent reflections, such that the significant effects were observed only for people with low intentions to change. It is possible that when processing a counter-attitudinal message, the "deliberative argumentation" and the "negative position" ROIs are more likely to be activated when an individual is least likely to change. Thus, the activity within their ROIs may be more indicative of the subsequent deliberation as well as negative valence expressed in the descriptions of the messages. In addition, an interesting pattern we found while decomposing the observed interaction in Model 6 was that, for individuals with high intentions to change, their neural activity within the "negative position" ROI was slightly but positively correlated with the positivity score in their descriptions. This result may be explained by the argument proposed by Raju and Unnava (2006), that individuals with different levels of commitment to prior beliefs may resort to different means to reduce negative arousal provoked by counter-attitudinal messages, and that the post-scan language measures may have captured people's thoughts after dissonance resolution, rather than mirroring the online processing during initial exposure. Specifically, for those more committed individuals, more negative thoughts may be observed because they reduce discomfort through actively generating counterarguments; for those less committed individuals, more positive thoughts may be identified because they may have modified their prior attitudes to align with the position advocated by the messages. This may help explain why the main effect of neural activity was not observed in the model, but a significant interaction effect was detected. Another potentially promising area for future investigations would be to examine how stable, chronic individual differences such as smokers' tendencies in risk and appetitive reactivity in response to freedom-threatening messages or situations may influence their likelihood and degree of reactance experience (Clayton, Lang, Leshner, & Quick, 2018).

Methodological implications

Our study extends prior work on thought-listing (Cacioppo et al., 1997; Ivanov, Parker, & Dillingham, 2013; Miller & Baron, 1973; Osterhouse & Brock, 1970) by linking neural activity during message exposure with quantitative linguistic analysis of post-scan reflections in a health context. Neural activity within the a priori defined DLPFC ROIs, which were identified by an independent localizer task in a different sample of participants, allows us to link neurocognitive mechanisms of counterarguing when exposure to counter-attitudinal messages occurs, to naturalistic thoughts that follow. The functional localization approach identifies the neural regions that are most robustly associated with the specific, manipulated cognitive functions, and is considered one way to help lessen, though not eliminate, concerns related to the use of reverse inference (Poldrack, 2006).

The computational linguistic measures used in the study go beyond prior thought-listing procedures by providing a scalable assessment of thoughts at the level of language use. The use of linguistic analysis advanced our understanding of message-prompted counterarguing, such that the degree of deliberation during counterarguing depends on the quality of thoughts (i.e., cognitive depth) instead of the mere amount of thoughts. The fact that the theoretically meaningful post-scan linguistic constructions are associated with real-time brain activity in the functionally identified ROIs is important in both theoretical and applied research, considering that while language data is relatively easy to collect and analyze (particularly with computational automated methods), neuroimaging data lacks such scalability (O'Donnell & Falk, 2015).

The methodological integration of linking brain and language data adds to the body of literature investigating the underlying mechanisms and components of counterarguing during real-time message exposure that would not be possible using either method in isolation.

Practical implications and future directions

The assessment of counterarguing combining neuroimaging and linguistic analysis methods may complement and expand the capacity of qualitative (e.g., focus groups, in-depth interviews; Bradley, Thorson, Bothner, & Allen, 2000; McCausland et al., 2009) and quantitative approaches (e.g., belief-intention ranking, perceived effectiveness; Hornik & Woolf, 1999; Zhao, Strasser, Cappella, Lerman, & Fishbein, 2011) in message formative evaluation. Further, complementing past research (e.g., Petty, Williams, Harkins, & Latané, 1977), our results also emphasize that parameters and instruments that gauge the depth of thought may be key to counterarguing. The practicality of the suggestions above, however, should be interpreted in the context of limitations in the current study that suggest opportunities for future research. For example, we performed quantitative linguistic analysis with dictionary-based approach (Grimmer & Stewart, 2013). Specifically, we extracted the polarity or opinion slant expressed in each text based on the two pre-defined emotion word categories. However, these categories may not be able to capture negative valence in situations without overt presence of positive or negative words, but apparently conveying the speaker's negative stance toward the message (e.g., "*I feel it's talking to me like a child*."). In addition, although we controlled for individual differences in verbosity in our analyses, word count may still not be the optimal measure of breadth of deliberative argumentation. For example, a high word count may either indicate a thorough, extensive and detailed account of an ad or, on the other hand, a fluffy summary that is not focused on the core arguments. Similarly, while the cognitive mechanism word category in LIWC captures cognitive depth, it is not a measure of argument strength; other linguistic measures that can directly tap into the thought quality or argument strength in individuals' post-exposure descriptions would be highly informative and complement the results reported here. On a related note, similar to other thought-listing techniques, linguistic analysis relies on participants' self-reported and retrospective post-exposure descriptions. Therefore, the extent to which the participants are willing to and capable of accurately reporting their internal dialogs will affect the language measures (Cacioppo et al., 1997). Moreover, although we quantified the breadth and depth dimensions of cognitive deliberation using the LIWC measures, other brain regions might also capture these dimensions across contexts (i.e., outside of the regions identified by our counterarguing localizer). In other words, our data do not imply that breadth of elaboration cannot be associated with counterarguing, but rather that the neural processes we used to track counterarguing in this study are not associated with this linguistic dimension during subsequent message description. Finally, we chose to compare the anti-smoking ads to one another (i.e., focusing on variation among the ads, rather than whether the processes observed are specific to smoking ads, which would have only been possible with the use of an active comparison baseline condition). Future studies hoping to interpret study findings in light of potential boundary conditions (e.g., by domain or argument type), may benefit from using an active baseline (e.g., nonsmoking-related ads, or ads that differ in persuasiveness, effectiveness, or argument quality), to allow variation of intensity in message features. Future studies may consider larger and more representative samples to extend the generalizability of the findings.

It is also important to note that the DLPFC ROIs identified by our functional localizer do not comprise exhaustive or comprehensive brain bases of counterarguing, but instead represent a conservative test of processes likely to be engaged during counterarguing. Indeed, the Weber, Huskey et al. (2015) found the evidence that the interaction of message sensation value and argument strength yields significantly greater activation in the middle frontal gyrus and superior temporal gyrus among people likely to engage in counterarguing. Huskey et al. (2017) further observed that message features (message sensation values and argument strength) interacted with audience characteristics (issue involvement) to modulate persuasion network connectivity patterns in predicting perceived message effectiveness among high-risk subjects. They inferred that counterarguing was likely based on similar logic to the study reported here (i.e., that heavier users should be more likely to counterargue). Their results suggest additional brain regions that are implicated in perceptions of message effectiveness at a large scale, particularly under conditions when counterarguing is likely. The processes we identified with our localizer task may reflect differences in the methods used to identify the brain regions (their studies examined interactions between message features or between message features and audience characteristics, and focused on a different measure of perceived message effectiveness). Future research could test relationships between intersecting design elements from these studies.

In sum, our findings corroborate and extend social psychological theories by showing evidence for specific underlying components of counterarguing. The combination of neuroimaging and quantitative linguistic analysis could be utilized to examine issues ranging from identifying neural precursors of message propagation (Falk, O'Donnell, & Lieberman, 2012) to investigating the underlying moral judgment and decision-making processes leading to evaluations of fictional media products (Weber, Tamborini, Lee, & Stipp, 2008). Future work at this intersection will inform our understanding of persuasion and successful communication more broadly.

Notes

1. Each categorical word score is the proportion of words in each text belonging to the specific category.
2. While more participants were on the lower end of the scale, the intention to change variable still demonstrated sufficient variation in our sample. See Online Supplementary Materials Appendix G and Figure S1 for the frequency distribution of the variable. Zero-order correlations of the focal variables (independent, dependent, and moderator variables) at the person and ad levels are summarized in Table S1 on Online Supplementary Materials.
3. Although a more generally and frequently used multiple-item perceived message effectiveness scale (e.g., Zhao et al., 2011) would be ideal, the use of a multi-item scale was not feasible in the fMRI scanner due to limitations related to cost, fatigue, and body

movement in the scanner. We thus only used a one-item measure to assess the degree to which each ad made the smoker want to quit.

4. Sensitivity analyses which excluded these control variables from the models were also conducted. The results confirmed that the patterns observed still held stable, further indicating the robustness of our findings (see Online Supplementary Materials Table S4 for details).

Acknowledgments

We would also like to acknowledge the American Legacy Foundation for providing the banner ads. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, the FDA, or the American Legacy Foundation.

Funding

Research reported in this publication was supported by National Institutes of Health (NIH) Director's New Innovator Award (1DP2DA035156-01), and the National Cancer Institute (NCI) of the NIH and the Food and Drug Administration (FDA) Center for Tobacco Products (CTP) under Award Number P50CA179546.

ORCID

Jiaying Liu http://orcid.org/0000-0001-8461-0954

References

Berkman, E. T., & Falk, E. B. (2013). Beyond brain mapping using neural measures to predict real-world outcomes. *Current Directions in Psychological Science*, *22*, 45–50. doi:10.1177/0963721412469394

Bradley, T., Thorson, E., Bothner, V., & Allen, T. (2000). When the target audience is hostile to the behavior change: A case study in strategy development in social marketing. *Social Marketing Quarterly*, *6*(3), 35–38. doi:10.1080/15245004.2000.9961115

Brehm, S. S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control*. New York, NY: Academic Press.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002, June). Region of interest analysis using an SPM toolbox. *Presented at the 8th International Conference on Functional Mapping of the Human Brain*, Sendai, Japan.

Brown, M. G., & Gold, R. S. (2014). Cognitive responses to positively and negatively framed health messages: A thought-listing study. *Psychology, Health & Medicine*, *19*, 724–729. doi:10.1080/13548506.2013.871305

Burleson, B. R. (1987). Cognitive complexity. In J. C. McCroskey & J. A. Daly (Eds.), *Personality and interpersonal communication* (pp. 305–349). Beverly Hills, CA: Sage.

Cacioppo, J. T., & Sandman, C. A. (1981). Psychophysiological functioning, cognitive responding, and attitudes. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses to persuasion* (pp. 81–103). Hillsdale, NJ: Erlbaum.

Cacioppo, J. T., & Petty, E. (1981). Social psychological procedures for cognitive response assessment: The thought-listing technique. In T. V. Merluzzi, C. R. Glass, & M. Genest (Eds.), *Cognitive assessment* (pp. 309–342). New York, NY: Guilford.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (2007). *Handbook of psychophysiology*. Cambridge, UK: Cambridge University Press.

Cacioppo, J. T., von Hippel, W., & Ernst, J. M. (1997). Mapping cognitive structures and processes through verbal content: The thought-listing technique. *Journal of Consulting and Clinical Psychology*, *65*, 928–940. doi:10.1037/0022-006X.65.6.928

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In (Eds.), *Unintended thought* (pp. 212–252). New York, NY: Guilford Press.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York, NY: Guilford Press.

Chung, C. K., & Pennebaker, J. W. (2012). Linguistic inquiry and word count (LIWC): Pronounced "Luke," … and other useful facts. In P. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 206–229). Hershey, PA: IGI Global.

Cialdini, R. B., Levy, A., Herman, C. P., Kozlowski, L. T., & Petty, R. E. (1976). Elastic shifts of opinion: Determinants of direction and durability. *Journal of Personality and Social Psychology*, *34*, 663. doi:10.1037/0022-3514.34.4.663

Clayton, R., Lang, A., Leshner, G., & Quick, B. L. (2018). Who fights, who flees? An integration of the LC4MP and psychological reactance theory. *Media Psychology*, *22*, 545–571. doi:10.1080/15213269.2018.1476157

Clinton, V., Carlson, S. E., & Seipel, B. (2016). Linguistic markers of inference generation while reading. *Journal of Psycholinguistic Research*, *45*, 553–574. doi:10.1007/s10936-015-9360-8

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, N.J: Routledge.

Cooper, N., Tompson, S., O'Donnell, M. B., & Falk, E. B. (2015). Brain activity in self-and value-related regions in response to online anti-smoking messages predicts behavior change. *Journal of Media Psychology*, *27*, 93–108. doi:10.1027/1864-1105/a000146

Davison, G. C., Navarre, S. G., & Vogel, R. S. (1995). The articulated thoughts in simulated situations paradigm: A think-aloud approach to cognitive assessment. *Current Directions in Psychological Science*, *4*, 29–33. doi:10.1111/1467-8721.ep10770963

Dillard, J. P., & Shen, L. (2005). On the nature of reactance and its role in persuasive health communication. *Communication Monographs*, *72*, 144–168. doi:10.1080/03637750500111815

Eagly, A. H., & Chaiken, S. (2005). Attitude research in the 21st century: The current state of knowledge. In D. Albaraccín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 743–767). Mahwah, NJ: Erlbaum.

Erceg-Hurn, D. M., & Steed, L. G. (2011). Does exposure to cigarette health warnings elicit psychological reactance in smokers? *Journal of Applied Social Psychology*, *41*, 219–237. doi:10.1111/j.1559-1816.2010.00710.x

Falk, E. B. (2012). Can neuroscience advance our understanding of core questions in communication studies? An overview of communication neuroscience. In S. Jones (Ed.), *Communication at the center* (pp. 77–94). New York, NY: Hampton Press.

Falk, E. B., O'Donnell, M. B., & Lieberman, M. D. (2012). Getting the word out: Neural correlates of enthusiastic message propagation. *Frontiers in Human Neuroscience*, *6*, 1–14. doi:10.3389/fnhum.2012.00313

Fransen, M. L., Smit, E. G., & Verlegh, P. W. J. (2015). Strategies and motives for resistance to persuasion: An integrative framework. *Frontiers in Psychology*, *6*, 1–12. doi:10.3389/fpsyg.2015.01201

Freeman, M. A., Hennessy, E. V., & Marzullo, D. M. (2001). Defensive evaluation of antismoking messages among college-age smokers: The role of possible selves. *Health Psychology*, *20*, 424–433. doi:10.1037/0278-6133.20.6.424

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*, 267–297. doi:10.1093/pan/mps028

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

Heimberg, R. G., Nyman, D., & O'Brien, G. T. (1987). Assessing variations of the thought-listing technique: Effects of instructions, stimulus intensity, stimulus modality, and scoring procedures. *Cognitive Therapy and Research*, *11*, 13–24. doi:10.1007/BF01183129

Hornik, R., & Woolf, K. D. (1999). Using cross-sectional surveys to plan message strategies. *Social Marketing Quarterly*, *5*(2), 34–41. doi:10.1080/15245004.1999.9961044

Huskey, R., Mangus, J. M., Turner, B. O., & Weber, R. (2017). The persuasion network is modulated by drug-use risk and predicts anti-drug message effectiveness. *Social Cognitive and Affective Neuroscience*, *12*, 1902–1915. doi:10.1093/scan/nsx126

Ivanov, B., Parker, K. A., & Dillingham, L. L. (2013). Measuring counter-argument: A review and critique of the most popular techniques. *The International Journal of Interdisciplinary Studies in Communication*, *7*, 59–74. doi:10.18848/2324-7320/CGP

Kang, M. Y., Cappella, J., & Fishbein, M. (2006). The attentional mechanism of message sensation value: Interaction between message sensation value and argument quality on message effectiveness. *Communication Monographs*, *73*, 351–378. doi:10.1080/03637750601024164

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. Retrieved from http://CRAN.R-project.org/package=lmerTest

Layoun, N., Salameh, P., Waked, M., Aoun Bacha, Z., Zeenny, R. M., El Hitti, E., … Dramaix, M. (2017). Motivation to quit smoking and acceptability of shocking warnings on cigarette packages in Lebanon. *Patient Preference and Adherence*, *11*, 331–342. doi:10.2147/PPA.S122877

Lieberman, M. D. (2010). Social cognitive neuroscience. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 143–193). Hoboken, NJ: John Wiley & Sons, Inc.

Loeber, S., Vollstädt-Klein, S., Wilden, S., Schneider, S., Rockenbach, C., Dinter, C., … Kiefer, F. (2011). The effect of pictorial warnings on cigarette packages on attentional bias of smokers. *Pharmacology, Biochemistry, and Behavior*, *98*, 292–298. doi:10.1016/j.pbb.2011.01.010

Mandl, H., & Ballstaedt, S.-P. (1982). Effects of elaboration on recall of texts. In A. Flammer & W. Kintsch (Eds.), *Advances in psychology* (pp. 482–494). Amsterdam, Netherlands: North-Holland Publishing Co.

McCausland, K. L., Allen, J. A., Duke, J. C., Xiao, H., Asche, E. T., Costantino, J. C., & Vallone, D. M. (2009). Piloting EX, a social marketing campaign to prompt smoking cessation. *Social Marketing Quarterly*, *15*, 80–101. doi:10.1080/15245000902989121

Miller, N., & Baron, R. S. (1973). On measuring counterarguing. *Journal for the Theory of Social Behaviour*, *3*, 101–118. doi:10.1111/j.1468-5914.1973.tb00317.x

Noar, S. M., Palmgreen, P., Zimmerman, R. S., Lustria, M. L. A., & Lu, H.-Y. (2010). Assessing the relationship between perceived message sensation value and perceived message effectiveness: Analysis of PSAs from an effective campaign. *Communication Studies*, *61*, 21–45. doi:10.1080/10510970903396477

O'Donnell, M. B., Coronel, J., Cascio, C. N., Lieberman, M. D., & Falk, E. B. (2018, May). *An fMRI localizer for deliberative counterarguing. Paper presented at The Social & Affective Neuroscience Society Annual Meeting*, Brooklyn, NY.

O'Donnell, M. B., & Falk, E. B. (2015). Linking neuroimaging with functional linguistic analysis to understand processes of successful communication. *Communication Methods and Measures*, *9*, 55–77. doi:10.1080/19312458.2014.999751

O'Keefe, D. (2016). *Persuasion: Theory and research* (3rd ed.). Thousand Oaks, CA: Sage.

Osterhouse, R. A., & Brock, T. C. (1970). Distraction increases yielding to propaganda by inhibiting counterarguing. *Journal of Personality and Social Psychology*, *15*, 344–358. doi:10.1037/h0029598

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.Net.

Pennebaker, J. W., Mayne, T. J., & Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology*, *72*, 863–871. doi:10.1037//0022-3514.72.4.863

Petty, R. E., Tormala, Z. L., & Rucker, D. D. (2004). Resisting persuasion by counterarguing: An attitude strength perspective. In J. T. Jost, M. R. Banaji, & D. A. Prentice (Eds.), *Perspectivism in social psychology: The yin and yang of scientific progress* (pp. 37–51). Washington, DC: American Psychological Association.

Petty, R. E., & Cacioppo, J. T. (1979). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology*, *37*, 1915–1926. doi:10.1037/0022-3514.37.10.1915

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205. doi:10.1016/S0065-2601(08)60214-2

Petty, R. E., Williams, K. D., Harkins, S. G., & Latané, B. (1977). Social inhibition of helping yourself: Bystander response to a cheeseburger. *Personality and Social Psychology Bulletin*, *3*, 575–578. doi:10.1177/014616727700300405

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63. doi:10.1016/j.tics.2005.12.004

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*, 67–70. doi:10.1093/scan/nsm006

Potter, R. F., & Bolls, P. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. New York, NY: Routledge.

Rains, S. A. (2013). The nature of psychological reactance revisited: A meta-analytic review. *Human Communication Research*, *39*, 47–73. doi:10.1111/j.1468-2958.2012.01443.x

Rains, S. A., & Turner, M. M. (2007). Psychological reactance and persuasive health communication: A test and extension of the intertwined model. *Human Communication Research*, *33*, 241–269. doi:10.1111/j.1468-2958.2007.00298.x

Raju, S., & Unnava, H. R. (2006). The role of arousal in commitment: An explanation for the number of counterarguments. *Journal of Consumer Research*, *33*, 173–178. doi:10.1086/506298

Ringold, D. J. (2002). Boomerang effects in response to public health interventions: Some unintended consequences in the alcoholic beverage market. *Journal of Consumer Policy*, *25*, 27–63. doi:10.1023/A:1014588126336

Roberts, D. F., & Maccoby, N. (1973). Information processing and persuasion: Counterarguing behavior. In P. Clarke (Ed.), *New models for mass communication research* (Vol. 2, pp. 269–307). Beverly Hills, CA: Sage Publications.

Slater, M. D., & Rouner, D. (2002). Entertainment - education and elaboration likelihood: Understanding the processing of narrative persuasion. *Communication Theory*, *12*, 173–191. doi:10.1111/j.1468-2885.2002.tb00265.x

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24–54. doi:10.1177/0261927X09351676

Tormala, Z. L., & Petty, R. E. (2002). What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty. *Journal of Personality & Social Psychology*, *83*, 1298–1313. doi:10.1037//0022-3514.83.6.1298

Tormala, Z. L., & Petty, R. E. (2004). Resistance to persuasion and attitude certainty: The moderating role of elaboration. *Personality and Social Psychology Bulletin*, *30*, 1446–1457. doi:10.1177/0146167204264251

Vallone, D. M., Duke, J. C., Mowery, P. D., McCausland, K. L., Xiao, H., Costantino, J. C., … Allen, J. A. (2010). The impact of EX®: Results from a pilot smoking-cessation media campaign. *American Journal of Preventive Medicine*, *38*, S312–S318. doi:10.1016/j.amepre.2009.11.021

Weber, R., Eden, A., Huskey, R., Mangus, J. M., & Falk, E. (2015). Bridging media psychology and cognitive neuroscience. *Journal of Media Psychology*, *27*, 146–156. doi:10.1027/1864-1105/a000163

Weber, R., Huskey, R., Mangus, J. M., Westcott-Baker, A., & Turner, B. O. (2015). Neural predictors of message effectiveness during counterarguing in antidrug campaigns. *Communication Monographs*, *82*, 4–30. doi:10.1080/03637751.2014.971414

Weber, R., Tamborini, R., Lee, H. E., & Stipp, H. (2008). Soap opera exposure and enjoyment: A longitudinal test of disposition theory. *Media Psychology*, *11*, 462–487. doi:10.1080/15213260802509993

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*, 48–75. doi:10.1080/19312458.2010.547822

Zuwerink, J. R., & Devine, P. G. (1996). Attitude importance and resistance to persuasion: It's not just the thought that counts. *Journal of Personality and Social Psychology*, *70*, 931–944. doi:10.1037/0022-3514.70.5.93