Big Data under the Microscope and Brains in Social Context: Integrating Methods from Computational Social Science and Neuroscience

By
MATTHEW BROOK O'DONNELL
and
EMILY B. FALK

Methods for analyzing neural and computational social science data are usually used by different types of scientists and generally seen as distinct, but they strongly complement one another. Computational social science methodologies can strengthen and contextualize individual-level analysis, specifically our understanding of the brain. Neuroscience can help to unpack the mechanisms that lead from micro- through meso- to macro-level observations. Integrating levels of analysis is essential to unified progress in social research. We present two example areas that illustrate this integration. First, combining egocentric social network data with neural variables from the "egos" provides insight about why and for whom certain types of antismoking messages may be more or less effective. Second, combining tools from natural language processing with neuroimaging reveals mechanisms involved in successful message propagation, and suggests links from microscopic to macroscopic scales.

Keywords: fMRI; neuroscience; social network analysis; linguistic analysis; natural language processing; big data; computational social science

Much of the initial excitement and purported promise of big data has come from discovery of correlations between events that were difficult to observe using traditional analytical approaches and datasets. For exam-

Matthew Brook O'Donnell is a research assistant professor at the Annenberg School for Communication at the University of Pennsylvania. His research combines tools and insights from computational linguistics with neuroimaging and experimental and large-scale descriptive methods to examine how ideas and influence spread through online and offline social networks.

Emily B. Falk is an assistant professor at the Annenberg School for Communication at the University of Pennsylvania. Her research integrates cognitive neuroscience, psychology, and communication science to understand media effects and social influence at the individual, group, and population levels. Her lab examines the spread of messages, social norms, values, and culture through social networks.

DOI: 10.1177/0002716215569446

ple, data obtained from web search log files (i.e., objective records of which products people search for online during a specific timeframe) have been predictive of outcomes ranging from consumer demand for cars, travel destinations, employment trends, and the movement of financial markets ahead of and beyond the traditionally constructed models (Carrière-Swallow and Labbé 2013; Choi and Varian 2012; Preis, Moat, and Stanley 2013). One major advantage of these big data is their ability to directly observe certain kinds of behavior (e.g., car purchases, ticket purchases, etc.); however, such big data often lack the ability to pinpoint the socio-psychological mechanisms that lead to the relationships observed. We argue that affordances of new technology, such as the ability to directly examine neural activity in real time, and the ability to directly observe online behaviors, strengthen our ability to test theoretical claims and that this can be optimally achieved by combining computational social science methods with methods optimized for deep interrogation of individual psychology. Indeed, fields such as communication science, political science, and public health have long triangulated between large-scale observation of population-level phenomena and laboratory investigations of causal pathways. One defining feature of the current moment, however, is the convergence of methods for aggregating individual behaviors into macro-level trends,[1] methods for recording information about meso-level social environments, and methods for gaining deeper insights about micro-level psychological and neural mechanisms that produce individual cognitive, affective, and social characteristics and behaviors. In this discussion we circumscribe the levels of analysis that we will treat as follows:

1. Micro-level: psychological processes and neural mechanisms within an individual that are related to that individual's behavior;
2. Meso-level: proximal social contextual factors such as social network ties that influence an individual's behavior; and
3. Macro-level: large-scale patterns of observed behavior, generally formed by aggregating individual-level behaviors at a large scale.

More specifically, we focus on the promise of integrating one form of data that provides a window into the individual (micro) level processes—neural measures—with two key types of data—language and networks—commonly analyzed at meso and macro scales. These latter two types of data and associated computational analytic techniques are core elements of computational social science (Lazer et al. 2009). We argue that both neuroscience and population sciences have much to gain in this dialogue (Falk et al. 2013). From the perspective of neuroscientists, growing evidence points to the importance of social and environmental factors in shaping neural development, structure, and function (Pfeifer and Blakemore 2012; Gianaros et al. 2007; Hanson et al. 2013; Kanai et al. 2012);

NOTE: The authors gratefully acknowledge Joe Bayer, Chris Cascio, Nicole Cooper, Jason Coronel, and Teresa Pegors for their contributions to projects discussed here and for helpful feedback on ideas contained in this manuscript. This work was generously supported by an NIH Director's New Innovator Award NIH 1DP2DA03515601 (PI Falk).

however, few neuroscience investigations sample participants with respect to their social network characteristics or other variables emphasized by computational social science or make use of such data in interpreting brain function (Falk et al. 2013). Likewise, many models in computational social science acknowledge individual differences in person-level parameters and make assumptions about processes that by definition arise in the brains of individuals (cognition, affect, communication, and interpretation of social signals), but these assumptions are rarely tested (Pfeiffer et al. 2013). In the following sections, we provide examples of investigations that have begun to integrate neural measures with two (of many) types of data that form some of the backbone of computational social science. In the first example, we present data regarding media effects at macro and micro levels of analysis, and suggest that data gathered at the meso level regarding participants' social networks can help to contextualize the effects observed. Second, we present data regarding neural precursors of idea transmission and suggest ways in which linguistic analysis might scale the findings observed from micro to macro levels of analysis. We conclude by emphasizing the potential for both zooming in to consider so-called big data under a finer-focused microscope and zooming out to consider the brain in a wider meso-level social context, as well as related aggregate macro-level effects.

Example 1: Understanding the Effectiveness of Antismoking Messages

Questions about what makes messages effective span levels of analysis from the individual ([micro] e.g., What factors of the message are likely to make a person change their behavior?) to the population level ([macro] e.g., How should a mass media campaign be framed to make the greatest impact?). In this example we focus on persuasive health messages and specifically those relating to smoking behaviors. We suggest that integrating techniques from computational social science and neuroscience may help to build understanding of how mediated inputs produce behavioral outputs across levels of analysis.

The data reported here are part of a larger study of image-based ads designed to promote engagement with online smoking cessation aids. The dataset contains both a macro/population-level (logged data on clicks generated by emails sent to eight hundred thousand smokers in New York State) and a micro/individual-level component (a group of fifty smokers who underwent a neuroimaging protocol where they viewed the ads) (Falk et al. 2014). The images were modeled after the graphic warning labels for tobacco recently proposed by the U.S. Food and Drug Administration (FDA) and framed with the text "Stop Smoking. Start Living." The images portrayed either social or health consequences of smoking or the social or health benefits of quitting smoking (twelve social and twenty-eight nonsocial/health framed images). For example, an image of a smoker having to stand separately from their coworkers was classified as a negative social framing compared to a negative nonsocial/health image that showed the effects of smoking on teeth. At the population level, these images were used in an email distribution campaign in which eight hundred thousand smokers in New York State received a single email containing one of the forty images along with links to online quit-smoking resources managed by the New York Department of Public Health. Each image was distributed to twenty thousand unique smokers and the click through rate (CTR) was recorded for each image. The CTR, a measure of the effectiveness of the campaign, for the nonsocial (health) images was higher (16.2 percent) than for the socially framed images (14.2 percent) ($t(38) = 2.12$, $p = .04$) In other words, smokers receiving messages framed with health content were more likely to click to get more information than those receiving socially framed messages. These differences were not explained by measures relating to argument strength and thought favorability (Zhao et al. 2011) that were gathered from ratings by an independent group of smokers. This was a puzzle for the research team; although strong health risk messages have been demonstrated to be effective in promoting positive smoking relevant outcomes (Hammond et al. 2006), social norms are also known to be strong motivators of behavior (Cialdini and Goldstein 2004). Why then, despite producing similar levels of favorability (i.e., thoughts about quitting or not quitting smoking), did the socially framed messages not persuade smokers to seek more information in the large-scale campaign? One possibility is that the social messages failed to elicit social thinking, their key ingredient. A second possibility, however, is that the social messages may have primed social thinking (i.e., thinking about instances of social interaction), but the social referents called to mind may not have called to mind norms that would discourage smoking.

To further probe these possibilities, and examine the mechanisms involved in processing the two types of messages, we conducted a separate study in which we collected micro-level neural data in fifty smokers during real-time exposure to each of the forty images used in the population-level email campaign, as well as an additional twenty images to more evenly balance across social and health categories (twenty-seven of the sixty images were classified as social and the remaining thirty-three as nonsocial/health related). We were particularly interested in neural mechanisms given the biases inherent in retrospective self-reports of psychological reactions to messages (Noar 2006; Wilson and Nisbett 1978; Wilson and Schooler 1991). Functional magnetic resonance imaging (fMRI) allows monitoring of neurocognitive responses as they occur. The data presented here are from forty of these fifty smokers for whom we were able to collect network data as described below.

To test whether the social messages might have been ineffective due to a failure to elicit social processes, compared to the alternative hypothesis that social processes may have been elicited but failed to have the desired effect, we drew on a large body of literature that has characterized brain regions involved in one type of social thinking—considering the mental states of others, often called "mentalizing"; for example, mentalizing might involve understanding that another person has different information or beliefs than one's own and working to decode that other person's motives. Mentalizing consistently recruits neural activity in the brain's dorsomedial prefrontal cortex (DMPFC) and posterior cingulate cortex (PCC) and the tempoparietal junction (TPJ) (Atique et al. 2011; Denny et al. 2012; Saxe and Kanwisher 2003). Thus, if social and health messages did not differ in the neural response they produced within these regions, we might conclude that the social messages failed to elicit the type of mentalizing on which their effectiveness should be based. However, if social messages do elicit greater mentalizing activity, other explanations would need to be explored.

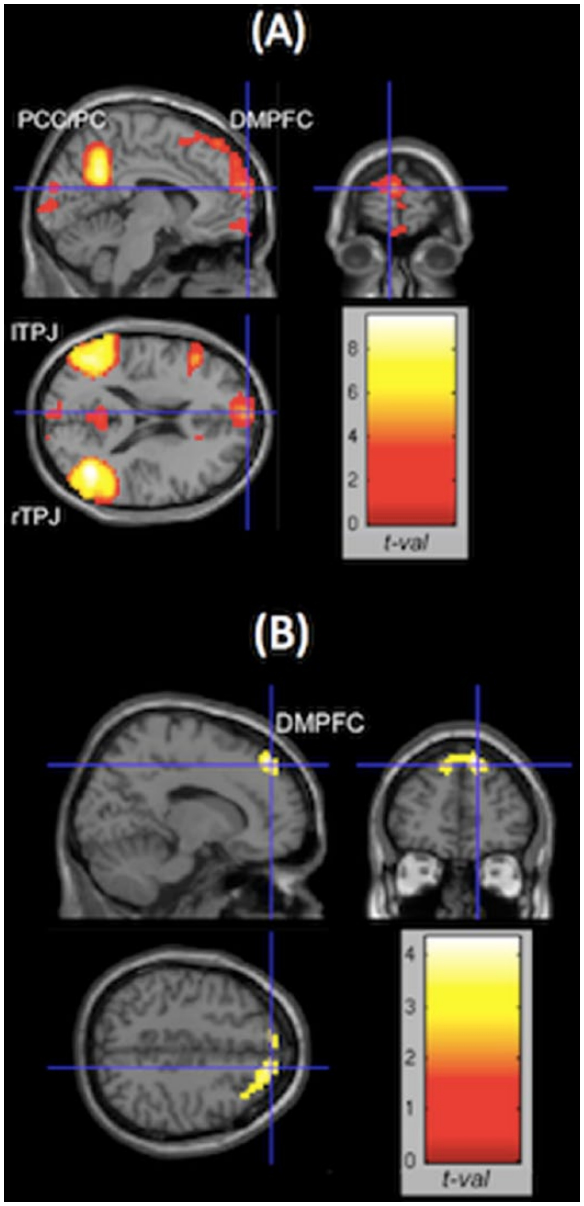
To examine these possibilities, we focused on data from the fMRI scan in which each smoker was exposed to all sixty images along with the text "Stop Smoking. Start Living." After viewing each image for four seconds, subjects made a rating responding to the prompt, "This makes me want to quit" (1 = *definitely does not* to 5 = *definitely does*). Neural activity was modeled for each subject using the general linear model using Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK). We compared neural activity during exposure to social versus nonsocial images, as well as the interaction between proportion of smokers in a smoker's ego network (a meso-level feature) and their neural response to socially framed (vs. health-framed) smoking cessation messages.

Preliminary analysis of the smokers' responses to the socially framed versus nonsocially framed messages revealed robust activity within the mentalizing network (i.e., a series of neural regions that have been associated with thinking about the mental states of others) within the DMPFC, TPJ, and the precuneus (Atique et al. 2011; Denny et al. 2012; Saxe and Kanwisher 2003) (see Figure 1A; Table 1). This is consistent with the hypothesis that the social messages did prime social thinking to a greater degree than messages focusing on personal health and other topics. This suggests that a different explanation is needed to explain the population-level response to the social images. Our alternative hypothesis was that despite activating social thoughts, the social referents called to mind may not have discouraged smoking. In line with this possibility, we hypothesized that meso-level features of a smoker's social environment might influence how smokers respond to socially framed health messages.

More specifically, we focused on egocentric networks (or egonets), which are social networks focused on a particular individual (the ego, in this case, was our smoking participants) and consist of the immediate connections (e.g., direct relationships) between the ego and other people with whom they are connected (referred to as alters) within a specific realm (e.g., an organizational context, an activity group such as a sports team, or an online social network such as Facebook) and the connections between the alters (friend-to-friend connections) (Everett and Borgatti 2005; Marsden 2002; Burt et al. 2012). Such network resources can be viewed as individual differences or personality measures (Burt 2011; Burt, Kilduff, and Tasselli 2013), and have been shown to be critical for a number of important outcomes including individual health and well-being and civic engagement (Christakis and Fowler 2008).

To quantify our participants' egonets, we developed a web-based application designed to collect the three components of individual ego networks (alters, friends they have communicated with recently; alter characteristics, whether these friends are smokers; and alter-alter connections, whether two friends know

FIGURE 1
Neural Activity



NOTE: (A) Neural activity in group of smokers (n = 40) when viewing socially framed anti-smoking images contrasted to viewing nonsocial (health related) images ($p < .005$, $k = 40$, $x = -6$, $y = 67$, $z = 19$). (B) Neural activity in DMPFC associated with higher proportion of smokers in self-reported ego network in social > nonsocial contrast ($p < .005$, $k = 40$, $x = 15$, $y = 46$, $z = 46$). DMPFC = dorsomedial prefrontal cortex; PC = precuneus; PCC = posterior cingulate cortex; TPJ = temporoparietal junction.

TABLE 1
Neural Activity in Smokers' Brains Associated with Viewing Social > Nonsocial Images

	Local Max				
Region	*x*	*y*	*z*	*K*	*t*-stat
Left TPJ	–50	–64	25	1,546	8.79
Right TPJ	42	–64	19	1,536	9.56
Precuneus	–2	–57	31	571	8.92
DMPFC	–6	67	19	356	4.97
Left IFG	–50	26	19	240	4.72
Left MFG	–50	1	52	181	3.87
Calcarine (left & right)	8	–102	4	157	3.73
Right IFG	35	15	25	126	4.08
VMPFC	1	56	–17	67	3.77
Right cerebellum	22	–78	–32	56	5.24

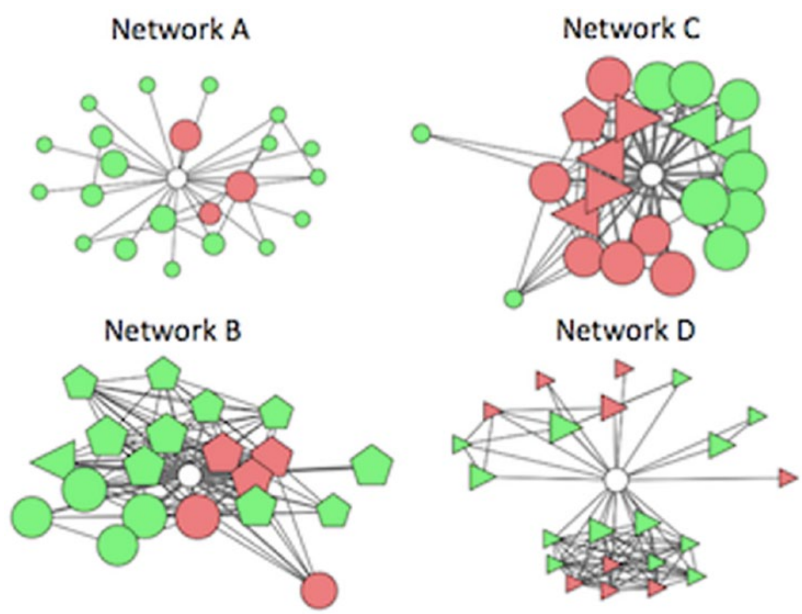
NOTE: See Figure 1A. DMPFC = dorsomedial prefrontal cortex; TPJ = temporoparietal junction; MFG = middle frontal gyrus; IFG = inferior frontal gyrus; VMPFC = ventromedial prefrontal cortex.

each other). Among other elements, this allows us to reconstruct selected features of an individual participant's social environment. In this study, we collected ego networks from the group of forty smokers using the web-based application. The smokers were asked to list the names of key family members who they felt particularly close to, up to twenty people they had called or received a call from on their cell phone in the past week, up to twenty people they had texted or received a text from in the past week, the names of up to twenty Facebook friends they had most recently interacted with (this was collected automatically for participants who gave permission), and the names of any other individuals not included in these lists but with whom they felt particularly close. They were asked to merge duplicates from across these lists and then answered a series of questions about each unique name including whether the individual was a regular smoker. Finally, they indicated whether each of the alters knew each other. Figure 2 shows the ego networks from four example smokers in the study; Table 2 summarizes some of the statistics for these networks.

Notably, the proportion of smokers within our participants' egonets varied substantially (0–80 percent, mean 29 percent, SD = 18.15), as did the position of smokers within the network (e.g., average closeness of the ego to smokers compared to nonsmokers) (in twenty-three of the thirty-five egonets that contained both smokers and nonsmokers, smokers were rated as closer than nonsmokers).

We next examined the effect of this social environmental variable on neural responses during exposure to the social images compared to health images. Figure 1B shows a cluster in the DMPFC, indicating that smokers with a greater proportion of smokers in their reported ego networks show greater levels of activity in this region when viewing socially framed images compared to when they

FIGURE 2
Example Egonets Reported by Smokers in Study



NOTE: The individual (EGO) is represented by the white node in the center. The size of friend (ALTER) nodes represents a closeness rating made using a graphical "distance from self" interface and converted to a 1 (*not at all close*) to 7 (*extremely close*) scale). Green nodes represent nonsmokers and red nodes represent friends who are smokers. The four different node shapes capture the recency (the last seen rating), with the circle for friends seen in the last week, pentagon for those seen within the last month, the left pointing triangle ("<") for in the last year (i.e., "less than a year") and the right pointing (">") for more than a year ago. The edges in the network indicate that two individuals know each other. Compare ego networks A and D both with twenty-two ALTERS. Table 2 contains a range of network measures from the four egonets.

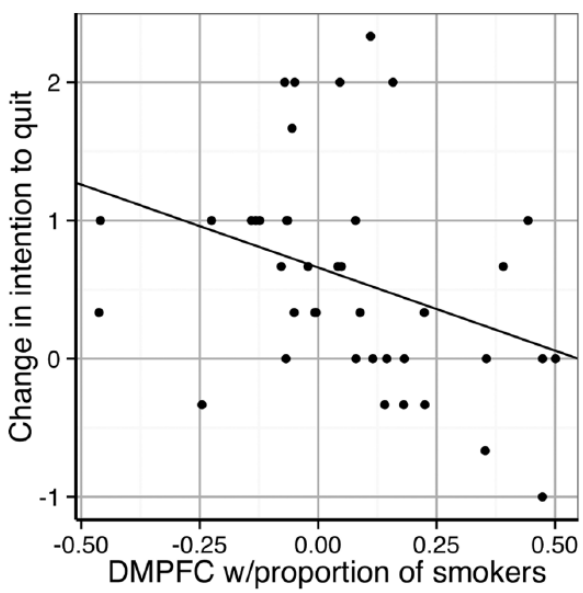
view nonsocial (health) framed images. Furthermore, activity within this specific subregion of the mentalizing system was negatively associated with changes in smokers' intentions to quit ($t(38) = -2.28$, $p = .028$) (Figure 3). Thus, smokers with a higher proportion of smokers in their ego network who engaged in greater levels of activity within the hypothesized mentalizing regions while viewing socially framed antismoking images were less likely to increase their intentions (or more likely to decrease their intentions) to quit in the month following the scan.

Although one of many possible explanations, taken together, these data are consistent with the idea that when viewing socially framed antismoking images smokers with a high proportion of smokers in their ego networks engage in greater levels of mentalizing compared to those with a lower proportion of

TABLE 2
Network Measures for Example Smoker Egonets

Measure	Network A	Network B	Network C	Network D
Friends	22	20	22	22
Smokers (%)	3 (0.14)	5 (0.25)	11 (0.5)	9 (0.41)
Average closeness	1.68	5.3	5.73	1.32
Smoker average closeness	3.33	5.6	6.18	1.11
Nonsmoker average closeness	1.42	5.2	5.27	1.46

FIGURE 3
Neural Activity Associated with Proportion of Smokers in Ego Network in DMPFC Predicts Change in Self-Reported Intention to Quit between Initial and Final Study Appointments



NOTE: Initial study appointment = preexposure; final study appointment = 1 month after scan session. $t(38) = -2.28$, $p = .028$.

smokers in their network; consideration of the mental states of others may not reduce smoking, though, precisely because a high proportion of the people that smokers are thinking about are smokers themselves.

Additional work is clearly needed to more definitively support several of the inferential leaps that we have suggested (e.g., that smokers with higher proportions of smokers in their networks automatically call these referents to mind), and to link large scale responses to campaign messages with both micro-level psychological and neurocognitive responses to messages, and to interactions with meso-level social environments. We view this example as suggestive, however, of the types of links that could be fruitful to explore further. Combining meso-views of individuals' social networks and their micro-level neural responses to social stimuli could offer insights into the mechanisms associated with behaviors observable across levels, and might also inform the design of campaigns that are optimized for the social environments of target audiences. Additional methods for connecting network measures with neural variables might include sampling strategically from larger-scale network studies or scanning individuals who are embedded within different experimentally manipulated social structures (Centola 2011; Contractor and DeChurch 2014).

Example 2: Understanding the Spread of Ideas

The spread of ideas from individual to individual and eventually to larger groups is a second area that engages analysis at the micro-, meso- and macro-levels. At the macro-level, tools from computational social science, specifically those that quantify language (e.g., categorical word counts, sentiment analysis, and other more sophisticated tools from the field of natural language processing; see Grimmer and Stewart [2013] for an overview), have been used to examine patterns of language associated with ideas that go viral in social media, news items that are more likely to be shared, and features of tweets that are associated with higher probability of being retweeted (Bhattacharya and Ram 2012; Go, Bhayani, and Huang 2009; Gruzd, Doiron, and Mai 2011; Suh et al. 2010). With large volumes of data, automated analysis of language to identify these features is necessary. Opinion mining or sentiment analysis tools, particularly those making use of machine learning approaches, have been particularly popular for synthesizing language data at large scales to extract meaning. For example, both the volume and sentiment (measured using linguistic analysis) of weblog discussion of movies has been shown to predict box office performance (Asur and Huberman 2013; Mishne 2006). Higher volumes of positive evaluative discussion (e.g., Tweets) about a movie are associated with better ongoing performance at the box office. Findings such as these cannot, however, speak in any detail to questions about the specific mechanisms involved at the individual level. What factors make an idea particularly salient to an individual who subsequently describes it to others in an enthusiastic manner?

Psychologists have begun to document the features of communicators, messages, and recipients that make it most likely that ideas will successfully spread from one person to the next (Berger and Schwartz 2011). Yet these investigations are sometimes limited by individuals' lack of ability to identify the internal mental processes that precede successful sharing. As noted above, functional neuroimaging allows recording of neural activity in real time during a psychological experience such as learning about an idea that one might later share with others. We, and others, have leveraged this ability to examine the underlying neurocognitive processes that take place during initial idea exposure and go on to predict successful idea sharing.

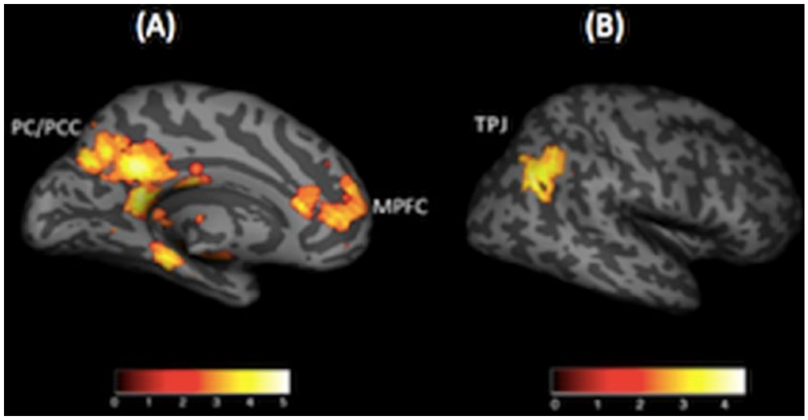
For example, in one recent study (Falk, O'Donnell, and Lieberman 2012), we examined neural processes associated with enthusiastic propagation of ideas from one person to the next using a combination of fMRI data and automated linguistic sentiment analysis. Our goal was to determine what types of cognitive processes are active that go on to predict later enthusiastic sharing of ideas. By employing automated sentiment analysis to characterize this enthusiasm, we positioned the study to link what happens in the brains of a small number of participants to larger-scale enthusiasm expressed in other contexts.

More specifically, in a message propagation task, we scanned participants who were directed to imagine they were acting as interns in a TV production company. Their role was to read ideas for potential new shows and to indicate whether they would recommend the show to their producer. After seeing twenty-four show ideas during a scanning session they were asked to talk about each show on video so that their producer could respond to their recommendations. The videos produced by the participants were transcribed and classified using a sentiment analysis (SA) algorithm trained on texts from a movie review corpus. The SA consisted of two separate binary classifiers. One distinguished between evaluative and descriptive texts, trained using texts from the IMDb (descriptive) and Rotten Tomatoes (evaluative) websites. And the second captured polarity (i.e., positive/negative) for evaluative texts, trained using the star ratings and text from a standard movie review corpus (Pang, Lee, and Vaithyanathan 2002). The SA returned both category labels (e.g., neutral, for descriptive texts, or positive/negative for evaluative texts) and classification probabilities.

We combined these quantitative language scores with the neural activity recorded when participants were exposed to the TV show ideas and examined which areas of the brain were associated with (1) the use of positive language overall and (2) highly evaluative positive language (see Figure 4). Show descriptions with high positivity scores showed an association with increased neural activity in brain regions that are most commonly associated with self-related processes, medial prefrontal cortex (MPFC) and the PCC and precuneus (PC) (Lieberman 2010). These data are consistent with the possibility that neural activity connected with participants' own traits and motivations (e.g., "I like this show" or "This show is relevant to me") may have positioned participants to later talk about show ideas using more highly positive language. That is, participants who found specific TV show ideas particularly self-relevant may have tended to use language patterns in their subsequent description of those shows that are associated with highly positive recommendation reviews.

Next we combined the two scores from the SA (evaluative*positivity) to produce a score between –1 (*highly negatively evaluative*) and 1 (*highly positively evaluative*). This score relates to show descriptions that are strongly evaluative and positively framed. We found that more positive and evaluative language was associated with neural activity in one of the primary regions discussed in example one as supporting mental state inference—right TPJ. In this context, the tendency to consider the social value of the idea—"Will this idea appeal to others?"

FIGURE 4
Linking Linguistic Measures with Neural Data in the Context of Message Propagation



SOURCE: Figure originally published in Falk, O'Donnell, and Lieberman (2012).
NOTE: (A) Neural activity associated with higher positivity scores from automatic SA classification ($p < .005, k = 37$). (B) Neural activity associated with higher combined evaluative*positivity scores from automatic SA classification ($p < .005$, $k = 37$) MPFC = medial prefrontal cortex; PC = precuneus; PCC = posterior cingulate cortex; TPJ = temporoparietal junction.

or "I think this would appeal to lots of people/these types of people"—may have positioned participants to package the ideas in ways that would communicate subjective opinions to others through more evaluative language (i.e., a strong recommendation of the idea).

These data provide evidence to establish a link between the neural mechanisms involved in initial idea evaluation and the (linguistic) manner in which it is retransmitted. Our use of automated linguistic tools was designed to lay the foundation for moving from the micro-level to the meso-level of person to person sharing where language becomes a predominant carrier of information and social intent and eventually to more direct connections with larger-scale retransmission. Future studies might directly manipulate the psychological processes suggested by this initial work (i.e., mentalizing) and observe the extent to which it changes later enthusiastic propagation.

Conclusion

Although in early phases of development, the combination of neural and computational social science tools may allow new connections among multiple levels of analysis. The examples presented here are highly suggestive of the potential of integrating data representing the neurocognitive processes of individuals engaged in laboratory tasks, with analytic methods from computational social science such as social network and quantitative linguistic analysis. In our first example, neural and egonet data suggest a novel explanation for why certain socially framed messages may not have achieved the large-scale media effects expected; in turn, this highlights a broader theoretical issue regarding interactions between individual message recipients and their close social referents. This example also highlights that the individuals studied in social neuroscience live surrounded by social structures that can be quantified using the tools of computational social science, with neural activity in key brain systems being moderated by social network variables. More broadly, integration of neural and computational social science tools may offer a richer picture of how the brain works and could reveal other important interactions between neural function and broader social environments. Together, these combinations provide contextualization and potential explanations of the patterns and association discovered at scale.

Our second example highlights the use of a different tool from the computational social science toolbox (linguistic analysis) to help link processes that can be measured both in controlled laboratory settings and at large scales in the real world. As a primary communicative tool, language reflects psychological processes experienced by the communicator (Gonzales, Hancock, and Pennebaker 2010; O'Donnell, Falk, and Lieberman 2015; Pennebaker et al. 2007). Practically, it can be collected in controlled laboratory studies, and also makes up a significant component of the computational social science/big data record that results from individuals interacting and engaging in a social contexts (O'Connor et al. 2010). Thus, language is an ideal tool to link such levels of analysis. In current work, our team is extending the findings described here to more directly link neural activity and language expressed at the micro level with larger-scale data on selection and retransmission (of news articles).

The examples reviewed here are just initial investigations of potential methodological combinations, and further work is needed to pinpoint the potential as well as boundaries and limitations. At a practical level these goals can be achieved through strategic partnerships among researchers who traditionally work at different levels of analysis, and through the addition of biological (e.g., neuroimaging) measures to selected subsamples of population level computational social science (i.e., larger representative surveys and big data–style scrapping [Falk et al. 2013]). In addition, computational social science methods can be applied to data collected on individuals within upcoming neuroimaging studies. Beyond the benefits for each set of disciplines on their own, collecting such data across levels of analysis will also aid in explicitly theorizing about both how observed macro processes might be the product of specific neural mechanisms and how different neural mechanisms might be moderated by macro-level processes.

Note

1. We wish to emphasize the importance of considering other macro-level social structures and institutions that influence both individuals and populations (e.g., of the type traditionally studied by sociologists and that may not consist only of aggregate individual behaviors); however, treatment of these ideas and interactions is beyond the scope of this article. In addition, treatment of the vast technology developing to

understand meso-level social environments (Swan 2013) is similarly essential to a complete picture of social, cognitive, and affective processes, but beyond the scope of this article.

References

Asur, Sitaram, and Bernardo A. Huberman. 2013. Predicting the future with social media. *Applied Energy* 112 (December): 1536–43.

Atique, Bijoy, Michael Erb, Alireza Gharabaghi, Wolfgang Grodd, and Silke Anders. 2011. Task-specific activity and connectivity within the mentalizing network during emotion and intention mentalizing. *NeuroImage* 55 (4): 1899–1911.

Berger, Jonah, and Eric M Schwartz. 2011. What drives immediate and ongoing word of mouth? *Journal of Marketing Research* 48 (5): 869–80.

Bhattacharya, Devipsita, and Sudha Ram. 2012. Sharing news articles using 140 characters: A diffusion analysis on Twitter. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, 966–71. Washington, DC: IEEE Computer Society.

Burt, Ronald S. 2011. *Neighbor networks: Competitive advantage local and personal*. Reprint ed. New York, NY: Oxford University Press.

Burt, Ronald S., Martin Kilduff, and Stefano Tasselli. 2013. Social network analysis: Foundations and frontiers on advantage. *Annual Review of Psychology* 64 (1): 527–47.

Burt, Ronald S., David O. Meltzer, Michael Seid, Amy Borgert, Jeanette W. Chung, Richard B. Colletti, George Dellal, Stacy A. Kahn, Heather C. Kaplan, Laura E. Peterson, and Peter Margolis. 2012. What's in a name generator? Choosing the right name generators for social network surveys in healthcare quality and safety research. *BMJ Quality & Safety* 21 (12): 992–1000.

Carrière-Swallow, Yan, and Felipe Labbé. 2013. Nowcasting with Google Trends in an emerging market. *Journal of Forecasting* 32 (4): 289–98.

Centola, Damon. 2011. An experimental study of homophily in the adoption of health behavior. *Science* 334 (6060): 1269–72.

Choi, Hyunyoung, and Hal Varian. 2012. Predicting the present with Google Trends. *Economic Record* 88 (June): 2–9.

Christakis, Nicholas A., and James H. Fowler. 2008. The collective dynamics of smoking in a large social network. *New England Journal of Medicine* 358 (21): 2249–58.

Cialdini, Robert B., and Noah J. Goldstein. 2004. Social influence: Compliance and conformity. *Annual Review of Psychology* 55:591–621.

Contractor, Noshir S., and Leslie A. DeChurch. 2014. Integrating social networks and human social motives to achieve social influence at scale. *Proceedings of the National Academy of Sciences* 111 (Suppl. 4): 13650–57.

Denny, Bryan T., Hedy Kober, Tor D. Wager, and Kevin N. Ochsner. 2012. A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience* 24 (August): 1742–52.

Everett, Martin, and Stephen P. Borgatti. 2005. Ego network betweenness. *Social Networks* 27 (1): 31–38.

Falk, Emily B., Luke W. Hyde, Colter Mitchell, Jessica Faul, Richard Gonzalez, Mary M. Heitzeg, Daniel P. Keating, Kenneth M. Langa, Meghan E. Mertz, and Julie Maslowsky, et al. 2013. What is a representative brain? Neuroscience meets population science. *Proceedings of the National Academy of Sciences* 110 (October): 17615–22.

Falk, Emily B., Matthew Brook O'Donnell, and Matthew D Lieberman. 2012. Getting the word out: Neural correlates of enthusiastic message propagation. *Frontiers in Human Neuroscience* 6:313. doi:10.3389/fnhum.2012.00313.

Falk, Emily B., Matthew Brook O'Donnell, Steven Tompson, Richard Gonzales, Sonya Dal Cin, Victor Strecher, and Lawerence C. An. 2014. Neural systems associated with self-related processing predict population success of health messages. Paper presented at the 64th Annual Conference of the International Communication Association Communication, Seattle, WA.

Gianaros, Peter J., Jeffrey A. Horenstein, Sheldon Cohen, Karen A. Matthews, Sarah M. Brown, Janie D. Flory, Hugo D. Critchley, Stephen Manuck, and Ahmad R. Hariri. 2007. Perigenual anterior cingulate morphology covaries with perceived social standing. *Social Cognitive and Affective Neuroscience* 2 (September): 159–60.

Go, Alec, Richa Bhayani, and Lei Huang. 2009. *Twitter sentiment classification using distant supervision*. CS224N Project Report. Stanford, CA: Stanford University.

Gonzales, Amy L., Jeffrey T. Hancock, and James W. Pennebaker. 2010. Language style matching as a predictor of social dynamics in small groups. *Communication Research* 37 (1): 3–19.

Grimmer, Justin, and Brandon M. Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21 (3): 267–97.

Gruzd, Anatoliy, Sophie Doiron, and Philip Mai. 2011. Is happiness contagious online? A case of Twitter and the 2010 Winter Olympics. In *2011 44th Hawaii International Conference on System Sciences (HICSS)*, 1–9. Manoa, HI.

Hammond, David, Geoffrey T. Fong, Ann McNeill, Ron Borland, and K. Michael Cummings. 2006. Effectiveness of cigarette warning labels in informing smokers about the risks of smoking: Findings from the International Tobacco Control (ITC) Four Country Survey. *Tobacco Control* 15 (Suppl 3., June): iii19–iii25.

Hanson, Jamie, Nicole Hair, Amitabh Chandra, Ed Moss, Jay Bhattacharya, Seth D. Polk, and Barbara Wolfe. 2013. Brain development and poverty: A first look. In *The biological consequences of socioeconomic inequalities*, eds. Barbara Wolfe, W. Evans, and T. E. Seeman. New York, NY: Russell Sage Foundation.

Kanai, Ryota, Bahador Bahrami, Brad Duchaine, Agnieszka Janik, Michael J. Banissy, and Geraint Rees. 2012. Brain structure links loneliness to social perception. *Current Biology* 22 (20): 1975–79.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, and Myron Gutmann, et al. 2009. Computational social science. *Science* 323 (5915): 721–23.

Lieberman, Matthew D. 2010. Social cognitive neuroscience. In *Handbook of social psychology*, 5th ed., eds. Susan T. Fiske, Daniel T. Gilbert, and Gardner Lindzey, 143–93. New York, NY: McGraw-Hill.

Marsden, Peter V. 2002. Egocentric and sociocentric measures of network centrality. *Social Networks* 24 (4): 407–22.

Mishne, Gilad. 2006. Predicting movie sales from blogger sentiment. In *Proceedings of the AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*. Stanford, CA: AAAI. Available from http://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-03/SS06-03-030.pdf.

Noar, Seth M. 2006. A 10-year retrospective of research in health mass media campaigns: Where do we go from here? *Journal of Health Communication* 11:21–42.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan Routledge, and Noah Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*. Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

O'Donnell, Matthew Brook, Emily B. Falk, and Matthew D. Lieberman. 2015. Social in, social out: How the brain responds to social language with more social language. *Communication Monographs* 82 (1): 31–63.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, 79–86. Stroudsburg, PA: Association for Computational Linguistics.

Pennebaker, James W., Cindy K. Chung, Molly Ireland, Amy Gonzales, and Roger J. Booth. 2007. *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net.

Pfeifer, Jennifer H., and Sarah-Jayne Blakemore. 2012. Adolescent social cognitive and affective neuroscience: Past, present, and future. *Social Cognitive and Affective Neuroscience* 7 (January): 1–10.

Pfeiffer, Ulrich J., Bert Timmermans, Kai Vogeley, Chris Frith, and Leonhard Schilbach. 2013. Towards a neuroscience of social interaction. *Frontiers in Human Neuroscience* 22. doi:10.3389/fnhum.2013.00022.

Preis, Tobias, Helen Susannah Moat, and H. Eugene Stanley. 2013. Quantifying trading behavior in financial markets Using Google trends. *Scientific Reports* 3 (April). doi:10.1038/ srep01684.

Saxe, Rebecca, and Nancy Kanwisher. 2003. People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *Neuroimage* 19:1835–42.

Suh, Bongwon, Lichan Hong, Peter Pirolli, and Ed H. Chi. 2010. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *2010 IEEE Second International Conference on Social Computing (SocialCom)*, 177–84. Washington, DC: IEEE Computer Society.

Swan, Melanie. 2013. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data* 1 (2): 85–99.

Wilson, Timothy De Camp, and Richard E. Nisbett. 1978. The accuracy of verbal reports about the effects of stimuli on evaluations and behavior. *Social Psychology* 41:118–31.

Wilson, Timothy De Camp, and Jonathan W. Schooler. 1991. Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology* 60:181–92.

Zhao, Xiaoquan, Andrew Strasser, Joseph N. Cappella, Caryn Lerman, and Martin Fishbein. 2011. A measure of perceived argument strength: Reliability and validity. *Communication Methods and Measures* 5:48–75.