Marketing Science Institute Working Paper Series 2010
Report No. 10-114

Social Transmission, Emotion, and the Virality of Online Content

Jonah Berger and Katherine L. Milkman

Report Summary

Why are certain pieces of online content more viral than others? Companies often create websites, online ad campaigns, or videos in the hopes that consumers will share them with others. Some of these attempts succeed while others fail. Is success just random, as some have argued, or might certain characteristics predict whether content will be viral?

In this report, Jonah Berger and Katherine Milkman take a psychological approach to understanding diffusion. Using a unique dataset of all the *New York Times* articles published over a three-month period, they examine the link between integral affect (i.e., the emotion evoked) and whether content is highly shared. Specifically, they examine how content valence (i.e., whether an article is more positive or negative) as well as the specific emotions it evokes (anxiety, anger, awe, disgust, sadness) relate to whether content is highly shared. Their data include information about the content of each *Times* article published online over a three-month period, and whether it made the newspaper's "most emailed" list.

Their results suggest a strong relationship between emotion and virality: affect-laden content—regardless of whether it is positive or negative—is more likely to make the most emailed list. Further, positive content is more viral than negative content; however, this link is complex. While more awe-inspiring and more surprising content are more likely to make the most emailed list, and sadness-inducing content is less viral, some negative emotions are positively associated with virality. More anxiety- and anger-inducing content are both more likely to make the most emailed list. In fact, the most powerful predictor of virality in their model is how much anger an article evokes. There was no significant relationship between disgust and virality.

These results hold controlling for how surprising, interesting, or practically useful content is (all of which are positively linked to virality), as well as external drivers of attention (e.g., how prominently articles were featured or the fame of the article's author).

While common wisdom assumes that people tend to pass along negative news more than positive news, these results indicate that, in general, positive news is actually more viral. While being featured prominently or being written by a famous author increases the likelihood that articles are highly shared, these results suggest that content characteristics are of similar importance.

These findings shed light on how to design successful viral marketing campaigns and craft contagious content. Given that affect-laden and awe-inspiring content is more likely to be shared, for example, campaigns that strive to evoke these emotions may be more successful. Similar points apply to managing consumer sentiment online, particularly when it is negative. Not all negative emotions are the same when it comes to sharing and some may be more likely to incite transmission. Brand transgressions that evoke anxiety or anger, for example, may be more likely to be shared than those that evoke sadness; this suggests that companies may want to take a more active role in managing situations that evoke these emotions.

Jonah Berger is Assistant Professor of Marketing and Katherine L. Milkman is Assistant Professor of Operations and Information Management, both at the Wharton School, University of Pennsylvania.

Sharing online content is an integral part of modern life. We forward newspaper articles to our friends, pass YouTube videos to our relatives, and send restaurant reviews to our neighbors. Indeed, 59% of people say they frequently share online content with others (Allsop, Bassett, and Hoskins 2007), and someone tweets a link to a *New York Times* story once every four seconds (Harris 2010). Such social transmission has important implications for both consumers and brands. Decades of research suggest that interpersonal communication affects attitudes and decision making (Asch 1956; Katz and Lazarsfeld 1955), and recent work has demonstrated the causal impact of word-of-mouth on product adoption and sales (Chevalier and Mayzlin 2006; Godes and Mayzlin 2009).

But while it is clear that social transmission is both frequent, and important, less is known about why certain pieces of online content are more viral than others. Some ads are frequently shared while others are not. Some newspaper articles earn a position on their website's "most emailed list" while others languish. Companies often create websites, online ad campaigns, or videos in the hopes that consumers will share them with others, but some of these attempts succeed while others fail. Is success just random, as some have argued (Cashmore 2009; also see Salganik, Dods, and Watts 2006), or might certain characteristics predict whether content will be viral?

This paper examines the link between integral affect (i.e., the emotion content evokes) and virality. We do so through analyzing a unique dataset of nearly 7,000 articles from one of the world's most popular newspapers: *The New York Times*. The *Times* covers a wide range of topics (i.e., world news, sports, and travel) and boasts the most frequented website of any newspaper in the U.S. (Nielson, 2008), making it an ideal venue for examining what types of online content are most frequently shared. Our data includes information about the content of each *Times* article published online over a three month period, and whether it made the newspapers “most emailed list.” Controlling for external drivers of attention, such as where an article was featured online, we test the relationship between emotion and virality. Specifically, we examine how content valence (i.e., whether an article is more positive or negative) as well as the specific emotions it evokes (e.g., anger, anxiety, and awe) relates to whether it is highly shared.

This research makes a number of important contributions. First, while there has been a great deal of recent research on word-of-mouth and viral marketing (Godes and Mayzlin 2004; 2009; Goldenberg, Libai, and Muller 2001; Stephen, Dover, and Goldenberg 2010; Wojnicki and Godes 2008), most of this work has focused on the causal impact of social transmission (e.g., on sales) and has given less attention to what types of things are shared. By looking at real transmission across a range of topics in a naturalistic setting, this investigation is the first to demonstrate characteristics of online content that are linked to virality. Second, our findings shed light on a relatively unexamined (and important) domain of consumer behavior. People often share online content, and the emergence of social media (e.g., Twitter and Facebook) has only increased the speed and reach of content dissemination. Yet little work has examined why people share certain things rather than others (though see Cheema and Kaikati 2010; Stephen and Lehmann 2010). Our results provide preliminary insight into potential underlying processes that drive people to share. Finally, our research sheds light on how to design successful word-of-mouth or viral marketing campaigns. Organizations now use social media (e.g., blogs) and viral marketing (e.g., YouTube videos) to increase customer engagement. These methods are not only cheaper than buying ads on television, but are thought to be more effective than traditional marketing channels. The effectiveness of these tactics, however, hinges on the ability to create content that consumers actually share. Marketers or organizations can put an ad on YouTube, for example, but if no one shares it, the benefit of social media is lost. Consequently, understanding what types of things are highly shared can help organizations and policy makers craft contagious content.

Content Characteristics and Social Transmission

Online content can provide information and evoke emotion. While our focus is on the link between emotion and social transmission, we first briefly discuss the role that information plays in social transmission before elaborating on how emotional characteristics of content may be associated with virality.

One reason content may be highly shared is because it is contains useful information. Articles about where to find low airfares or good restaurants help people save money and eat better. Consequently, consumers may share practically useful content for altruistic reasons (e.g., to help others) or for self-enhancement purposes (e.g., to appear knowledgeable, see Wojnicki and Godes 2008). Practically useful information also has social exchange value (Homans 1958), and people may share it to generate reciprocity (Fehr, Kirchsteiger, Riedl 1998).

Emotion and virality

Beyond the information it provides, however, we argue that emotional aspects of content may impact whether they get shared. People report discussing many of their emotional experiences with others, and the social sharing of emotion serves a variety of functions. First, emotional stimuli often elicit ambiguous sensations, and through talking about and sharing emotional content with others, people can gain a deeper understanding of how they feel (Rime, Mesquita, Philippot, and Boca 1991). Second, to the extent that emotional material challenges people's beliefs or way of seeing the world, they may share it with others to cope or reduce feelings of dissonance (Festinger, Riecken, and Schachter 1956). Third, sharing emotional content can strengthen social bonds, and people may share with others to deepen social connection (Peters and Kashima 2007). Consequently, we suggest that more emotion-laden content, regardless of valence, should be more viral.

But which is more likely to be highly shared, positive or negative content? While there is a lay belief that people are more likely to pass along negative news (Godes et al 2005), this has never actually been tested. Further, the study on which this idea is based actually focused on understanding what types of news people encounter, not what they transmit (see Goodman 1999; Godes et al 2005). Consequently, researchers have noted that "more rigorous research into the relative probabilities of transmission of positive and negative information would be valuable to both academics and managers" (Godes et al. 2005, p. 419), yet little work has examined the link between content valence and virality.

While some research suggests that negative information receives more attention, we hypothesize that positive content will be more viral. The old news business adage that "if it bleeds, it leads," is based on the idea that negative information generates more attention and interest. Negative information has survival value because it tells people what to avoid (e.g., that restaurant will make you sick), and across a range of domains bad things have a stronger impact than good ones (see Baumeister et al. 2001 for a review). That said, when considering transmission, there are a number of reasons to believe that positive content should be more viral. Consumers often share things to self-enhance (Wojnicki and Godes 2008) or communicate identity, and consequently positive things may be shared more because they reflect positively on the self. Most people would prefer to be known as someone who shares upbeat stories or makes others feel better rather than someone who shares things that makes people angry or upset. Further, people may share positive content to help boost the recipients' mood or provide information about potential rewards (i.e., this restaurant is worth trying).

Specific emotions. In addition to valence, emotional content also varies along a host of other dimensions (e.g., arousal or uncertainty; Smith and Ellsworth 1985). Anger and sadness are both negative emotions, for example, but while anger is characterized by a state of heightened activation, sadness is characterized by low arousal or deactivation (Barrett and Russell 1998). Such differences have important downstream effects, leading even emotions of the same valence to have different effects on behavior (Lerner and Keltner 2001; Lerner, Small, and Lowenstein 2004).

Consequently, we examine the link between specific emotions and virality. Negative emotions are better distinguished from one another and more generalized mood states (Keltner and Lerner 2010), and the basic or universal negative emotions are anger, anxiety, disgust, sadness, and surprise (Ekman, Friesen, and Ellsworth 1982; Sauter et al. 2009; though surprise may be positive in some cases). The link between surprise and virality seems relatively straightforward. Surprising things tend to be novel and entertaining. Consequently, consistent with self-enhancement motivations for transmission (Wojnicki and Godes 2008), people may share surprising content to inform others or because it reflects positively on the sharer (i.e., they know about interesting things). This suggests that more surprising content should be more viral.

The link between other specific negative emotions and virality, however, is less clear. One simple theory would be that content which evokes most negative emotions will be less viral.

As noted above, people usually want to make their friends feel good rather than bad, and may avoid passing along content that will make the recipient angry or disgusted. Similarly, people may avoid sharing sad or anxiety inducing content because it could result in negative halo effects and harm their image. That said, when we asked a cross-national sample of individuals (N = 57, mean age = 49) how evoking anger, anxiety, sadness, or disgust would influence their likelihood of sharing a newspaper article (7-point scale: -3 = greatly decreases likelihood of sharing, 3 = greatly increases likelihood of sharing), their responses were quite divided. While 40% of respondents thought they would be more likely to share stories that evoked anger, another 40% thought they would be *less* likely to share stories that evoked anger. Responses to anxiety (40% believed it would increase sharing, 37% decrease), sadness (44% believed it would increase sharing, 21% decrease), and disgust (47% believed it would increase sharing, 32% decrease) were similarly mixed. This indicates that it is not entirely clear how evoking anger, anxiety, sadness, or disgust impacts whether content is shared.

We suggest that even though they are negatively valenced, content that evokes anger, anxiety, or disgust will be more viral. While anger, anxiety, and disgust are negative emotions, they are characterized by high activation or arousal which drives people to action (Barrett and Russell 1998). Consequently, content that evokes these emotions may be more likely to be shared. Consistent with this, there are many examples of online social movements driven by anger towards a certain company, and the largest class of rumors are "wedge-drivers" that provoke anger towards specific demographic groups (Knapp 1947). Similarly consumers often spread urban legends that foster anxiety (e.g., Mountain Dew lowers sperm count, Brunvand 1981), and generalized anxiety is thought to boost circulation of social information more generally (Heath, Bell, and Sternberg 2001; Rosnow 1980). Finally, people report higher willingness to share urban legends that are more disgusting (Heath et al. 2001).

That said, we suggest that the relationship between sadness and virality will be negative. Sadness is characterized by deactivation (Barrett and Russell 1998) and associated with feeling lost and helpless (Lazarus 1991). Sad people often close up and become withdrawn. These characteristics should reduce the chance that people share content with others and lead sad content to be less viral.

Positive emotions are more difficult to distinguish from one another than negative emotions, but we also consider whether awe inspiring content is highly viral.[1] Awe is the emotion of self-transcendence characterized by a feeling of admiration and elevation in the face of something greater than the self (e.g., experiencing a beautiful work of art, Keltner and Haidt 2003). Stimuli that open the mind to unconsidered possibilities can inspire awe, and stories about a treatment that may cure AIDS or a hockey goalie that plays even though they have brain cancer may both inspire this emotion.

We suggest that awe-inspiring content will be more viral. Awe is a social emotion (Keltner et al 2006) that encourages people to connect with others and spread the word. People who have had religious epiphanies, for example, seem to have a deep need to talk about them or proselytize (James 1902; Keltner and Haidt 2003), and other awe-inducing experiences may activate similar psychological needs. Awe-inducing experiences also encourage people to look beyond themselves and deepen connections to the broader social world (Shiota et al 2007), which may promote transmission.

The Current Research

Overall then, we examine the link between integral affect and the virality. Is positive or negative content more viral? Beyond valence, how are specific emotions linked to virality? For example, is anger-, anxiety-, or sadness-inducing content more or less viral?

To answer these questions, we investigate the virality of almost 7,000 articles from one of the world's most popular newspapers: *The New York Times*. Because the *Times* covers a wide range of topics (i.e., world news, sports, and travel) and boasts the most frequented website of any newspaper in the U.S. (Nielson, 2008), it is an ideal venue for examining what types of content are most frequently shared. *Times* articles are also shared with a wide range of recipients. When we asked a sample of 343 *Times* readers whom they had most recently shared an article with, responses indicated a mix of friends (42%), relatives (40%), colleagues (10%), and others (7%). The *Times* continually reports which 25 articles from its website have been emailed most frequently in the last 24 hours. We examine how the amount of emotion an article evokes, as well as the valence and specific nature of those emotions (e.g., anger vs. sadness) relates to whether it makes the *Times'* most emailed list.

Importantly, our analysis includes a variety of controls. First, in addition to suggesting that practical utility drives transmission, one could argue that more interesting articles are more likely to be shared. Regardless of how much emotion they evoke, some articles may just be more fascinating or entertaining, and this might lead people to share them (e.g., to self-enhance; Wojnicki and Godes 2008). Consequently, we control for both practical utility and interest to examine the link between emotion and virality above and beyond these factors. Second, in addition to quantifying various content characteristics (e.g., the extent to which articles provide practical utility or evoke various emotions), our analyses include a number control that are unrelated to characteristics of the content itself. Articles that appear on the front page of the physical paper or spend more time in prominent positions on the homepage may receive more attention and thus mechanically have a better chance of making the most emailed list. Consequently we control for these, and other potential external drivers of attention, to ensure that any relationships between content characteristics and virality are not the result of editorial decisions about what to feature or author fame. Measuring these aspects also allows us to provide at least a preliminary investigation into the role of placement versus content characteristics in shaping virality. While being heavily advertised, or in this case prominently featured, certainly increases the chance cultural items succeed, we examine whether content characteristics are of similar importance.

Data

We collected information about articles written for the *Times* that appeared on the paper's homepage (www.nytimes.com) between August 30th and Nov 30th 2008 (6,956 articles). Data was captured by a webcrawler that visited the *Times* homepage every 15 minutes during the period. It recorded information about every article on the homepage and each article on the most emailed list (updated every 15 minutes). The content of AP, Reuters, and Bloomberg articles, as well as blogs, is not stored by the *Times*, and so was not available for our analyses. Videos and images with no text were also not included. We captured each article's title, full text, author(s), topic area (e.g., opinion or sports), and two sentence summary created by the *Times*. We also captured each article's section, page, and publication date if it appeared in the print paper, as well as the dates, times, locations and durations of all appearances it made on the *Times'* homepage. Twenty percent of articles in our final data set earned a position on the most e-mailed list.

Article coding

We relied on human coders to quantify the extent to which each article evoked specific emotions (i.e., anxiety, anger, awe, disgust, sadness, or surprise) and contained practically useful information or evoked interest. Coders were blind to our hypotheses. They received the title and summary of each article, a web link to the article's full text, and detailed coding instructions (see Supplementary Materials). Given the overwhelming number of articles in our data set, we selected a random subsample for coding (N = 2,566). For each dimension (*Awe, Anger, Anxiety, Disgust, Sadness, Surprise, Practical utility,* and *Interest*), a separate group of three independent raters rated each article on a five point Likert scale based on the extent to which it was characterized by that particular aspect (1 = not at all, 5 = extremely). Raters were given feedback on their coding of a test set of articles until it was clear they understood the relevant construct. Inter-rater reliability was high on all dimensions (all α's > .70),[2] and scores were averaged across coders (see Table 2 for summary statistics) and standardized. All uncoded articles were assigned a score of zero on each dimension after standardization, and a dummy was included in regression analyses to control for uncoded stories (see Cohen and Cohen [1983] for a discussion of this imputation methodology). This allowed us to use the full set of articles collected to analyze the relationship between other content characteristics (that did not require manual coding) and virality. We also report our results relying only on the coded subset of articles to show that they are meaningfully unchanged.

Automated sentiment analysis was used to quantify the positivity (i.e., valence) and emotionality (i.e., affect-ladenness) of each article. These methods are well-established (Pang and Lee 2008; Pennebaker, Mehl, and Niederhoffer 2003) and increase coding ease and objectivity.[3] A computer program counted the number of positive and negative words in each article using a list of 7,630 words classified as positive or negative by human readers (Pennebaker, Booth, and Francis 2007). Positivity was quantified as the difference between the percentage of positive and negative words in an article. Emotionality was quantified as the percentage of words that were classified as either positive or negative. These variables were also standardized to ease interpretation of our regression results (see Table 3, following References, for correlations between standardized variables).

See Table 1 (following References) for articles that scored highly on the different dimensions.

Additional controls

External factors (separate from content characteristics) may affect an article's virality by functioning like advertising. Appearing earlier or in certain sections of the physical paper, spending more time in a prominent position on the homepage, being released when readership is greater, and being written by a famous author all likely generate attention for an article and increase its chances of making the most emailed list. Consequently, we control for these factors (see Table 4, following References).

Appearance in the physical paper. To characterize where an article appeared in the physical paper, we created dummy variables to control for the article's section (e.g., Section A). We also create indicator variables quantifying the page in a given section (e.g., A1) where an article appeared in print to control for the possibility that appearing earlier in some sections has a different effect than appearing earlier in others.

Appearance on the homepage. To characterize how much time an article spent in prominent positions on the homepage, we created variables that indicated where, when, and for how long every article was featured on the *Times* homepage. The homepage layout remained the same throughout the period of data collection. Articles could appear in several dozen positions on the homepage, so we aggregated positions into seven general regions based on locations that likely receive similar amounts of attention (Figure 1). Variables indicating the amount of time an article spent in each of these seven regions were included as controls after winsorization of the top 1% of outliers (to prevent extreme outliers from exerting undue influence on our results; see Tables A1 and A2 in the Appendix for summary statistics).

Release timing. To control for the possibility that articles released at different times of day receive different amounts of attention, we created controls for the time of day (6 am – 6 pm or 6 pm – 6 am EST) when an article first appeared online.

Author fame. We control for author fame to ensure that our results are not driven by the tastes of particularly popular writers whose stories may be particularly likely to be shared. To quantify author fame, we capture the number of Google hits returned by a search for each first author's full name (as of February 15, 2009). Due to its skew, we use the logarithm of this variable as a control in our analyses.

We also control for variables that might both might influence transmission and the likelihood that an article possesses certain characteristics (i.e., evokes anger).

Writing complexity. We control for how difficult a piece of writing is to read, using the *SMOG Complexity Index* (McLaughlin 1969). This widely used index variable essentially measures the grade-level appropriateness of the writing. Alternate complexity measures yield meaningfully unchanged results.

Author gender. Since male and female authors have different writing styles (Koppel, Argamon, and Shimoni 2002; Milkman, Carmona and Gleason, 2007), we control for the gender of an article's first author (male, female or unknown due to a missing byline). We classify gender using a first name mapping list from prior research (Morton, Zettelmeyer, and Silva-Risso 2003). For names that were classified as gender neutral or did not appear on this list, research assistants determined author gender by looking the authors up online.

Article length. We also control for an article's length in words. Longer articles may be more likely to go into enough detail to inspire awe or evoke anger but may simply be more viral because they contain more information.

Competition. Finally, we control for the competition a given article faced to make the most emailed list or "cohort effects". As would be expected from a daily newspaper, most articles released on a given day do not appear on the homepage for more than 24 hours, as they are replaced by the next day's lead stories. In addition, articles that make the most emailed list do so soon after they are released (95% do so within 24 hours of appearing on the homepage). Consequently, any competition among articles for attention or sharing essentially occurs within a daily cohort of content. Thus we include day dummy variables to control for competition to make the most emailed list.

See Table 4, following References, for a list of the independent variables included in our analyses.

Analysis strategy

Since 96% of articles that make the most emailed list do so only once (i.e., they do not leave the list and then re-appear later), we model making the list as a single event (see Appendix). To analyze the relationship between an article's content characteristics and the likelihood that it will make the *New York Times'* most e-mailed list, we use the following logistic regression specification:

$$(1)\qquad makes\_it_{at} = \frac{1}{1+\exp\left\{-\left(\begin{array}{l}\alpha_t + \beta_1 * z\text{-}emotionality_{at} + \beta_2 * z\text{-}positivity_{at} + \\ \beta_3 * z\text{-}awe_{at} + \beta_4 * z\text{-}anger_{at} + \beta_5 * z\text{-}anxiety_{at} + \\ \beta_6 * z\text{-}sadness_{at} + \beta_7 * z\text{-}disgust_{at} + \theta' * X_{at}\end{array}\right)\right\}}$$

where $makes\_it_{at}$ is a variable that takes on a value of one when an article *a,* released online on day *t,* earns a position on the most e-mailed list and zero otherwise, and α_t is an unobserved day-specific effect. Our primary predictor variables quantify the extent to which an article *a* published on day *t* was coded as positive, emotional, awe-inspiring, anger-inducing, anxiety-inducing, sadness-inducing, or surprising. X_{at} is a vector of the other control variables described above (see Table 4, following References).[4] We estimate the equation using a fixed effects logistic regression and cluster standard errors by day.

Results

Results suggest a strong link between evoking emotion and whether online content becomes viral (Table 5, following References). First, looking at the results of emotionality, we find that more affect laden content, regardless of whether it is positive or negative, is more likely to make the most emailed list. Looked at another way, when both the percentage of positive and negative words in an article were entered in as predictors (instead of emotionality and valence), both were positively linked to making the most emailed list.

Second, looking at the link between valence and sharing indicates that positive content is more viral than negative content. The comprehensiveness of our dataset is particularly useful here because it allows us to disentangle preferential transmission from mere base rates. There might be more positive than negative WOM, for example, but without knowing the full frequency of events, this might just be a result of the fact that positive events are more common (Rozin, Berman, and Royzman 2010) and thus there are more of them to talk about. Access to the full corpus of articles published by the *Times* over the analysis period allows us separate these possibilities. Taking into account the distribution of valence, our results indicate that more positive content is more viral.

Third, looking at specific emotions shows that the link between emotion and virality is more complex than just valence alone. While more awe-inspiring and more surprising content is more likely to make the most emailed list, and sadness-inducing content is less viral, some negative emotions are positively associated with virality. More anxiety- and anger-inducing content are both more likely to make the most emailed list. This suggests that transmission is more than just about sharing positive things and avoiding sharing negative ones. There was no significant relationship between disgust and virality.[5]

It is worth noting that these results persist even controlling for interest and practical utility (Table 5, Model 3, following References). Not surprisingly, more practically useful or interesting articles are more likely to make the *New York Times*' most emailed list, but even after controlling for these content characteristics, the links between emotion and virality remain significant.

These relationships are also robust to including a host of other controls (Table 5, Model 4, following References). Not surprisingly, being featured on the homepage for longer is positively associated with making the most emailed list, and time in more prominent positions on the page (e.g., top vs. bottom) is more strongly linked to virality. Even controlling for this type of "advertising", however, the relationships between emotional characteristics of content and virality persist and are of similar magnitude. The robustness of our results to the inclusion of such controls ensures that the high transmission rates of awe-inspiring stories, for example, is not simply driven by editors tending to feature awe-inspiring news, which could mechanically increase the virality of such content.[6] Longer articles, articles by more famous authors, and articles written by women are also more likely than others to make the most emailed list, but controlling for these factors does not meaningfully change the relationship between psychological characteristics of content and virality.

The results are also robust to controlling for an article's general topic (20 areas classified by the *Times* such as opinion, science, or health; Table 5, Model 5, following References). This indicates that our findings are not merely driven by certain areas (e.g., science or health) tending to both contain highly surprising or awe-inspiring articles, for example, and being particularly likely to make the most e-mailed list. Rather, this more conservative test of our hypothesis demonstrates that the observed relationships between integral affect and virality hold not only across topics but also within them. Even among opinion or health articles, for example, awe-inspiring articles and surprising articles are more viral.

Finally, our results remain meaningfully unchanged in terms of magnitude and statistical significance if we: (1) restrict our analyses to include only those 2,566 articles that were randomly selected for hand-coding (Table 5, Model 6, following References); (2) add squared and/or cubed terms quantifying how long an article spent in each of seven homepage regions; (3) add dummies indicating whether an article ever appeared in a given homepage region; (4) split the region variables into time spent in each region during the day (6 am – 6 pm EST) and night (6 am – 6 pm EST); (5) control for the day of the week when an article was published in the physical paper (instead of online); (6) winsorize the top and bottom 1% of outliers for each control variable in our regression; (7) remove day fixed effects from our analyses; or (8) replace day fixed effects with controls for the average rating of practical utility, awe, anger, anxiety, sadness, surprise, positivity and emotionality in the day's published news stories. These checks indicate that the observed results are not an artifact of the particular regression specifications we rely on in our primary analyses. Results are also robust to alternate ways of quantifying emotion (e.g., using textual analysis to quantify the extent to which articles inspire awe or evoke anxiety).

More broadly, though our results suggest that external drivers of attention (e.g., being prominently featured) shape what becomes viral, they also indicate that content characteristics are of similar importance. For instance, the most powerful predictor of virality in our model is how much anger an article evokes: parameter estimates imply that a one standard deviation increase in an article's anger rating increases the odds that an article make the most e-mailed list by a factor of 1.5 (Table 5, Model 4, following References). This increase is equivalent to the effect of spending an additional 2.9 hours as the lead story on the *Times* website, which is nearly four times the average number of hours the average article spends in that position. Similarly, a one standard deviation increase in evoking awe (our second most powerful content predictor) increases the odds that an article will make the most e-mailed list by a factor of 1.4 (Table 5, Model 4). Even our weakest content predictor – positivity – meaningfully moves the needle. An increase of one standard deviation in positivity has an equivalent impact on an article's odds of making the most emailed list to spending 1.2 hours as the *Times'* lead story. See Figure 2 for an illustration of the magnitude of these detected effects.

General Discussion

The emergence social media (e.g., Facebook and Twitter) has boosted interest in word-of-mouth and viral marketing. But while it is clear that consumers often share online content, and that social transmission influences product adoption and sales, less is known why certain content becomes viral. Further, though diffusion research has examined how certain individuals (e.g., social hubs or influentials), or social network structures might influence social transmission, there has been less attention to how characteristics of content that spread across social ties might shape collective outcomes.

This paper takes an interdisciplinary approach to studying virality. Building on psychological theory, we conducted a broad analysis of social transmission that provides insight into the relationship between emotion and virality. First, our findings inform the ongoing debate about whether people tend to share positive or negative content. While common wisdom suggest that people tend to pass along negative news more than positive news, our results indicate that positive news is actually more viral. Further, by examining the full corpus of *New York Times* content (i.e., all articles available), we can say that positive content is more likely to be highly shared even controlling for how frequently it occurs.

That said, our results suggest that transmission is more complex than just valence. While more awe inspiring or surprising content is more viral, and sad content is less viral, content that evokes more anger or anxiety is more highly shared. Not surprisingly, more practically useful or interesting content is also more viral, but our results hold even controlling for these factors.

Demonstrating these relationships outside of the laboratory, and across a large and diverse body of content, underscores their generality. Further, the naturalistic setting allows us to test the relative importance of content characteristics and external drivers of attention in shaping virality. While being featured prominently or being written by a famous author increases the likelihood that articles are highly shared, our results suggest that content characteristics are of similar importance.

Theoretical implications

This research links psychological and sociological approaches to studying diffusion. While research has modeled product adoption (Bass 1969) or looked at how social networks shape the diffusion and sales (Stephen and Toubia 2010; Van den Bulte and Wuyts 2007), macro-level collective outcomes such as what becomes viral depend on micro-level individual decisions about what to share. Consequently, when trying to understand collective outcomes, it is important to consider the underlying individual-level psychological processes. Along these lines, this work suggests that the emotion content evokes in individuals (micro-level) helps determine which cultural items succeed in the marketplace of ideas (macro-level).

While our study focuses on collective outcomes, it also sheds some light on underlying drivers of social transmission. Awe, anger, and anxiety are all linked to increased virality, and each of these emotions is characterized by activation or arousal (Barrett and Russell 1998). In contrast, sadness is characterized by deactivation and is associated with decreased virality. This pattern of results suggests that arousal may shape what people share. Emotions characterized by activation may excite people, signal that activity is desired, and lead consumers to spread the word. Emotions characterized by deactivation may calm people, signal that inactivity is desired, and thus reduce social transmission. Future research into this possibility seems warranted.

Our findings also suggest that social transmission is about more than just value exchange or self-enhancement. Consistent with the notion that people share to entertain, surprising and interesting content is highly viral. Similarly, consistent with the notion that people share to inform others, or boost their mood, practically useful and positive content is more viral. These effects are all consistent with the idea that people may share valuable content to help others, generate reciprocity, or self-enhance (e.g., show they know entertaining or useful things). Even controlling for these effects, however, affect laden content in general, and anxiety- or anger-evoking content in particular, is more likely to make the most emailed list. Such content does not clearly produce immediate economic value in the traditional sense, or even necessarily reflect favorably on the self. Though more work is necessary to determine why such content is viral, one possibility may lie in social connection. Sharing affectively rich content can reinforce shared views and deepen social bonds (Heath, et al 2001; Peters and Kashima 2007), and people may share even affectively negative content to deepen connections with others.

Limitations and directions for future research

This research is not without limitations. The field setting allowed us to examine real transmission of a broad set of content by a diverse population, and to examine collective outcomes, but as is the case with most studies of archival data, we are limited in our ability to draw causal inferences from our results or investigate underlying psychological mechanisms.[7] Future experimental work, however, might more directly examine the underlying mechanisms behind the observed relationships. One potential avenue would be to investigate why people share content that induces particular emotions. As noted earlier, for example, people may share awe-inspiring content because it generates the need to proselytize or creates a desire to connect with others. Alternatively, given that awe appears to be a self-diminishing emotion (Shiota, et al 2007), people might share such experiences to bolster their own sense of self. Similarly, people might share anxiety-inducing content to calm themselves or reduce uncertainty. Directly manipulating emotional states and examining what people share, or examining the consequences of sharing various types of emotional content would shed light on the mechanisms behind the social transmission of emotional content.

It might also be interesting to examine how audience size moderates what people share. People often email online content to a particular friend or two, but in other cases they may broadcast content to a much larger audience (e.g., tweeting, blogging, or posting it on their Facebook wall). Though the former (i.e., narrowcasting) can involve niche information (i.e., sending an article about rowing technique to a friend who likes crew), broadcasting likely requires posting things that have broader appeal. One could also imagine that while narrowcasting is recipient focused (i.e., what they would enjoy), broadcasting is self-focused (i.e., what someone wants to say about themselves or show others). Consequently, self-presentation motives, identity signaling, or affiliation goals should play a stronger role in shaping what people share with larger audiences.

Though our data does not allow us to speak to this issue in great detail, we were able to investigate the link between article characteristics and blogging. Half-way into data collection, we built a supplementary web-crawler that captures the *Times'* list of the 25 articles that appeared in the most blogs over the previous 24 hours. Analysis suggests that similar factors drive both virality and blogging: more emotional, positive, interesting, and anger-inducing, and less sadness-inducing stories are more likely to make the most blogged list. Interestingly, the effect of practical utility reverses – though a practically useful story is more likely to make the most emailed list, practically useful content is marginally *less* likely to be blogged about. This may be due in part to the nature of blogs as commentary. While movie reviews, technology perspectives, and recipes all contain useful information, they are already commentary, and thus there may not be much added value from a blogger contributing his or her spin on the issue.

Future research might also examine how the effects observed here are moderated by situational or relationship factors. Given that the weather can affect people's moods, for example, it may affect the type of content that is shared. People might be more likely to share positive stories on overcast days, for example, to make others feel happier. Alternatively, people might be more likely to share more negative stories on overcast days due to mood congruence. More broadly, other cues in the environment might change what people share by making certain topics more accessible (Berger and Fitzsimons 2008; Nedungadi 1990). If the Yankees win the World Series, for example, that should be front page news, but as a result, people may also be more likely to share any sports story more generally because that topic is primed.

Marketing implications

These findings have a number of important marketing implications. First, online content providers may want to pay greater attention to the emotions their content evokes. Doing so should help companies maximize revenue for placing advertisements or pricing access to content (e.g., potentially charging more for content that is likely to be highly shared). It might also be useful to highlight, or design more content that evokes the emotional aspects noted here, as such content is likely to be shared (which increases page views).

More generally, our findings shed light on how to design successful viral marketing campaigns and craft contagious content. Given that affect-laden and awe-inspiring content is more likely to be shared, for example, campaigns that strive to evoke these emotions may be more successful. Similar points apply to managing consumer sentiment online. Social media (e.g., discussion forums or Twitter) allow consumers help co-create brand meaning (Kuksov and Shachar 2010), but they can also facilitate negative social movements, distributing negative stories and generating consumer backlashes. Moms offended by a Motrin ad campaign, for example, banded together and began posting negative YouTube videos and tweets (Petrecca 2008). Our findings, however, suggest that not all negative emotions are the same when it comes to sharing and that some may be more likely to incite transmission. Brand transgressions that evoke anxiety or anger, for example, may be more likely to be shared than those that evoke sadness, which suggest that companies may want to take a more active role in managing situations that evoke these emotions.

In conclusion, this research is the first broad analysis of how online content characteristics relate to virality. Our results suggest that in addition to practical utility, emotion plays an important role in what gets shared, though the relationships are more complex than mere valence alone. More generally, while much more work remains, this work highlights the value of considering how psychological processes may shape collective outcomes such as what becomes viral.

Appendix

Modeling Approach

We used a logistic regression model because of the nature of our question and the available data. While more complex panel-type models are appropriate when there is time variation in at least one independent variable and the outcome, we do not have period-by-period variation in the dependent variable. Rather than having the number of emails sent in each period, we only have a dummy variable that switches from 0 (not on the most emailed list) to 1 (on the most emailed list) at some point due to events that happened not primarily in the same period but several periods earlier (such as advertising in previous periods). Further, our interest is not in when an article makes the list but whether it ever does so. Finally, while one could imagine that when an article is featured might impact when it makes the list, such an analysis is far from straightforward. The effects are likely to be delayed (where an article is displayed in a given time period is extremely unlikely to have any effect on whether the article makes the most emailed list during that period), but it is difficult to predict a priori what the lag between being featured prominently and making the list would be. Thus, the only way to run an appropriate panel model would be to include the full lag structure on all of our time varying variables (times spent in various positions on the home page). Since we have no priors on the appropriate lag structure, the full lag structure would be the only appropriate solution. So, for instance, imagine there are two slots on the homepage (we actually have eight) and that they are position A and position B. Our model would then need to be something like:

> *Being on the list in period t =* β_1**(being in position A in period t) +* β_2**(being in position A in period t – 1) +* β_3**(being in position A in period t – 2) + … +* β_N**(being in position A in period t – N) +* β_{N+1}**(being in position B in period t) +* β_{N+2}**(being in position B in period t – 1) +* β_{N+3}**(being in position B in period t – 2) + … +* β_{2N}**(being in position B in period t – N) +*$\boldsymbol{\beta}$*(a vector of our other time-invariant predictors)*

If we estimated this model, we would actually end up with an equivalent model to our current logistic regression specification where we have summed all of the different periods for each position. The two are equivalent models unless we include interactions on the lag terms, and it is unclear what interactions it would make sense to include. In addition, there are considerable losses in efficiency from this panel specification when compared with our current model. Thus, we rely on a simple logistic regression model to analyze our data set.

Coding Instructions

Anger. Articles vary in how angry they make most readers feel. Certain articles might make people really angry while others do not make them angry at all. Here is a definition of anger http://en.wikipedia.org/wiki/Anger. Please code the articles based on how much anger they evoke.

Anxiety. Articles vary in how much anxiety they would evoke in most readers. Certain articles might make people really anxious while others do not make them anxious at all. Here is a definition of anxiety http://en.wikipedia.org/wiki/Anxiety. Please code the articles based on how much anxiety they evoke.

Awe. Articles vary in how much they inspire awe. Awe is the emotion of self-transcendence, a feeling of admiration and elevation in the face of something greater than the self. It involves the opening or broadening of the mind and an experience of wow that makes you stop and think. Seeing the Grand Canyon, standing in front of a beautiful piece of art, hearing a grand theory, or listening to a beautiful symphony may all inspire awe. So may the revelation of something profound and important in something you may have once seen as ordinary or routine or seeing a causal connection between important things and seemingly remote causes.

Disgust. Articles vary in how much disgust they evoke. Certain articles might make people really disgusted while others do not make them disgusted at all. Here is a definition of disgust http://en.wikipedia.org/wiki/Disgust. Please code the articles based on how much disgust they evoke.

Sadness. Articles vary in how much sadness they evoke. Certain articles might make people really sad while others do not make them sad at all. Here is a definition of sadness http://en.wikipedia.org/wiki/Sadness. Please code the articles based on how much sadness they evoke.

Surprise. *Articles* vary in how much surprise they evoke. Certain articles might make people really surprised while others do not make them surprised at all. Here is a definition of surprise http://en.wikipedia.org/wiki/Surprise_(emotion). Please code the articles based on how much surprise they evoke.

Practical Utility. Articles vary in how much practical utility they have. Some contain useful information that leads the reader to modify their behavior. For example, reading an article suggesting certain vegetables are good for you might cause a reader to eat more of those vegetables. Similarly, an article talking about a new Personal Digital Assistant may influence what the reader buys. Please code the articles based on how much practical utility they provide.

Interest. Articles vary in how much interest they evoke. Certain articles are really interesting while others are not interesting at all. Please code the articles based on how much interest they evoke.

TABLE A1
HOMEPAGE LOCATION ARTICLE SUMMARY STATISTICS

	% of Articles That Ever Occupy This Location	For Articles that Ever Occupy Location:		
		% That Make List	Mean Hrs	Hrs Std. Dev.
Top Feature	28%	33%	2.61	2.94
Near Top Feature	32%	31%	5.05	5.11
Right Column	22%	31%	3.85	5.11
Middle Feature Bar	25%	32%	11.65	11.63
Bulleted Sub-Feature	29%	26%	3.14	3.91
More News	31%	24%	3.69	4.18
Bottom List	88%	20%	23.31	28.40

Note: The average article in our data set appeared somewhere on the *Times* homepage for a total of 29 hours (standard deviation = 30 hours)

TABLE A2
PHYSICAL NEWSPAPER ARTICLE LOCATION SUMMARY STATISTICS

	% of Articles That Ever Occupy This Location	For Articles that Ever Occupy This Location:		
		% That Make List	Mean Pg #	Mean Pg # for Articles that Make List
Section A	39%	25%	15.84	10.64
Section B	15%	10%	6.59	5.76
Section C	10%	16%	4.12	5.38
Section D	7%	17%	3.05	2.27
Section E	4%	22%	4.78	7.62
Section F	2%	42%	3.28	3.43
Other Section	13%	24%	9.59	14.87
Never in Paper	10%	11%	-	-

Notes

1. We focus on awe in particular because preliminary analysis of the data suggested that science articles or other topics that might evoke awe frequently appeared on the most emailed list.
2. There is certainly some heterogeneity in what people find surprising, for example, or awe-inspiring. That said, the fact that multiple raters coded articles similarly suggests that content tends to evoke similar emotions across people.
3. Automated ratings were significantly correlated with manual coders ratings of a subset of articles
4. This includes: practical utility and interest scores, indicators of the number of hours an article spent in each of seven online locations, a dummy indicating whether the article first appeared online at night (6 pm – 6 am EST), a dummy indicating which section in the physical paper the article appeared in, an indicator of the page number an article appeared in for each of the given physical paper sections, the first author's fame, the article's complexity score, dummies indicating whether the first author is female or of unknown gender, wordcount, and a dummy indicating whether the article in question was one of the 3,000 coded manually on the characteristics: *practicality*, *surprise*, and *awe-inspiring*.
5. This may be due in part to the context examined. Most online content, and news articles in particular, tend not to evoke large amounts of disgust.
6. Further, regressing the various content characteristics on being featured suggest that topical section (e.g., national news vs. sports), rather than integral affect, determines where articles are featured. Results show that section, or even more general topical areas (e.g., opinion), are strongly related to whether and where articles are featured on the homepage, while emotional characteristics are not.
7. It is worth noting that the observed relationships between psychological characteristics and virality are at least consistent with the responses of 343 *New York Times* readers who were asked to list the article they had most recently shared and why they shared it. For example, numerous explanations highlighted that sharing was driven by practical utility (e.g., "My sister lives in Santa Fe and is an artist there; the article is 'The Art of Santa Fe'"), anger (e.g., "My daughter is fighting with her insurance to get a breast lump removed."), awe (e.g., "Because I admire the work Dr. Pepperberg has done on animal behavior and learning, and want other people to learn about animal behavior so they have a better understanding of themselves as human animals, and a better understanding of how the differences between animals and humans are of degree, not essence"), surprise (e.g., "It was a fun article about a sport that I had never heard of that is played in a state that I have lived in for eight years."), positivity (e.g., "I wanted to share with my brother the good news of the Obama resurgence."), and anxiety (e.g., "To warn her about a health risk"). While these examples are merely illustrative, they suggest at least some consistency between micro-level motives and our macro-level quantitative analysis.

REFERENCES

Allport, Gordon W. and Leo Postman (1947), *The Psychology of Rumor*, New York: Henry Holt.

Allsop, Dee T., Bryce R. Bassett, and James A. Hoskins (2007), "Word-of-Mouth Research: Principles and Applications," *Journal of Advertising Research*, 47, 388-411.

Asch, Solomon E. (1956), "Studies of Independence and Conformity: A Minority of One Against a Unanimous Majority," *Psychological Monographs,* 70 (416).

Barrett, L. F., & Russell, J. A. (1998), "Independence and Bipolarity in the Structure of Current Affect," *Journal of Personality and Social Psychology,* 74, 967-984.

Baumeister, R.F., Bratslavsky, E., Finkenauer, C., & Vohs, K.D. (2001), "Bad is Stronger than Good," *Review of General Psychology*, 5, 323-370.

Bass, Frank (1969). "A new product growth model for consumer durables". *Management Science* 15 (5): p215–227

Berger, Jonah and Gráinne M. Fitzsimons (2008), "Dogs on the Street, Pumas on Your Feet: How Cues in the Environment Influence Product Evaluation and Choice," *Journal of Marketing Research*, 45(1), 1-14.

Brunvand, J.H. (1981), *The Vanishing Hitchhiker: American Urban Legends and their Meanings*, New York: W.W. Norton & Company.

Cashmore, Pete (2009), "YouTube: Why Do We Watch?". http://www.cnn.com/2009/TECH/12/17/cashmore.youtube/index.html

Cheema, Amar and Andrew M. Kaikati (2010), "The Effect of Need for Uniqueness on Word of Mouth," *Journal of Marketing Research*, 47, 3 (June), 553-63.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43, 345-354.

Cohen, Jacob and Patricia Cohen (1983), *Applied Multiple Regression/ Correlation Analysis for the Behavioral Sciences*: *(2nd Ed,*. Hillsdale, NJ: Erlbaum.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1982). *What emotion categories or dimensions can observers judge from facial behavior? In P. Ekman (Ed.)*, *Emotion in the human face* (pp. 39-55), New York: Cambridge University Press.

Fehr, Ernst, Georg Kirchsteiger and Arno Riedl (1998), "Gift Exchange and Reciprocity in Competitive Experimental Markets," *European Economic Review*, 42, 1, 1-34.

Festinger, Leon, Henry W. Riecken, and Stanley Schachter (1956), *When Prophecy Fails*, New York: Harper and Row.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23, 545–60.

Godes, David, Dina Mayzlin, Yubo Chen, Sanjiv Das, Chrysanthos Dellarocas, Bruce Pfeiffer, Barak Libai, Subrata Sen, Mengze Shi and Peeter Verlegh (2005), "The Firm's Management of Social Interactions," *Marketing Letters*, 16 (3/4), 415-428.

Godes, David and Dina Mayzlin (2009), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, 28, 721-739.

Goldenberg, Jacob, Barak Libai, and Eitan Muller (2009),"The chilling effects of network externalities," *International Journal of Research in Marketing, forthcoming*.

Harris, Jacob (2010), "How Often Is The Times Tweeted," *New York Times Open Blog*, April 15, http://open.blogs.nytimes.com/2010/04/15/how-often-is-the-times-tweeted/

Heath, Chip, Chris Bell, and Emily Sternberg (2001), "Emotional Selection in Memes: The Case of Urban Legends," *Journal of Personality and Social Psychology,* 81, 1028-1041.

Homans, George C. (1958), "Social Behavior as Exchange," *American Journal of Sociology*, 63 (6), 597-606.
James, William (1902), *The Varieties of Religious Experience.* New York: Touchstone.
Kashima, Yoshihasa (2008), "A Social Psychology of Cultural Dynamics: Examining How Cultures are Formed, Maintained, and Transformed," *Social and Personality Psychology Compass,* 2, 107-120.
Katz, Elihu, Paul Felix Lazarsfeld, and Elmo Roper (1955), *Personal Influence: The Part Played by People in the Flow of Mass Communication*. Glencoe, IL: Free Press.
Keltner, Dacher and Jon Haidt (2003), "Approaching Awe: A Moral, Spiritual, and Aesthetic Emotion," *Cognition and Emotion,* 17, 297-314.
Keltner, Dacher and Jennifer S. Lerner (2010), "Emotion," To appear in *The Handbook of Social Psychology (5th edition) (D. Gilbert, S. Fiske, and G. Lindsey, Eds.).* New York: McGraw Hill.
Knapp, R. H. (1944). "A Psychology of Rumor," *Public Opinion Quarterly*, 8, 23-37.
Koppel, Moshe, Shlomo Argamon, and Anat Rachel Shimoni (2002), "Automatically Categorizing Written Texts by Author Gender," *Literary and Linguistic Computing,* 17, 401-412.
Kuksov, Dmitri and Ron Shachar (2010), "User Generated Content and Image Advertising," Working Paper.
Lazarus, Richard S. (1991), *Emotion and Adaptation,* London: Oxford University Press.
Lerner, J. S., & Keltner, D. (2001). "Fear, Anger, and Risk," *Journal of Personality & Social Psychology*, 81(1), 146-159.
Lerner, J., Small, D., Lowenstein, G. (2004), "Heart Strings and Purse Strings: Carryover Effects of Emotions on Economic Decisions", *Psychological Science*, Vol. 15, 337-41.
Markus, Hazel Rose and Shinobu Kitayama (1991), "Culture and the Self: Implications for Cognition, Emotion, and Motivation," *Psychological Review*, 98 (2), 224-253.
McLaughlin, G. Harry (1969), "SMOG Grading: A New Readability Formula," *Journal of Reading,* 12, 639-646.
Milkman, Katherine L., Rene Carmona, and William Gleason (2007), "A Statistical Analysis of Editorial Influence and Author-Character Similarities in 1990s New Yorker Fiction," *Journal of Literary and Linguistic Computing*, 22, 305-328.
Morton, Fiona Scott, Florian Zettelmeyer, and Jorge Silva-Risso (2003), "Consumer Information and Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities?" *Quantitative Marketing and Economics,* 1, 65-92.
Nedungadi, Prakash (1990), "Recall and Consumer Consideration Sets: Influencing Choice Without Altering Brand Evaluations," *Journal of Consumer Research*, 263–76.
Nielson (2008). *NetRatings NetView*.
Pang, Bo and Lillian Lee (2008), "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, 2, 1-135.
Pennebaker, James W., Roger J. Booth, and Martha E. Francis (2007), *LIWC2007: Linguistic Inquiry and Word Count,* Austin, Texas: liwc.net
Pennebaker, J. W., Mehl, M. R., Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54,* 547-577.
Peters, Kim and Yoshihasa Kashima (2007), "From Social Talk to Social Action: Shaping the Social Triad with Emotion Sharing," *Journal of Personality and Social Psychology*, 93, 780-797.

Petrecca, Laura (2008), "Offended Moms Get Tweet Revenge over Motrin Ads," *USAToday.com*, November 19.
Rime, Bernard, Batja Mesquita, Pierre Philippot, and Stefano Boca (1991), "Beyond the Emotional Event: Six Studies on the Social Sharing of Emotion," *Cognition and Emotion*, 5 (September-November), 435-465.
Rosnow, Ralph L. (1980), "Psychology of Rumor Reconsidered," *Psychological Bulletin* 87:578-591
Rosnow, Ralph L. and Gary A. Fine (1976), *Rumor and Gossip: The Social Psychology of Hearsay*, New York: Elsevier.
Rozin, P., Berman, L., & Royzman, E.B. (2010), "Biases in use of positive and negative words across twenty natural languages." *Cognition and Emotion*, 24, 536-548.
Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts (2006), "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market," *Science*, 311, 854-856
Sauter, Disa A., Frank Eisner, Paul Ekman, and Sophie K. Scott. 2009. "Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations," *Proceedings of the National Academy of Sciences*, 107(6), 2408-2412.
Schaller, Mark and Christian S. Crandall (2004), *The Psychological Foundations of Culture,* Mahwah, NJ: Lawrence Erlbaum Associates.
Shiota, Michelle, Dacher Keltner, and Amanda Mossman (2007), "The Nature of Awe: Elicitors, Appraisals, and Effects on Self-Concept," *Cognition and Emotion,* 21, 944-963.
Smith, C. A., & Ellsworth, P. C. (1985), "Patterns of Cognitive Appraisal in Emotion," *Journal of Personality and Social Psychology*, 48, 813-838.
Stephen, Andrew T. and Olivier Toubia (2010), "Deriving Value from Social Commerce Networks," *Journal of Marketing Research*, 47 (2), 215-228.
Stephen, Andrew T. and Donald R. Lehmann (2010), "Why Do People Spread Word-of-Mouth? Effects of Recipient and Relationship Characteristics on Transmission Behaviors," INSEAD working paper.
Stephen, Andrew T., Dover, Yaniv and Goldenberg, Jacob (2010), "A Comparison of the Effects of Transmitter Activity and Connectivity on the Diffusion of Information Over Online Social Networks," INSEAD Working Paper.
Van den Bulte, Christophe and Stefan Wuyts (2007), *Social Networks and Marketing,* Cambridge, MA: Marketing Science Institute.
Wojnicki, Andrea C. and Dave Godes (2008), "Word-of-Mouth as Self-Enhancement," University of Toronto Working Paper.

FIGURE 1
HOMEPAGE LOCATION CLASSIFICATIONS. PORTIONS WITH "X'S" THROUGH THEM ALWAYS FEATURED AP AND REUTERS NEWS STORIES, VIDEOS, BLOGS, OR ADVERTISEMENTS RATHER THAN ARTICLES BY *TIMES* REPORTERS

FIGURE 2
PERCENT CHANGE IN FITTED PROBABILITY OF MAKING THE LIST FOR A 1 STANDARD DEVIATION INCREASE ABOVE THE MEAN IN AN ARTICLE CHARACTERISTIC

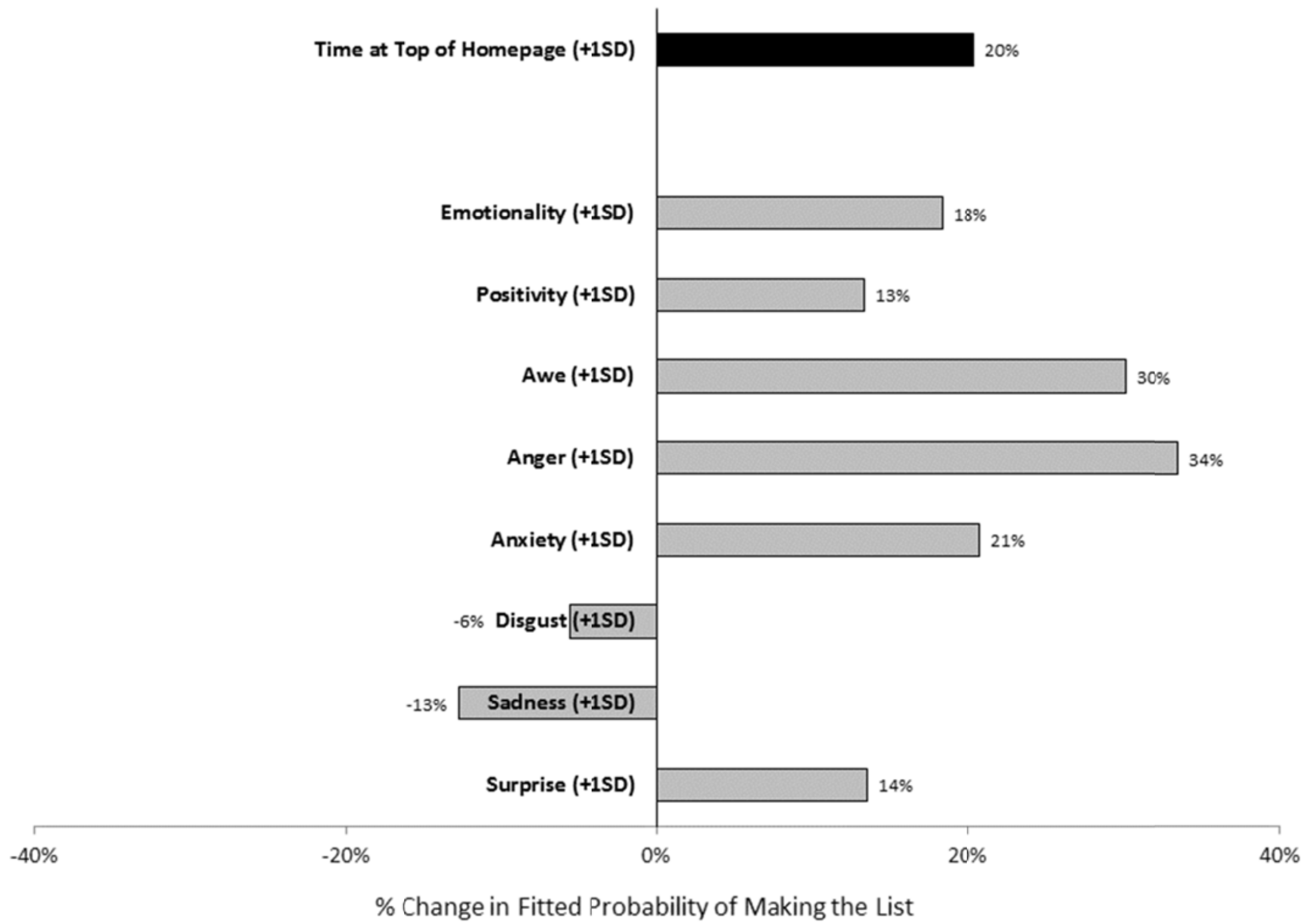

TABLE 1

	Primary Predictors
Emotionality	*High Scoring:* • "Redefining Depression as Mere Sadness" • "When All Else Fails, Blaming the Patient Often Comes Next"
Positivity	*High Scoring:* • "Wide-Eyed New Arrivals Falling in Love With the City" • "Tony Award for Philanthropy" *Low Scoring:* • "Web Rumors Tied to Korean Actress's Suicide" • "Germany: Baby Polar Bear's Feeder Dies"
Awe	*High Scoring:* • "Rare Treatment Is Reported to Cure AIDS Patient" • "The Promise and Power of RNA"
Anger	*High Scoring:* • "What Red Ink? Wall Street Paid Hefty Bonuses" • "Loan Titans Paid McCain Adviser Nearly $2 Million"
Anxiety	*High Scoring:* • "For Stocks, Worst Single-Day Drop in Two Decades" • "Home Prices Seem Far From Bottom"
Disgust	*High Scoring:* • "Dogfighting Ring Is Broken Up in Texas" • "Brooklyn Woman Is Questioned in Death of Daughter, 11"
Sadness	*High Scoring:* • "Maimed on 9/11, Trying to Be Whole Again" • "Obama Pays Tribute to His Grandmother After She Dies"
Surprise	*High Scoring:* • "Passion for Food Adjusts to Fit Passion for Running" (a story about a restaurateur who runs marathons) • "Pecking, but No Order, on Streets of East Harlem" (a story about chickens in Harlem)
	Control Variables
Practical Utility	*High Scoring:* • "Voter Resources" • "It Comes in Beige or Black, but You Make It Green" (a story about being environmentally friendly when disposing of old computers)
Interest	*High Scoring:* • "Love, Sex and the Changing Landscape of Infidelity" • "Teams Prepare for the Courtship of LeBron James"

TABLE 2
PREDICTOR VARIABLE SUMMARY STATISTICS

		Mean	Std. Dev.
Primary Predictor Variables	**Emotionality***	7.43%	1.92%
	Positivity*	0.98%	1.84%
	Awe*	1.81	0.71
	Anger*	1.47	0.51
	Anxiety*	1.55	0.64
	Disgust*	1.27	0.51
	Sadness*	1.31	0.41
	Surprise*	2.25	0.87
Other Control Variables	**Practical Utility***	1.66	1.01
	Interest*	2.71	0.85
	Wordcount	1,021.35	668.94
	Complexity*	11.08	1.54
	Author Fame	9.13	2.54
	Author Female	0.29	0.45
	Author Male	0.66	0.48

*Note that these summary statistics pertain to the variable in question prior to standardization.

TABLE 3
CORRELATIONS BETWEEN PREDICTOR VARIABLES

	Emotionality	Positivity	Awe	Anger	Anxiety	Disgust	Sadness	Surprise	Practical Utility	Interest	Word Count x 10^{-3}	Complexity	Author Fame	Author Female	Missing	Top Feature	Top Feature	Right Column	Sub-Feature	More News	Feature Bar
Emotionality	1.00																				
Positivity	0.04*	1.00																			
Awe	-0.02	0.02	1.00																		
Anger	0.04*	-0.16*	-0.21*	1.00																	
Anxiety	0.03*	-0.18*	-0.11*	0.50*	1.00																
Disgust	-0.01	-0.15*	-0.06*	0.52*	0.26*	1.00															
Sadness	0.00	-0.18*	0.08*	0.42*	0.45*	0.28*	1.00														
Surprise	-0.10*	-0.04*	0.24*	-0.01	0.00	0.11*	0.05*	1.00													
Practical Utility	0.06*	0.04*	-0.11*	-0.12*	0.07*	-0.14*	-0.05*	-0.05*	1.00												
Interest	0.054*	0.07*	0.26*	-0.13*	-0.24*	0.01	-0.19*	0.18*	-0.06*	1.00											
Word Count x 10^{-3}	0.06*	0.05*	0.04*	0.02	0.00	-0.02*	0.00	0.02*	-0.01	0.06*	1.00										
Complexity	0.05*	-0.05*	-0.04*	0.10*	0.13*	0.06*	0.05*	0.04*	0.01	-0.11*	-0.06*	1.00									
Author Fame	-0.09*	-0.03*	0.06*	0.01	0.03*	0.02	0.01	0.02	-0.02	0.00	0.01	0.01	1.00								
Author Female	-0.07*	0.06*	0.01	-0.03*	0.00	0.02	0.00	0.07*	0.05*	-0.01	0.00	-0.02*	0.00	1.00							
Missing	0.21*	0.03*	-0.06*	0.03*	-0.02	0.00	0.00	-0.09*	0.01	0.02	-0.01	0.02*	-0.71*	-0.15*	1.00						
Top Feature	0.01	-0.02	-0.03*	0.06*	0.06*	0.00	0.05*	-0.02*	0.02	-0.03*	0.28*	0.01	0.00	-0.02	0.01	1.00					
Near Top Feature	-0.01	-0.06*	-0.02	0.15*	0.07*	0.07*	0.07*	0.01	-0.03*	-0.05*	0.27*	0.06*	0.06*	-0.01	-0.05*	0.27*	1.00				
Right Column	0.16*	0.05*	0.04*	0.00	-0.02	0.02	-0.02	-0.02*	0.05*	0.06*	0.05*	-0.01	-0.03*	-0.02	0.16*	0.02	-0.04*	1.00			
Bulleted Sub-Feature	0.00	-0.02	-0.05*	0.09*	0.08*	0.04*	0.06*	-0.04*	0.04*	-0.05*	0.07*	0.03*	0.03*	0.01	-0.04*	0.12*	0.12*	-0.03*	1.00		
More News	-0.08*	-0.11*	-0.01	0.07*	0.06*	0.07*	0.06*	0.07*	-0.08*	-0.04*	-0.02	0.09*	0.05*	-0.01	-0.07*	0.01	0.10*	-0.06*	-0.05*	1.00	
Middle Feature Bar	0.11*	0.10*	0.06*	-0.06*	-0.06*	-0.06*	-0.05*	0.04*	0.00	0.10*	0.16*	-0.06*	-0.13*	0.00	0.13*	0.02	-0.05*	0.07*	-0.04*	-0.08*	1.00
Bottom List	0.03*	0.15*	0.07*	-0.11*	-0.09*	-0.08*	-0.06*	0.04*	0.06*	0.09*	0.29*	-0.04*	-0.06*	0.05*	0.00	0.04*	-0.05*	0.10*	0.00	-0.09*	0.13*

*Significant at 5% level.

TABLE 4
PREDICTOR VARIABLES

Variable	**Where it Came from**
Main Independent Variables	
Emotionality	Coded through textual analysis (LIWC)
Positivity	Coded through textual analysis (LIWC)
Awe	Coded by hand
Anger	Coded by hand
Anxiety	Coded by hand
Disgust	Coded by hand
Sadness	Coded by hand
Surprise	Coded by hand
Practical Utility	Coded by hand
Interest	Coded by hand
Control Variables	
Word Count	Coded through textual analysis (LIWC)
Author Fame	Log of # of hits returned by Google search of author's name
Writing Complexity	SMOG Complexity Index
Author Gender	List mapping names to genders (Morton & Zettelmeyer '03)
Author Byline Missing	Captured by webcrawler
Article Section Dummies	Captured by webcrawler
Hours Spent in Different Places on the Homepage	Captured by webcrawler
Section of the Physical Paper (e.g., A)	Captured by webcrawler
Page in Section in the Physical Paper (e.g., A1)	Captured by webcrawler
Time of Day the Article Appeared	Captured by webcrawler
Day the Article Appeared	Captured by webcrawler
Category of the Article (e.g., sports)	Captured by wecbrawler

TABLE 5
ARTICLE'S LIKELIHOOD OF MAKING THE *NEW YORK TIMES'* MOST E-MAILED LIST BASED ON PSYCHOLOGICAL CHARACTERISTICS OF ITS CONTENT, AS WELL AS VARIOUS CONTROL VARIABLES

		(1)	(2)	(3)	(4)	(5)	(6)
Emotion Predictors	**Emotionality**	0.27***	0.26***	0.25***	0.22***	0.09*	0.29***
		(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.06)
	Positivity	0.11***	0.16***	0.16***	0.16***	0.14***	0.23***
		(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)
Specific Emotions	**Awe**	-	0.46***	0.39***	0.34***	0.30***	0.36***
		-	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
	Anger	-	0.45***	0.52***	0.38***	0.29**	0.37***
		-	(0.07)	(0.07)	(0.09)	(0.10)	(0.10)
	Anxiety	-	0.20***	0.20***	0.24***	0.21***	0.27***
		-	(0.05)	(0.05)	(0.07)	(0.07)	(0.07)
	Disgust	-	-0.04	-0.07	-0.07	-0.03	-0.04
		-	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)
	Sadness	-	-0.19***	-0.15***	-0.17*	-0.12^	-0.16*
		-	(0.05)	(0.06)	(0.07)	(0.07)	(0.07)
	Surprise	-	-	0.17***	0.16**	0.24***	0.18**
		-	-	(0.05)	(0.06)	(0.06)	(0.06)
Content Controls	**Practical Utility**	-	-	0.25***	0.34***	0.18**	0.27***
		-	-	(0.05)	(0.06)	(0.07)	(0.06)
	Interest	-	-	0.26***	0.29***	0.31***	0.27***
		-	-	(0.05)	(0.06)	(0.07)	(0.07)
Homepage Location	**Top Feature**	-	-	-	0.13***	0.11***	0.11***
Control Variables		-	-	-	(0.02)	(0.02)	(0.03)
	Near Top Feature	-	-	-	0.11***	0.10***	0.12***
		-	-	-	(0.01)	(0.01)	(0.01)
	Right Column	-	-	-	0.14***	0.10***	0.15***
		-	-	-	(0.01)	(0.02)	(0.02)
	Middle Feature Bar	-	-	-	0.06***	0.05***	0.06***
		-	-	-	(0.00)	(0.01)	(0.01)
	Bulleted Sub-Feature	-	-	-	0.04**	0.04**	0.05*
		-	-	-	(0.01)	(0.01)	(0.02)
	More News	-	-	-	0.01	0.06***	-0.01
		-	-	-	(0.01)	(0.01)	(0.02)
	Bottom List x 10	-	-	-	0.06**	0.11***	0.08**
		-	-	-	(0.02)	(0.03)	(0.03)
Other Control	**Word Count x 10^{-3}**	-	-	-	0.52***	0.71***	0.57***
Variables		-	-	-	(0.11)	(0.12)	(0.18)
	Complexity	-	-	-	0.05	0.05	0.06
		-	-	-	(0.04)	(0.04)	(0.07)
	First Author Fame	-	-	-	0.17***	0.15***	0.15***
		-	-	-	(0.02)	(0.02)	(0.03)
	Female First Author	-	-	-	0.36***	0.33***	0.27*
		-	-	-	(0.08)	(0.09)	(0.13)
	Uncredited	-	-	-	0.39	-0.56*	0.50
		-	-	-	(0.26)	(0.27)	(0.37)
Newspaper Location & Web Timing Controls		No	No	No	Yes	Yes	Yes
Article Section Dummies (arts, books, etc.)		No	No	No	No	Yes	No
Observations		6,956	6,956	6,956	6,956	6,956	2,566
McFadden's R^2		0.04	0.07	0.08	0.28	0.36	0.32
Log pseudolikelihood		-3,118.45	-3,034.34	-2,998.96	-2,331.37	-2,084.85	-904.76

Logistic regressions include day fixed effects. ^Significant at the 10% level. *Significant at 5% level. **Significant at 1% level. ***Significant at the 0.1% level. Relative effect sizes of coded variables should be interpreted with care, as these variables are necessarily proxies for underlying constructs rather than exact measures of those constructs. Of greater interest is the large relative estimated effect of each of these proxy variables on an article's likelihood of making the most e-mailed list compared to the cleanly measured effects of external drivers of attention (e.g., time spent on various positions on the Times homepage).