What Makes Online Content Viral?

Jonah Berger

Katherine L. Milkman*

Forthcoming, *Journal of Marketing Research*

What Makes Online Content Viral?

ABSTRACT

Why are certain pieces of online content more viral than others? This article takes a psychological approach to understanding diffusion. Using a unique dataset of all the *New York Times* articles published over a three month period, the authors examine how emotion shapes virality. Results indicate that positive content is more viral than negative content, but that the relationship between emotion and social transmission is more complex than valence alone. Virality is driven, in part, by physiological arousal. Content that evokes high-arousal positive (awe) or negative (anger or anxiety) emotions is more viral. Content that evokes low arousal, or deactivating emotions (e.g., sadness) is less viral. These results hold even 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 content was featured). Experimental results further demonstrate the causal impact of specific emotion on transmission, and illustrate that it is driven by the level of activation induced. Taken together, these findings shed light on why people share content and provide insight into designing effective viral marketing campaigns.

KEYWORDS: Word-of-Mouth, Viral Marketing, Social Transmission, Online Content

Sharing online content is an integral part of modern life. People forward newspaper articles to their friends, pass YouTube videos to their relatives, and send restaurant reviews to their neighbors. Indeed, 59% of individuals 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 also has an important impact on 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 customer service experiences spread throughout the blogosphere while others are never shared. Some newspaper articles earn a position on their website's "most emailed list" while others languish. Companies often create online ad campaigns or encourage consumer-generated content in the hopes that people will share this content with others, but some of these efforts takeoff while others fail. Is virality just random, as some have argued (Cashmore 2009), or might certain characteristics predict whether content will be highly shared?

This paper examines how content characteristics impact virality. In particular, we focus on how emotion shapes social transmission. We do so in two ways. First, we analyze a unique dataset of nearly 7,000 *New York Times* articles to examine which articles make the newspaper's "most emailed list." Controlling for external drivers of

attention, such as where an article was featured online and for how long, we examine how content's valence (i.e., whether an article is positive or negative) as well as the specific emotions it evokes (e.g., anger, sadness, and awe) impact whether it is highly shared. Second, we experimentally manipulate the specific emotion evoked by content to directly test the causal impact of arousal on social transmission.

This research makes a number of important contributions. First, research on word-of-mouth and viral marketing has focused on its *impact* (i.e., on diffusion and sales; Godes and Mayzlin 2004; 2009; Goldenberg, et al., 2009), not its *causes*. But what drives people to share content with others and what type of content is more likely to be shared? By combining a large-scale examination of real transmission in the field with tightly controlled experiments, we both demonstrate characteristics of viral online content and shed light on the underlying processes that drive people to share. Second, our findings provide insight into how to design successful viral marketing campaigns. Word-of-mouth and social media are seen as cheaper and more effective than traditional media, but their utility hinges on people transmitting content that helps the brand. If no one shares a company's content, or if consumers share content that portrays the company negatively, the benefit of social transmission is lost. Consequently, understanding what drives people to share can help organizations and policy makers avoid consumer backlashes and craft contagious content.

CONTENT CHARACTERISTICS AND SOCIAL TRANSMISSION

One reason people may share stories, news, and information is because they contains useful information. Coupons or articles about good restaurants help people save money and eat better. Consumers may share such 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 content also has social exchange value (Homans 1958), and people may share it to generate reciprocity (Fehr, Kirchsteiger, Riedl 1998).

Emotional aspects of content may also impact whether it is shared (Heath, Bell, and Sternberg 2001). People report discussing many of their emotional experiences with others, and customers report greater word-of-mouth at the extremes of satisfaction (i.e., highly satisfied or highly dissatisfied, Anderson 1998). People may share emotionally charged content to make sense of their experiences, reduce dissonance, or deepen social connections (Festinger, Riecken, and Schachter 1956; Peters and Kashima 2007; Rime, et al. 1991).

Emotional Valence and Social Transmission

These observations suggest that emotionally evocative content may be particularly viral, but which is more likely to be 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 been tested. Further, the study on which this notion is based actually focused on understanding what types of news people encounter, not what they

transmit (see Goodman 1999). 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).

We hypothesize that more positive content will be more viral. Consumers often share content for self-presentation purposes (Wojnicki and Godes 2008) or to communicate identity, and consequently positive content may be shared more because it reflects positively on the sender. Most people would prefer to be known as someone who shares upbeat stories or makes others feel good rather than someone who shares things that makes others angry or upset. Sharing positive content may also help boost others' mood or provide information about potential rewards (i.e., this restaurant is worth trying).

The Role of Activation in Social Transmission

Importantly, however, the social transmission of emotional content may be driven by more than just valence. In addition to being positive or negative, emotions also differ on the level of physiological arousal or activation they evoke (Smith and Ellsworth 1985). Anger, anxiety, and sadness are all negative emotions, for example, but while anger and anxiety are characterized by states of heightened arousal or activation, sadness is characterized by low arousal or deactivation (Barrett and Russell 1998).

We suggest that these differences in arousal shape social transmission (also see Berger 2011). Arousal is a state of mobilization. While low arousal or deactivation is characterized by relaxation, high arousal or activation is characterized by activity (see Heilman 1997 for a review). Indeed, this excitatory state has been shown to increase

action related behaviors like getting up to help others (Gaertner and Dovidio 1977) or responding faster to offers in negotiations (Wood and Schweitzer 2011). Given that sharing information requires action, we suggest that activation should have similar effects on social transmission and boost the likelihood that content is highly shared.

If this is the case, then even two emotions of the same valence may have different effects on sharing if they induce different levels of activation. Consider something that makes people sad versus something that makes people angry. Both emotions are negative, so a simple valence-based perspective would suggest that content that induces either should be less viral (e.g., people want to make their friends feel good rather than bad). An arousal or activation based analysis, however, provides a more nuanced perspective. Even though both emotions are negative, anger might increase transmission (because it is characterized by high activation), while sadness might actually decrease transmission (because it is characterized by deactivation or inaction).

THE CURRENT RESEARCH

We examine how content characteristics drive social transmission and virality. In particular, we not only look at whether positive content is more viral than negative content, but go beyond mere valence to examine how specific emotions evoked by content, and the activation they induce, drive social transmission.

We study transmission in two ways. First, we investigate the virality of almost 7,000 articles from one of the world's most popular newspapers: *The New York Times* (Study 1). Controlling for a host of factors (e.g., where articles are featured and how

much interest they evoke), we examine how the emotionality, valence, and specific emotions evoked by an article impact its likelihood of making *The New York Times'* most emailed list. Second, we conduct a series of lab experiments (Study 2A, 2B, and 3) to test the underlying process we believe is responsible for the observed effects. By directly manipulating specific emotions and measuring the activation they induce, we test our hypothesis that content which evokes high arousal emotion is more likely to be shared.

STUDY 1: A FIELD STUDY OF EMOTIONS AND VIRALITY

Our first study investigates what types of *New York Times* articles are highly shared. The *Times* covers a wide range of topics (i.e., world news, sports, and travel), and its articles are shared with a mix of friends (42%), relatives (40%), colleagues (10%), and others (7%)[1], making it an ideal venue for examining the link between content characteristics and virality. They continually report which articles from their website have been most emailed in the last 24 hours, and we examine how (1) an article's valence and (2) the extent to which it evokes various specific emotions (e.g., anger or sadness), impact whether it makes the *Times'* most emailed list.

The state of the emotions literature is such that negative emotions have been much better distinguished from one another than positive emotions (Keltner and Lerner 2010). Consequently, while we also examine article valence, when considering specific emotions our archival analysis focuses on negative emotions because they are straightforward to differentiate and classify. Anger, anxiety, and sadness, are often

---

[1] Based on 343 *Times* readers who were asked with whom they had most recently shared an article.

described as basic or universal emotions (Ekman, Friesen, and Ellsworth 1982), and based on our earlier theorizing about activation, we predict that negative emotions characterized by activation (i.e., anger and anxiety) will be positively linked to virality, while negative emotions characterized by deactivation (i.e., sadness) will be negatively linked to virality.

We also examine whether awe, a high-arousal positive emotion, is linked to virality. Awe is characterized by a feeling of admiration and elevation in the face of something greater than the self (e.g., a new scientific discovery or someone overcoming adversity, Keltner and Haidt 2003). It is generated by stimuli that open the mind to unconsidered possibilities and the arousal it induces may promote transmission

Importantly, our empirical analyses control for a number of potentially confounding variables. First, as noted above, practically useful content may be more viral because it provides information. Self-presentation motives also shape transmission (Wojnicki and Godes 2008), and people may share interesting or surprising content because it is entertaining and reflects positively on them (i.e., suggests that they know interesting or entertaining things). Consequently, we control for these factors to examine the link between emotion and virality above and beyond them (though their relationships with virality may be of interest to some scholars and practitioners).

Second, our analyses also control for things beyond the content itself. Articles that appear on the front page of the newspaper or spend more time in prominent positions on the *Times'* 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.[2] Including these controls also allows us compare

the relative impact of placement versus content characteristics in shaping virality. While

being heavily advertised, or in this case prominently featured, should likely increase the

chance content makes the most emailed list, we examine whether content characteristics

(e.g., whether an article is positive or awe-inspiring) 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 in question. It recorded information about every article on the

homepage and each article on the most emailed list (updated every 15 minutes). 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 data set earned a position on the most e-mailed list.

Article Coding

We coded the articles on a number of dimensions. 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) and increase

---

[2] Discussion with the newspaper indicated that editorial decisions about how to feature articles on the homepage are made independently of (and well before) their appearance on the most emailed list.

coding ease and objectivity. Automated ratings were also significantly positively

correlated with manual coders ratings of a subset of articles A computer program

(LIWC) 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.

We relied on human coders to classify the extent to which content exhibited other,

more specific characteristics (e.g., evoked anger), as automated coding systems were not

available for these variables. In addition to coding whether articles contained practically

useful information or evoked interest or surprise (control variables), coders quantified the

extent to which each article evoked anxiety, anger, awe, or sadness.[3] 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 Appendix). 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, 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

the construct in question (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), indicating that

---

[3] Given that prior work has examined how disgust might impact the transmission of urban legends (Heath et al 2001) we also include disgust in our analysis (the rest of the results remain unchanged regardless of whether or not it is in the model). While we do not find any significant relationship between disgust and virality, this may be due in part to the fact that *New York Times* articles elicit little of this emotion.

content tends to evoke similar emotions across people. Scores were averaged across

coders and standardized. See Table 1 for sample articles that scored highly on the

different dimensions, Table 2 for summary statistics, and Table 3 for correlations

between variables. All uncoded articles were assigned a score of zero on each dimension

after standardization (meaning uncoded articles were assigned the mean value), and a

dummy was included in regression analyses to control for uncoded stories (see Cohen

and Cohen [1983] for a discussion of this standard 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. Using only

the coded subset of articles provides similar results.

Additional Controls

As discussed previously, external factors (separate from content characteristics)

may affect an article's virality by functioning like advertising. Consequently, we

rigorously control for such factors in our analyses (See Table 4 for a list of all

independent variables including controls).

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 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 similar 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.

Day-Dummies. Finally, we use day dummies to control for both competition among articles to make the most emailed list (i.e., other content that came out the same day) as well as any other time-specific effects (e.g., world events that might impact article characteristics and reader interest).

Analysis Strategy

Almost all articles that make the most emailed list do so only once (i.e., they do not leave the list and re-appear), so we model list making as a single event (see Appendix for further discussion). We rely on the following logistic regression specification:

$$
(1) \quad makes\_it_{at} = \frac{1}{1+\exp\left\{-\left(\begin{array}{l} \alpha_t + \beta_1 * \textit{z-emotionality}_{at} + \beta_2 * \textit{z-positivity}_{at} + \\ \beta_3 * \textit{z-awe}_{at} + \beta_4 * \textit{z-anger}_{at} + \beta_5 * \textit{z-anxiety}_{at} + \\ \beta_6 * \textit{z-sadness}_{at} + \theta' * X_{at} \end{array}\right)\right\}}
$$

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, or sadness-inducing. *X*$_{at}$ is a vector of the other control variables described above (see Table 4). We estimate the equation with fixed effects for the day of an article's release, clustering standard errors by day of release (results are similar if fixed effects are not included).

Results

   Is Positive or Negative Content More Viral? First, we examine content valence. Results indicate that content is more likely to become viral the more positive it is (Table 5, Model 1). Model 2 shows that more affect-laden content, regardless of valence, is more likely to make the most emailed list, but the returns to increased positivity persist even controlling for controlling for emotionality more generally. Looked at another way, when both the percentage of positive and negative words in an article are included as separate predictors (instead of emotionality and valence), both are positively associated with making the most emailed list. However, the coefficient on positive words is considerably larger than that on negative words. This indicates that while more positive *or* more negative content is more viral than content that does not evoke emotion, positive content is more viral than negative content.

   The nature of our dataset is particularly useful here because it allows us to disentangle preferential transmission from mere base rates (see Godes et al. 2005). Say

one found that there is more positive than negative WOM overall. It would be unclear whether this outcome was driven by (1) what people encounter (e.g., maybe people come across more positive events than negative ones) or (2) what people prefer to pass on (i.e., positive or negative content). Thus without knowing what people *could* have shared, it is hard to say much about what they *prefer* to share. Access to the full corpus of articles published by the *Times* over the analysis period as well as all content that made the most emailed list allows us separate these possibilities. Taking into account all published articles, our results show that an article is more likely to make the most emailed list the more positive it is.

How Do Specific Emotions Impact Virality? The relationships between specific emotions and virality suggest that the role of emotion in transmission is more complex than mere valence alone (Table 5, Model 3). While more awe-inspiring (a positive emotion) content is more viral, and sadness-inducing (a negative emotion) content is less viral, some negative emotions are positively associated with virality. More anxiety- and anger-inducing stories are both more likely to make the most emailed list. This suggests that transmission is about more than simply sharing positive things and avoiding sharing negative ones. In particular, consistent with our theorizing, content that evokes high-arousal emotions (i.e., awe, anger, and anxiety), regardless of their valence, is more viral.

Other Factors. These results persist controlling for a host of other factors (Table 5, Model 4). More interesting, informative (practically useful), and surprising articles are more likely to make the *Times'* most emailed list, but our focal results are significant even after controlling for these content characteristics. Similarly, being featured for longer in more prominent positions on the *Times* homepage (e.g., the lead story vs. at the

bottom of the page) is positively associated with making the most emailed list, but the relationships between emotional characteristics of content and virality persist even controlling for this type of "advertising." This suggests that the heightened virality of stories that evoke certain emotions is not simply driven by editors featuring those types of stories, which could mechanically increase their virality.[4] Longer articles, articles by more famous authors, and articles written by women are also more likely to make the most emailed list, but our results are robust to including these factors as well.

Robustness Checks. Results are also robust to controlling for an article's general topic (20 areas classified by the *Times* such as science or health; Table 5, Model 5). This indicates that our findings are not merely driven by certain areas tending to both evoke certain emotions and be particularly likely to make the most e-mailed list. Rather, this more conservative test of our hypothesis suggests that the observed relationships between emotion and virality hold not only across topics but also within them. Even among opinion or health articles, for example, awe-inspiring articles are more viral.

Finally, our results remain meaningfully unchanged in terms of magnitude and significance if we perform a host of other robustness checks including only analyzing the 2,566 hand-coded articles (Table 5, Model 6), removing day fixed effects, and using alternate ways of quantifying emotion (see Appendix for more robustness checks and analyses using article rank or time on the most emailed list as alternate dependent measures). These results indicate that the observed results are not an artifact of the particular regression specifications we rely on in our primary analyses.

---

[4] 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 general topical areas (e.g., opinion), are strongly related to whether and where articles are featured on the homepage, while emotional characteristics are not.

Discussion

Analysis of over three months of *New York Times* articles sheds light on what types of online content become viral and why. Contributing to the debate on whether positive or negative content is more likely to be shared, our results demonstrate that more positive content is more viral. Importantly, however, our findings also reveal that virality is driven by more than just valence. Sadness, anger, and anxiety are all negative emotions, but while sadder content is less viral, content that evokes more anxiety or anger is actually more viral. These findings are consistent with our hypothesis about how arousal shapes social transmission. Positive and negative emotions characterized by activation or arousal (i.e., awe, anxiety, and anger) are positively linked to virality, while emotions characterized by deactivation (i.e., sadness) are negatively linked to virality.

More broadly, our results suggest that while external drivers of attention (e.g., being prominently featured) shape what becomes viral, content characteristics are of similar importance (see Figure 2). For instance, a one standard deviation increase in the amount of anger an article evokes increases the odds that it will make the most e-mailed list by 34% (Table 5, Model 4). This increase is equivalent to spending an additional 2.9 hours as the lead story on the *Times* website, which is nearly four times the average number of hours articles spend in that position. Similarly, a one standard deviation increase in awe increases the odds of making the most e-mailed list by 30%.

These field results are consistent with the notion that activation drives social transmission, but to more directly test the process behind our specific emotions findings, we next turn to the laboratory.

STUDY 2: HOW HIGH-AROUSAL EMOTIONS AFFECT TRANSMISSION

Our experiments had three main goals. First, we wanted to directly test the causal impact of specific emotions on social transmission. The field study illustrates that content which evokes activating emotions is more likely to be viral, but by manipulating specific emotions in a more controlled setting, we can more cleanly examine how they affect transmission. Second, we wanted to test the hypothesized mechanism behind these effects, namely whether the arousal induced by content drives transmission. Third, while the *New York Times* provided a broad domain to study transmission, we wanted to test whether our findings would generalize to other marketing content.

We asked participants how likely they would be to share a story about a recent advertising campaign (Study 2a) or customer service experience (Study 2b) and manipulated whether the story in question evoked more or less of a specific emotion (amusement in Study2a and anger in Study2b). To test the generalizability of the effects, we looked at how both positive (amusement, Study2a) and negative (anger, Study2b) high-arousal emotions characterized influence transmission. If arousal increases sharing, then consistent with our field study, content that evokes more of an activating emotion (amusement or anger) should be more likely to be shared. Finally, we also measured experienced activation to test whether it drives the effect of emotion on sharing.

Study2A - Amusement

Participants (N = 49) were randomly assigned to read either a high or low amusement version of a story about a recent advertising campaign for Jimmy Dean

sausages. The two versions were adapted from prior work (McGraw and Warren 2010) showing that they differed on how much humor they evoked (a pre-test showed they do not differ in how much interest they evoked). In the low amusement condition, Jimmy Dean decides to hire a farmer as the new spokesperson for the company's line of pork products. In the high amusement condition, Jimmy Dean decides to hire a rabbi (which is funny given that they make pork products and pork is not considered kosher). After reading about the campaign, participants were asked how likely they would be to share it with others (1 = not at all likely, 7 = extremely likely).

Participants also rated their level of arousal using three 7-point scales ("How do you feel right now?" very passive-very active; very mellow-very fired up; very low energy-very high energy, $\alpha = 82$, adapted from Berger 2011 and averaged to form an Activation Index).

Results. As predicted, participants said they would be more likely to share the advertising campaign when it induced more amusement, and this was driven by the arousal it evoked. First, participants said they would be more likely to share the advertisement if they were in the high (M = 3.97) as opposed to low amusement condition (M = 2.92; $F(1, 47) = 10.89, p < .005$). Second, the results were similar for arousal; the high amusement condition (M = 3.73) evoked more arousal than the low amusement condition (M = 2.92; $F(1, 47) = 5.24, p < .05$). Third, as predicted, this boost in arousal mediated the effect of the amusement condition on sharing. Condition was linked to arousal ($\beta_{high\_amusement} = .39$, SE = .17, $t(47) = 2.29, p < .05$), arousal was linked to sharing ($\beta_{activation} = .58$, SE = .11, $t(47) = 5.06, p < .001$), and when both the

amusement condition and arousal were included in a regression predicting sharing, arousal mediated the effect of amusement on transmission (Sobel $z = 2.02$, $p < .05$).

Study2B - Anger

Participants ($N = 45$) were randomly assigned to read either a high or low anger version of a story about a (real) negative customer service experience with United Airlines (Negroni 2009). The two versions were pretested to evoke different amounts of anger but not other specific emotions, interest, or positivity in general. In both conditions, the story described a music group traveling on United Airlines to begin a week-long-tour of shows in Nebraska. As they were about to leave, however, they noticed that the United baggage handlers were mishandling their guitars. They asked for help from flight attendants, but by the time they landed, the guitars had been damaged. In the high anger condition, the story was entitled "United Smashes Guitars," and described how the baggage handlers seemed not to care about the guitars and how United was unwilling to pay for the damages. In the low anger condition, the story was entitled "United Dents Guitars," and described the baggage handlers as dropping the guitars but United being willing to help pay for the damages.

After reading the story, participants rated how likely they would be to share the customer service experience as well as their arousal using the scales from Study 2A.

Results. As predicted, participants said they would be more likely to share the customer service experience when it induced more anger, and this was driven by the arousal it evoked. First, participants reported being more likely to share the customer service experience if they were in the high anger ($M = 5.71$) as opposed to low anger

condition (M = 3.37; $F(1, 43) = 18.06$, $p < .001$). Second, the results were similar for arousal; the high anger condition (M = 4.48) evoked more arousal than the low anger condition (M = 3.00; F(1, 43) = 10.44, p < .005). Third, as in study 2A, this boost in arousal mediated the effect of condition on sharing. Regression analyses show that condition was linked to arousal ($\beta_{high\_anger} = .74$, SE = .23, t(44) = 3.23, p < .005), arousal was linked to sharing ($\beta_{activation} = .65$, SE = .17, t(44) = 3.85, p < .001), and when both anger condition and arousal were included in a regression, arousal mediated the effect of anger on transmission (Sobel z = 1.95, p = .05).

Discussion

The experimental results reinforce the findings from our archival field study, support our hypothesized process, and generalize our findings to a broader range of content. First, consistent with our analysis of the *New York Times'* most emailed list, the amount of emotion content evoked influenced transmission. People said they would be more likely to share an advertisement when it evoked more amusement (Study2a) and a customer service experience when it evoked more anger (Study2b). Second, the results underscore our hypothesized mechanism: Arousal mediated the impact of emotion on social transmission. Content that evokes more anger or amusement is more likely to be shared, and this is driven by the level of activation it induces.

STUDY 3: HOW DEACTIVATING EMOTIONS AFFECT TRANSMISSION

Our final experiment further tests the role of arousal by examining how *de*activating emotions affect transmission. Studies 2a and 2b show that increasing the amount of high arousal emotions boosts social transmission due to the activation it induces, but if our theory is correct, these effects should reverse for low arousal emotions. Content which that evokes more sadness, for example, should be *less* likely to be shared because it *deactivates* rather than activates.

Note that this is a particularly strong test of our theory because the prediction goes against a number of alternative explanations for our findings in Study 2. One could argue that evoking more of any specific emotion makes content better, or more compelling, but such an explanation would suggest evoking more sadness should increase (rather than decrease) transmission.

Method

Participants (N = 47) were randomly assigned to read either a high or low sadness version of a news article. The two versions were pretested to evoke different amounts of sadness but not other specific emotions, interest, or positivity in general. In both conditions, the article described someone who had to have titanium pins implanted in her hands and relearn her grip after sustaining injuries. The difference between conditions was the source of the injury. In the high sadness condition, the story was taken directly from our *New York Times* dataset. It was entitled "Maimed on 9/11, Trying to be Whole Again," and detailed how someone who worked in the World Trade Center sustained an injury during 9/11. In the low sadness condition, the story was entitled "Trying to be Better Again," and detailed how the person sustained the injury falling down the stairs at

her office. After reading one of these two versions of the story, participants answered the

same sharing and arousal questions as in Study 2.

As predicted, when the context evoked a deactivating (i.e., de-arousing) emotion,

the effects on transmission were reversed. First, participants were *less* likely to share the

story if they were in the high sadness (M = 2.39) as opposed to the low sadness condition

(M = 3.80; $F(1, 46) = 10.78$, $p < .005$). Second, the results were similar for arousal; the

high sadness condition (M = 2.75) evoked *less* arousal than the low sadness condition (M

= 3.89; $F(1, 46) = 10.29$, $p < .005$). Third, as hypothesized, this decrease in arousal

mediated the effect of condition on sharing. Condition was linked to arousal ($\beta_{high\_sadness}$

= -.57, SE = .18, $t(46) = -3.21$, $p < .005$), arousal was linked to sharing ($\beta_{activation}$ = .67, SE

= .15, $t(46) = 4.52$, $p < .001$), and when both sadness condition and arousal were included

in a regression predicting sharing, arousal mediated the effect of sadness on transmission

(Sobel $z = -2.32$, $p < .05$).

Discussion

Results of Study 3 further underscore the role of arousal in social transmission.

Consistent with the findings of our field study, when content evoked more of a low

arousal emotion it was actually *less* likely to be shared. Further, these effects were again

driven by arousal. When a story evoked more sadness it decreased arousal which, in turn,

decreased transmission. The fact that the effect of specific emotion intensity on

transmission reversed when the emotion was deactivating provides even stronger

evidence for our theoretical perspective. While one could argue that content which

evokes more emotion is more interesting or engaging (and indeed, pretest results suggest

that this is the case in this experiment), these results show that such increased emotion may actually decrease transmission if the specific emotion evoked is characterized by deactivation.

GENERAL DISCUSSION

The emergence of 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 about why consumers share content or why certain content becomes viral. Further, though diffusion research has examined how certain individuals (e.g., social hubs or influentials) and 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 a multi-method approach to studying virality. By combining a broad analysis of virality in the field with a series of controlled laboratory experiments, we document characteristics of viral content while also shedding light on what drives social transmission.

Our findings make a number of contributions to the existing literature. First, they 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.

Second, our results illustrate that the relationship between emotion and virality is about by more than just valence, and that arousal drives social transmission. Consistent with our theorizing, online content that evoked high-arousal emotions was more viral, regardless of whether those emotions were of a positive (i.e., awe) or negative (i.e., anger or anxiety) nature. Online content that evoked more of a deactivating emotion (i.e., sadness), however, was actually less likely to be viral. Experimentally manipulating specific emotions in a controlled environment confirms the hypothesized causal relationship between activation and social transmission. When marketing content evoked more of specific emotions characterized by arousal (i.e., amusement, Study 2a or anger Study 2b) it was more likely to be shared, but when it which evoked specific emotion characterized by deactivation (i.e., sadness, Study 3) it was less likely to be shared. In addition, these effects were mediated by arousal, further underscoring its impact on social transmission.

Demonstrating these relationships in both the laboratory and the field, as well as across a large and diverse body of content, underscores their generality. Further, although not a focus of our analysis, our field study also adds to the literature by demonstrating that more practically useful, interesting, and surprising content is more viral. Finally, the naturalistic setting allows us to measure the relative importance of content characteristics and external drivers of attention in shaping virality. While being featured prominently, for example, increases the likelihood that content will be 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 past research has modeled product adoption (Bass 1969) and examined how social networks shape diffusion and sales (Van den Bulte and Wuyts 2007), macro-level collective outcomes such as what becomes viral also 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 that drive social transmission (Berger 2011; Berger and Schwartz 2011). Along these lines, this work suggests that the emotion (and activation) that content evokes in individuals helps determine which cultural items succeed in the marketplace of ideas.

Our findings also suggest that social transmission is about more than just value exchange or self-presentation (also see Berger & Schwartz, 2011). Consistent with the notion that people share to entertain others, 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 content to help others, generate reciprocity, or boost their reputation (e.g., show they know entertaining or useful things). Even controlling for these effects, however, we find that highly arousing content (e.g., anxiety- or anger-evoking) 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. This suggests that social transmission may be less about motivation and more about the transmitter's internal states.

It is also interesting to consider these findings in relation to literature on characteristics of effective advertising. Just as certain characteristics of advertisements may make them more effective, certain characteristics of content may make it more likely to be shared. While there is likely some overlap in these factors (e.g., creative ads are more effective, Goldenberg, Mazursky, and Solomon 1999, and likely shared more), there may also be some important differences. For example, while negative emotions may hurt brand and product attitudes (Edell and Burke 1987), we have shown that some negative emotions can actually increase social transmission.

Directions for Future Research

Future work might 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 to a friend who likes crew), broadcasting likely requires posting content that has broader appeal. One could also imagine that while narrowcasting is recipient-focused (i.e., what a recipient would enjoy), broadcasting is self-focused (i.e., what someone wants to say about themselves or show others). Consequently, self-presentation motives, identity signaling (e.g., Berger and Heath 2007), or affiliation goals may 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

our data collection, we built a supplementary web-crawler to capture the *Times'* list of the 25 articles that had 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 factors. Given that the weather can affect people's moods (Keller et al. 2005), 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. Other cues in the environment might also shape social transmission by making certain topics more accessible (Berger and Fitzsimons 2008; Berger and Schwartz 2011; Nedungadi 1990). When the World Series is going on, for example, people may be more likely to share any sports story more generally because that topic is primed.

These findings also raise broader questions, such as how much of social transmission is driven by the sender versus the receiver, and how much of it is motivated versus unmotivated. While intuition might suggest that much of transmission is motivated (i.e., wanting to look good to others) and based on the receiver and what they would find of value, the current results highlight the important role that the sender's

internal states play in whether something gets shared. That said, deeper understanding of these issues requires further research.

Marketing Implications

These findings also have a number of important marketing implications. Considering the specific emotions content evokes should help companies maximize revenue when placing advertisements and should help online content providers when pricing access to content (e.g., potentially charging more for content that is likely to be highly shared). It might also be useful to feature or design content that evokes activating emotions, as such content is likely to be shared (thus increasing page views).

Our findings also shed light on how to design successful viral marketing campaigns and craft contagious content. While marketers often produce content that paints their product in a positive light, our results suggest that content will be more likely to be shared if it evokes high-arousal emotions. Ads that make consumers content or relaxed, for example, will not be as viral as those that amuse them. Further, while some marketers might shy away from ads that evoke negative emotions, our results suggest that negative emotion can actually increase transmission if it is characterized by activation. BMW, for example, created a series of short online films called "The Hire" that they hoped would go viral, and which included car chases and story lines that often evoked anxiety (with such titles as "Ambush" and "Hostage"). While one might be concerned that negative emotion would hurt the brand, our results suggest that it should increase transmission because anxiety induces arousal. (Incidentally, "The Hire" was highly successful, generating millions of views). Following this line of reasoning, public health

information should be more likely to be passed on if it is framed to evoke anger or anxiety rather than sadness.

Similar points apply to managing online consumer sentiment. While some consumer-generated content (e.g., reviews and blog posts) is positive, much is negative, and can build into consumer backlashes if it is not carefully managed. Moms offended by a Motrin ad campaign, for example, banded together and began posting negative YouTube videos and tweets (Petrecca 2008). While it is impossible to address all negative sentiment, our results suggest that certain types of negativity may be more important to address because they are more likely to be shared. Customer experiences that evoke anxiety or anger, for example, should be more likely to be shared than those that evoke sadness (and textual analysis can be used to distinguish different types of posts). Consequently, it may be more important to rectify experiences that make consumers anxious rather than disappointed.

In conclusion, this research illuminates how content characteristics shape whether it becomes viral. When looking to generate word-of-mouth, marketers often try targeting "influentials" or opinion leaders (i.e., some small set of special individuals who, whether through having more social ties or being more persuasive, theoretically have more influence than others). But while this approach is pervasive, recent research casts doubt on its value (Bakshy, et al., 2011; Watts 2007) and suggests it is far from cost-effective. Rather than targeting "special" people, the current research suggests that it may be more beneficial to focus on crafting contagious content. By considering how psychological processes shape social transmission, one can gain deeper insight into collective outcomes, such as what becomes viral.

REFERENCES

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.

Anderson, Eugene W. (1998), "Customer Satisfaction and Word-of-mouth," *Journal of Service Research*, 1(1), 5-17.

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

Bakshy, Eytan, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts (2011), "Everyone's An Influencer: Quantifying Influence on Twitter," *Proceedings of WSDM'2011*. 65-74.

Barrett, Lisa Feldman and James A. Russell (1998), "Independence and Bipolarity in the Structure of Current Affect," *Journal of Personality and Social Psychology,* 74, 967-984.

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

Berger, Jonah (2011), "Arousal Increases Social Transmission of Information," *Psychological Science*, 22(7), 891-893.

Berger, Jonah and Chip Heath (2007), "Where Consumers Diverge from Others: Identity-Signaling and Product Domains," *Journal of Consumer Research*, 34(2), 121-134.

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.

Berger, Jonah and Eric Schwartz (2011), "What Drives Immediate and Ongoing Word-of-Mouth?" *Journal of Marketing Research.*

Brooks, Alison Wood and Maurice E. Schweitzer, "Can Nervous Nelly negotiate? How anxiety causes negotiators to make low first offers, exit early, and earn less profit," *Organizational Behavior and Human Decision Processes*, 115, 43-54.

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

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.

Edell, Julie A. and Marian C. Burke (1987), "The Power of Feelings in Understanding Advertising Effects," *Journal of Consumer Research*, 14 (December), 421-433.

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.

Gaertner, Samuel L. and John F. Dovidio (1977), "The subtlety of white racism, arousal, and helping behavior," *Journal of Personality and Social Psychology*, 35, 691-707.

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, Han Sangman, Donald R. Lehmann, and Jae W. Hong (2009), "The Role of Hubs in the Adoption Process," *Journal of Marketing*, 73 (2), 1-13.

Goldenberg, Jacob, David Mazursky, and Sorin Solomon (1999), "Creativity Templates: Towards Identifying the Fundamental Schemes of Quality Advertisements," *Marketing Science*, 18(3), 333-351.

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.

Heilman, Kenneth M. (1997), "The Neurobiology of Emotional Experience," *Journal of Neuropsychiatry and Clinical Neuroscience*, 9, 439-448.

Homans, George C. (1958), "Social Behavior as Exchange," *American Journal of Sociology*, 63 (6), 597-606.

Katz, Elihu, and Paul Felix Lazarsfeld (1955), *Personal Influence: The Part Played by People in the Flow of Mass Communication*. Glencoe, IL: Free Press.

Keller, Matthew C., Barbara L. Fredrickson, Oscar Ybarra, Stephane Cote, Kareem Johnson, Joe Mikels, Anne Conway, and Tor Wager (2005), " AWarm Heart and a Clear Head: The Contingent Effects of Weather on Mood and Cognition," *Psychological Science*, 16(9), 724-731.

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.

Koppel, Moshe, Shlomo Argamon, and Anat Rachel Shimoni (2002), "Automatically Categorizing Written Texts by Author Gender," *Literary and Linguistic Computing,* 17, 401-412.

McLaughlin, G. Harry (1969), "SMOG Grading: A New Readability Formula," *Journal of Reading,* 12, 639-646.

McGraw, A. Peter and Caleb Warren (2010), "Benign Violations: Making Immoral Behavior Funny," *Psychological Science*, 21, 1141-1149.

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.

Negroni, Christine (2009), "With Video, a Traveler Fights Back," *New York Times,* October 28.

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

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.

Smith, C. A., & Ellsworth, P. C. (1985), "Patterns of Cognitive Appraisal in Emotion," *Journal of Personality and Social Psychology*, 48, 813-838.

Van den Bulte, Christophe and Stefan Wuyts (2007), *Social Networks and Marketing,* Cambridge, MA: Marketing Science Institute.

Watts, Duncan J. (2007), "Challenging the Influentials Hypothesis," *WOMMA Measuring Word of Mouth*, 3, 201-211.

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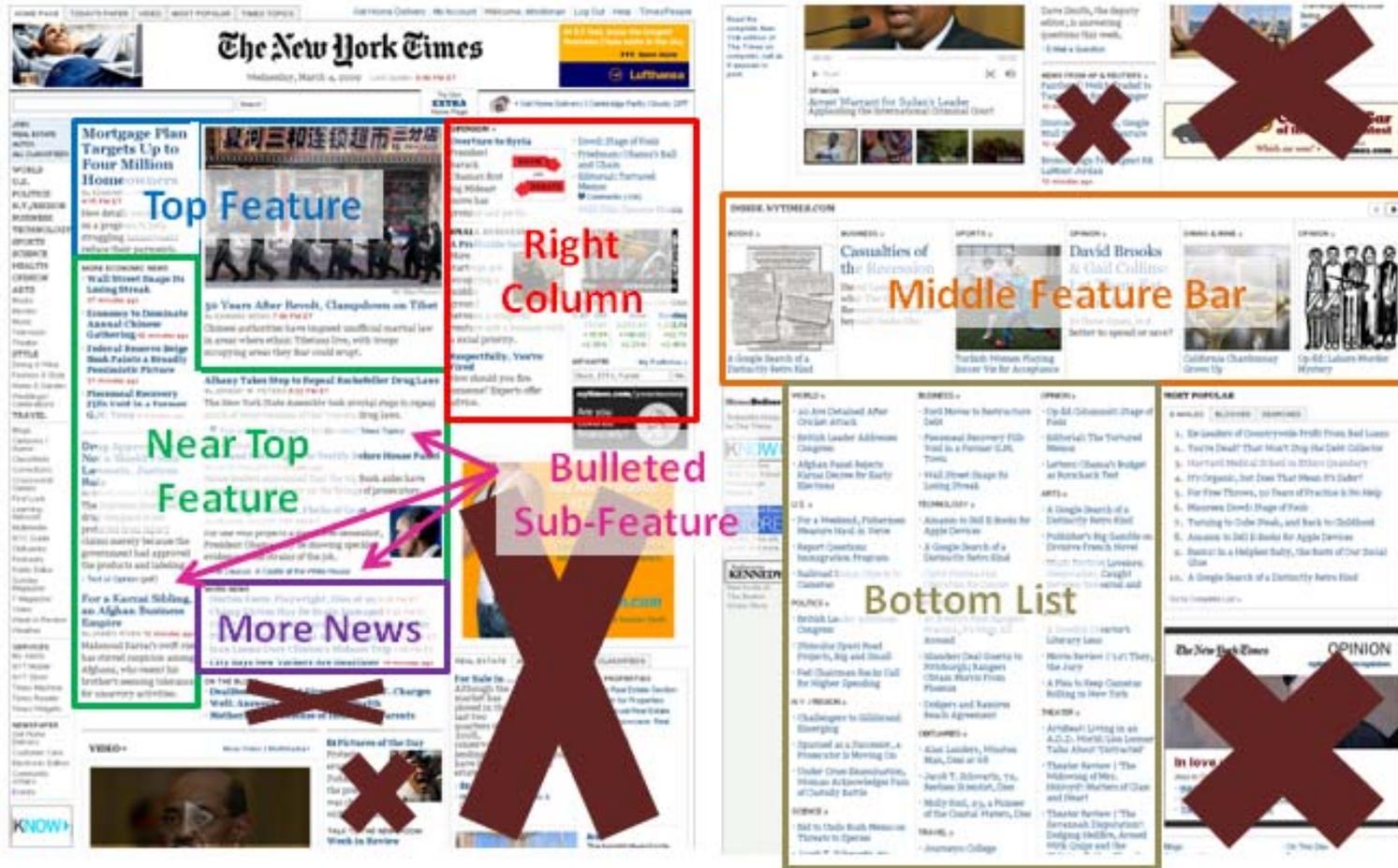
FIGURE 2
PERCENT CHANGE IN FITTED PROBABILITY OF MAKING THE LIST FOR A 1 STANDARD DEVIATION INCREASE
ABOVE THE MEAN IN AN ARTICLE CHARACTERISTIC

% Change in Fitted Probability of Making the List
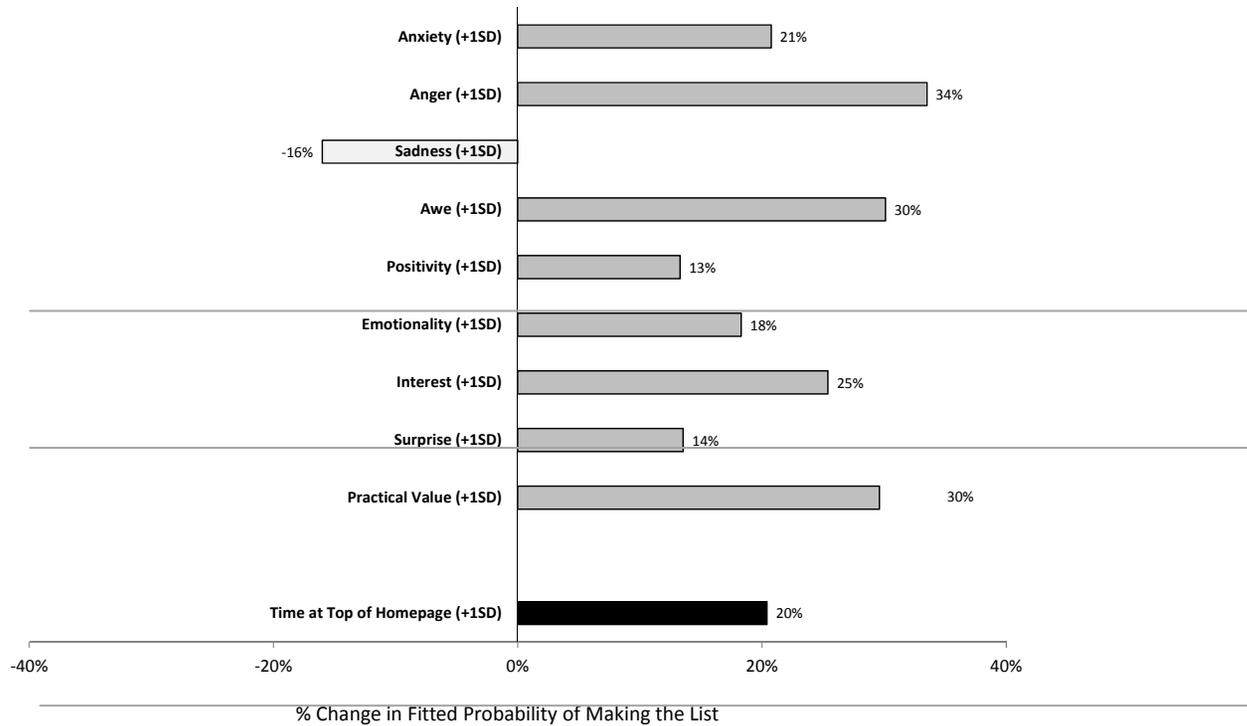
TABLE 1

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"
Sadness	*High Scoring:* • "Maimed on 9/11, Trying to Be Whole Again" • "Obama Pays Tribute to His Grandmother After She Dies"
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"
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)

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
	Sadness*	1.31	0.41
Other Control Variables	**Practical Utility***	1.66	1.01
	Interest*	2.71	0.85
	Surprise*	2.25	0.87
	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	Sadness	Practical Utility	Interest	Surprise	Word Count x 10^-3	Complexity	Author Fame	Author Female	Missing	Top Feature	Near Top Feature	Right Column	Bulleted Sub-Feature	More News	Middle 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															
Sadness	0.00	-0.18*	0.08*	0.42*	0.45*	1.00														
Practical Utility	0.06*	0.04*	-0.11*	-0.12*	0.07*	-0.05*	1.00													
Interest	0.054*	0.07*	0.26*	-0.13*	-0.24*	-0.19*	-0.06*	1.00												
Surprise	-0.10*	-0.04*	0.24*	-0.01	0.00	0.05*	-0.05*	0.18*	1.00											
Word Count x 10^-3	0.06*	0.05*	0.04*	0.02	0.00	0.00	-0.01	0.06*	0.02*	1.00										
Complexity	0.05*	-0.05*	-0.04*	0.10*	0.13*	0.05*	0.01	-0.11*	0.04*	-0.06*	1.00									
Author Fame	-0.09*	-0.03*	0.06*	0.01	0.03*	0.01	-0.02	0.00	0.02	0.01	0.01	1.00								
Author Female	-0.07*	0.06*	0.01	-0.03*	0.00	0.00	0.05*	-0.01	0.07*	0.00	-0.02*	0.00	1.00							
Missing	0.21*	0.03*	-0.06*	0.03*	-0.02	0.00	0.01	0.02	-0.09*	-0.01	0.02*	-0.71*	-0.15*	1.00						
Top Feature	0.01	-0.02	-0.03*	0.06*	0.06*	0.05*	0.02	-0.03*	-0.02*	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*	-0.03*	-0.05*	0.01	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	0.05*	0.06*	-0.02*	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*	0.06*	0.04*	-0.05*	-0.04*	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*	0.06*	-0.08*	-0.04*	0.07*	-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*	-0.05*	0.00	0.10*	0.04*	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*	-0.06*	0.06*	0.09*	0.04*	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	Manually coded
Anger	Manually coded
Anxiety	Manually coded
Sadness	Manually coded
Practical Utility	Manually coded
Interest	Manually coded
Surprise	Manually coded
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, et al. 2003)
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
AN ARTICLE'S LIKELIHOOD OF MAKING THE *NEW YORK TIMES'* MOST E-MAILED LIST AS A FUNCTION OF ITS CONTENT CHARACTERISTICS

		Positivity	Emotionality	Specific Emotions	Including Controls	Including Section Dummies	Only Coded Articles
		(1)	(2)	(3)	(4)	(5)	(6)
Emotion Predictors	**Positivity**	0.13***	0.11***	0.17***	0.16***	0.14***	0.23***
		(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)
	Emotionality	-	0.27***	0.26***	0.22***	0.09*	0.29***
		-	(0.03)	(0.03)	(0.04)	(0.04)	(0.06)
Specific Emotions	**Awe**	-	-	0.46***	0.34***	0.30***	0.36***
		-	-	(0.05)	(0.05)	(0.06)	(0.06)
	Anger	-	-	0.44***	0.38***	0.29**	0.37***
		-	-	(0.06)	(0.09)	(0.10)	(0.10)
	Anxiety	-	-	0.20***	0.24***	0.21***	0.27***
		-	-	(0.05)	(0.07)	(0.07)	(0.07)
	Sadness	-	-	-0.19***	-0.17*	-0.12^	-0.16*
		-	-	(0.05)	(0.07)	(0.07)	(0.07)
Content Controls	**Practical Utility**	-	-	-	0.34***	0.18**	0.27***
		-	-	-	(0.06)	(0.07)	(0.06)
	Interest	-	-	-	0.29***	0.31***	0.27***
		-	-	-	(0.06)	(0.07)	(0.07)
	Surprise	-	-	-	0.16**	0.24***	0.18**
		-	-	-	(0.06)	(0.06)	(0.06)
Homepage Location Control Variables	**Top Feature**	-	-	-	0.13***	0.11***	0.11***
		-	-	-	(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 Variables	**Word Count x 10^{-3}**	-	-	-	0.52***	0.71***	0.57***
		-	-	-	(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		0.00	0.04	0.07	0.28	0.36	0.32
Log pseudolikelihood		-3,245.85	-3,118.45	-3,034.17	-2,331.37	-2,084.85	-904.76

Logistic regressions models appear above predicting whether an article makes the *New York Times'* most emailed list. Successive models include added control variables with th e exception of Model 6. Model 6 presents our primary regression specification (see Model 4) including only observations of articles whose content was hand-coded by research assistants. All models include day fixed effects. ^Significant at the 10% level. *Significant at 5% level. **Significant at 1% level. ***Significant at the 0.1% level. Models (4)-(6) include disgust (hand-coded) as a control, as disgust has been linked to transmission in previous research (Heath et al., 2001), and including this control thus allows for a more conservative test of our hypotheses. Its effect is never signficant, and dropping this control variable does not change any of our results.

APPENDIX

Data Used

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.

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 seven) 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.

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.

Additional Robustness Checks.

Results are robust to (1) adding squared and/or cubed terms quantifying how long an article spent in each of seven homepage regions; (2) adding dummies indicating whether an article ever appeared in a given homepage region; (3) splitting the homepage region control variables into time spent in each region during the day (6 am – 6 pm EST) and night (6 am – 6 pm EST); (4) controlling for the day of the week when an article was published in the physical paper (instead of online); (5) winsorizing the top and bottom 1% of outliers for each control variable in our regression; (6) controlling for the first homepage region in which an article was featured on the *Times'* site; (7) 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; or (8) including interaction terms for each our primary predictor variables with dummies for each of the 20 topic areas classified by the *Times*.

Alternate Dependent Measures.

Making the 24-hour most emailed list is a binary variable (an article either makes it or it does not), and while we do not have access to the actual number of times articles are emailed, we do know the highest rank an article achieves on the most emailed list. Drawing strong conclusions from an analysis of this outcome measure is problematic, however, for a number of reasons. First, once an article earns a position on the most emailed list, it receives considerably more "advertising" than other stories. Some people look to the most emailed list every day to determine what articles to read. It is unclear, however, exactly how to properly control for this issue. For example, the top 10 most emailed stories over 24 hours are featured prominently on the *Times'* homepage, but readers must then click on a link to see the rest of the most emailed list (articles 11-25). This suggests that it may be inappropriate to assume that the same model predicts performance from rank 11 – 25 as rank 1 – 10. Second, any model assuming equal spacing between ranked categories is problematic, as the difference in virality between stories ranked 22 and 23 may be very small compared to the difference in virality between stories ranked 4 and 5, reducing the ease of interpretation of any results involving rank as an outcome variable. That said, using an ordered logit model, and coding articles that never make the most emailed list as earning a rank of "26" (leaving these articles out of the analysis introduces additional selection problems), we find nearly identical results to our primary analyses presented in Table 5 (Table A3).

Another question is persistence, or how long articles continue to be shared. This is an interesting issue, but unfortunately it cannot be easily addressed with our data. We do not have information about when articles were shared over time, only how long they spent on the most emailed list. Analyzing time spent on the most emailed list shows that both more affect-laden and more interesting content spends longer on the list (Table A3). However, this alternative outcome variable also has a number of problems. First, there is a selection problem: only articles that make the most emailed list have an opportunity to spend time on the list. This both restricts the number of articles available for analysis and ensures that all articles studied contain highly viral content. Second, as discussed above, articles that make the most emailed list receive different amounts of additional "advertising" on the *Times* homepage depending on what rank they achieve (top 10 articles are displayed prominently). Consequently, while it is difficult to infer too much from these ancillary results, they highlight an opportunity for future research.

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%	-	-

TABLE A3
AN ARTICLE'S HIGHEST RANK AND LONGEVITY ON THE *NEW YORK TIMES'*
MOST E-MAILED LIST AS A FUNCTION OF ITS CONTENT CHARACTERISTICS

Outcome Variable:		Highest Rank	Hours on List
		(7)	(8)
Emotion Predictors	**Emotionality**	0.22***	2.25**
		(0.04)	(0.85)
	Positivity	0.15***	0.72
		(0.04)	(0.81)
Specific Emotions	**Awe**	0.25***	-1.47
		(0.05)	(1.11)
	Anger	0.35***	0.35
		(0.08)	(1.14)
	Anxiety	0.19**	0.36
		(0.06)	(0.95)
	Sadness	-0.16**	-0.77
		(0.06)	(0.93)
Content Controls	**Practical Utility**	0.31***	0.38
		(0.05)	(1.07)
	Interest	0.27***	1.85^
		(0.06)	(1.00)
	Surprise	0.17***	1.04
		(0.05)	(0.85)
Homepage Location Control Variables	**Top Feature**	0.11***	-0.18
		(0.02)	(0.18)
	Near Top Feature	0.11***	0.21^
		(0.01)	(0.13)
	Right Column	0.15***	0.88***
		(0.01)	(0.17)
	Middle Feature Bar	0.05***	-0.01
		(0.00)	(0.06)
	Bulleted Sub-Feature	0.03*	-0.21
		(0.01)	(0.22)
	More News	0.01	0.32
		(0.01)	(0.24)
	Bottom List x 10	0.04*	0.07
		(0.02)	(0.22)
Other Control Variables	**Word Count x 10^{-3}**	0.37***	4.67*
		(0.08)	(1.99)
	Complexity	0.01	-1.10
		(0.03)	(0.95)
	First Author Fame	0.21***	1.89***
		(0.02)	(0.55)
	Female First Author	0.37***	4.07**
		(0.07)	(1.35)
	Uncredited	0.74***	13.29^
		(0.26)	(7.53)
Newspaper Location & Web Timing Controls		Yes	Yes
Article Section Dummies (arts, books, etc.)		No	No
Observations		6,956	1,391
Regression Modeling Approach		Ordered Logit	Ordinary Least Squares
Pseudo R^2/R^2		0.13	0.23
Log pseudolikelihood		-6,929.97	N/A

Regressions models above examine the content characteristics of an article associated with its highest rank achieved on the *New York Times'* most emailed list (reverse-scored such that 25 = the top of the list and 0 = never on the list) and its longevity on the list. Both models rely on our primary specification (see Table 5, Model 4) and include day fixed effects. ^Significant at the 10% level. *Significant at 5% level. **Significant at 1% level. ***Significant at the 0.1% level.