



# HHS Public Access

Author manuscript

*J Commun.* Author manuscript; available in PMC 2016 June 01.

Published in final edited form as:

*J Commun.* 2015 June 1; 65(3): 512–534. doi:10.1111/jcom.12160.

## Attracting Views and Going Viral: How Message Features and News-Sharing Channels Affect Health News Diffusion

Hyun Suk Kim

University of Pennsylvania

### Abstract

This study examined how intrinsic as well as perceived message features affect the extent to which online health news stories prompt audience selections and social retransmissions, and how news-sharing channels (e-mail vs. social media) shape what goes viral. The study analyzed actual behavioral data on audience viewing and sharing of *New York Times* health news articles, and associated article content and context data. News articles with high informational utility and positive sentiment invited more frequent selections and retransmissions. Articles were also more frequently selected when they presented controversial, emotionally evocative, and familiar content. Informational utility and novelty had stronger positive associations with e-mail-specific virality, while emotional evocativeness, content familiarity, and exemplification played a larger role in triggering social media-based retransmissions.

### Keywords

selective exposure; virality; selection; retransmission; diffusion; message effects; social media; big data; computational social science

---

The Internet and digital media technologies have turned news consumption into an increasingly more *selective* and *social* communication behavior (Napoli, 2011). People have the opportunity to exercise greater selectivity in their news choice than ever before in the current media landscape where news sources and channels proliferate (Bennett & Iyengar, 2008), and they often retransmit news stories to their social networks via e-mail and social media (Southwell, 2013). Then, what drives news diffusion in today's media environment? Why do certain news articles diffuse widely by triggering audience selection and social sharing, while others do not?

Previous research has identified social-psychological factors affecting audience selective exposure to and social flow of media content, such as confirmation bias (Knobloch-Westerwick, 2015) and social contagion (Rogers, 2003). This study aims to advance this line of research by addressing the following theoretical and empirical issues. First, as with a growing body of research (e.g., Berger & Milkman, 2012; Knobloch-Westerwick & Sarge, 2013), this study tests how message features shape what media content people choose to

consume and share, regardless of individual differences. Second, while content selection and sharing are sequentially connected communication behaviors, they have rarely been examined together. Third, little research has investigated actual content selection and retransmission *count* data observed in a natural setting as behavioral outcomes. Fourth, few studies have tested how retransmission channels such as e-mail and social media affect the kind of media content people share with their social networks.

Within the context of the online diffusion of New York Times (NYT) health news articles, the present study examines how message features affect the extent to which the articles (1) attract audience selection and (2) go ‘viral’ by inviting social sharing via email and social media (Facebook & Twitter). This study also tests how retransmission channels (e-mail vs. social media) shape news virality. Using both computational social science approaches (Lazer et al., 2009) and more traditional methods, this study collects and analyzes actual behavioral data of audience selections and propagations of the NYT articles, as well as the articles’ content and context data.

### Message-Level Drivers of Audience News Selection and Retransmission

Diffusion of media content involves two audience communication behaviors: selection and retransmission (Cappella, Kim, & Albarracín, 2014; Kim, Lee, Cappella, Vera, & Emery, 2013). That is, media content is most likely to diffuse widely when it both attracts audience selection and prompts subsequent social sharing. The current study thus focuses on message-level factors that previous research has suggested shape both of these communication behaviors: *informational utility*, *content valence*, *emotional evocativeness*, *novelty*, and *exemplification*. Except for novelty and exemplification, this study tests two types of message variations for each factor (O’Keefe, 2003): (1) *intrinsic* features that are independent of audience perceptions or responses (e.g., content valence in terms of words used in articles) and (2) *perceived* or *effect-based* features (e.g., content valence in terms of audience responses toward articles).

#### Informational utility

Scholars have identified informational utility as a key predictor of audience message selection (Knobloch-Westerwick, 2015) and sharing (Berger, 2014). A meta-analysis of selective exposure research (Hart et al., 2009) reveals that while there is an overall tendency for people to prefer congenial over uncongenial messages, the opposite is true when uncongenial ones have higher utility. The idea that messages with high informational utility enjoy a retransmission advantage has also received empirical support. For example, Berger and Milkman (2012) found that news articles conveying practically useful information are more likely to be retransmitted. An intrinsic message feature that taps into the notion of informational utility particularly in health contexts is the presence of efficacy information (Cappella et al., 2014; Knobloch-Westerwick & Sarge, 2013). Efficacy information can be defined as information on effective means to achieve health-related goals such as promoting health and overcoming (or reducing) health threats (Bandura, 2004; Moriarty & Stryker, 2008), and such information is effective in shaping health behaviors (Witte & Allen, 2000), all of which imply the high practical value of efficacy information. In sum, the current study predicts that health news articles presenting efficacy information are more frequently viewed

(H1a) and shared (H1b). As a perceived message feature of informational utility, this study examines the role of an overall sense of perceived content usefulness (Berger & Milkman, 2012) with hypotheses that articles conveying more useful content trigger more selections (H2a) and retransmissions (H2b).

### **Content valence**

The valence of media content plays a significant role in audience message selection and sharing. People tend to be hardwired for negative information, and this negativity bias is well documented in selective exposure research (Knobloch-Westerwick, 2015). In contrast, positivity bias likely operates in deciding what to share (Berger & Milkman, 2012; Kim et al., 2013). Compared to message selection, sharing is a more social behavior, and might thus involve more complex considerations (Berger, 2014; Huang, Lin, & Lin, 2009) such as anticipated responses from recipients (e.g., altruistic or socializing motivations) and expected perceptions of recipients about sharers (e.g., self-enhancement motivations). Positive messages are more likely to be passed on because sharing such messages makes recipients feel good and helps build or maintain the sharers' positive images (Berger, 2013, 2014). All in all, the current study predicts negativity bias in selection and positivity bias in retransmission, focusing on the following three message features. As with previous research (Berger & Milkman, 2012), the present study evaluates content valence in terms of both effect-based and intrinsic message properties: (1) positivity of emotional responses induced by articles (H3a for selection and H3b for sharing) and (2) positivity of emotions expressed in articles (positive vs. negative emotion word use; H4a for selection and H4b for sharing), respectively. This study further investigates how (3) perceived content controversiality (negative valence; Chen & Berger, 2013; Zillmann, Chen, Knobloch, & Callison, 2004) impacts audience selection (H5a) and retransmission (H5b).

### **Emotional evocativeness**

Independent of content valence, emotional evocativeness may shape audience message selection and sharing behaviors. Emotionally arousing messages tend to foster selective exposure (Knobloch-Westerwick, 2015), such that people seek out news stories with emotionally evocative frames (Zillmann et al., 2004). Media messages characterized by high emotional evocativeness are also more likely to go viral (Berger, 2014). The experience of emotional arousal tends to prompt social sharing of that emotion because emotion sharing has both intrapersonal and interpersonal benefits such as sense-making of the emotional experience and establishment (or strengthening) of social bonds (Rimé, 2009). Empirical evidence for the role of emotional evocativeness in enhancing virality is also robust (Berger & Milkman, 2012). This study therefore hypothesizes that emotional evocativeness increases the extent to which health news articles trigger audience selections and retransmissions, focusing on both effect-based and intrinsic message characteristics: (1) emotional arousal induced by articles (H6a for selection and H6b for sharing) and (2) emotionality expressed in articles (use of either positive or negative emotion words; H7a for selection and H7b for sharing), respectively.

## Novelty

This study posits that media messages are frequently selected and shared when their content is perceived as novel. Novel, surprising, or unusual news may attract selection as such news likely disturbs people's routine information processing (or violates schema-driven expectations), and leads them to 'stop and think' or view it as potentially threatening information (Knobloch-Westerwick, 2015). An experimental study showed that news articles with deviant or unusual content foster selective exposure (J. H. Lee, 2008). Novelty may also increase virality because unusual or surprising content has, in general, high social currency and makes for good conversation material (Berger, 2013, 2014). Research demonstrates that people are more likely to propagate novel or surprising messages – including news articles (Berger & Milkman, 2012) and antismoking arguments (Kim et al., 2013). In sum, this study predicts that health news stories providing more novel content prompt more selections (H8a) and retransmissions (H8b).

## Exemplification

Messages crafted in narrative form may also have a retransmission advantage because (1) stories are a fundamental form of human cognition and communication, and easier to comprehend and recall (Schank & Abelson, 1995), and (2) they deliver information in a vivid and engaging manner (Berger, 2013). Exemplification is an intrinsic message feature that makes news more vivid, engaging, and thereby more story-like (Zillmann & Brosius, 2000). Exemplars in a news article are personal stories (or experiences) of people related to the subject of the article. While news is a highly structured and conventionalized form of narrative (van Dijk, 1988), presenting relevant exemplars further enhances its narrativity (Kim, Bigman, Leader, Lerman, & Cappella, 2012). As such, this study posits that exemplification boosts virality (H9).<sup>1</sup>

## News-Sharing Channels and Virality

This study investigates how effects of message features on virality differ by online news-retransmission channels of different audience size (Barasch & Berger, 2014; Berger, 2014), focusing on the comparison between e-mail (narrowcasting) and social media (Facebook & Twitter; broadcasting). E-mail- and social media-based propagations tend to assume different types of recipients. E-mail-based forwarding usually targets an audience that is relatively small and narrow, while recipients of social media-based sharing (e.g., Facebook friends or Twitter followers) tend to be relatively large and diverse. As sharers' consideration of recipients (e.g., the nature/strength of the sharer-recipient relationship, recipients' preference) plays an important role in deciding what to share (e.g., Huang et al., 2009), it seems reasonable to expect that news-sharing channels varying in their target audience affect what goes viral by activating different motivations of the sharers (Berger, 2014). However, not enough empirical evidence has been assembled to allow specific predictions about the channel effect (cf. Barasch & Berger, 2014). Thus, this study poses a

---

<sup>1</sup>Exemplification may also affect selective exposure (e.g., Knobloch-Westerwick & Sarge, 2013). However, it is not tested here because exemplars are present in only a few teasers of the study sample (i.e., textual unit used when predicting news selections; see the Method section).

research question about the role of retransmission channels (e-mail vs. social media) in shaping the relationships between message features and news virality (RQ1).

## Method

This study used data on 760 NYT health news articles published online between July 11, 2012 and February 28, 2013 (about 7 ½ months; 33 weeks).<sup>2</sup> The unit of analysis is the article teaser (title + abstract) and the full text for selection and retransmission analyses, respectively.<sup>3</sup> The data were collected via three methodological tools: (1) *machine-based data mining* for selection and retransmission data, article metadata, and context information; (2) *content analysis* using both human and computerized coding methods for intrinsic message features; (3) *message evaluation survey* for perceived (or effect-based) message features. In what follows, details of each data collection method and associated variables are described. Table 1 shows the article-related data obtained through these tools. All the dependent, independent, and control variables of this study are measured using the data shown in Table 1. For the sake of brevity, however, this paper does not fully describe all control variables (see also the Analysis section).

### Machine-Based Data Mining: News Diffusion Tracker

In order to collect news diffusion-related data in an automated manner, the present study employed a ‘big data’ method as used in computational social science (Lazer et al., 2009). An automated software application, the *News Diffusion Tracker* (NDT), was developed to conduct two real-time data-mining tasks simultaneously: (1) importing data through the NYT’s Most Popular API (application programming interface), and (2) crawling the main page of the NYT website’s Health section. NDT was written in JavaScript and run on MS SQL server.

First, NDT fetched selection and sharing data via the API every 15 minutes. The API provided these data for articles that were published online *no earlier than 30 days* prior to the time of the request from NDT. The API returned data about the frequency with which an article had been viewed and shared (via e-mail, Facebook, Twitter; separately for each channel) by NYTimes.com readers *in the last 24 hours* as of the time of the request from NDT. Total selection count was obtained by summing 30 days of viewing data (i.e., every 24 hours) for each article. Total retransmission count was calculated by summing 30-day data for each article by platform (e-mail, Facebook, Twitter). The selection and retransmission variables followed a lognormal distribution (all *p*-values from the Shapiro-Wilk tests > .87), and thus were natural-log-transformed.<sup>4</sup> The average logged total selection count was 9.95 (*SD* = 1.44). The logged total sharing count was obtained by taking the logarithm of the sum of three retransmission items (e-mail, Facebook, Twitter;  $\alpha = .95$ ,  $M = 5.86$ ,  $SD = 1.48$ ). For RQ1, this study used the logged e-mail-sharing ( $M = 5.34$ ,  $SD =$

<sup>2</sup>Health news articles are defined as those published in the Health section of the NYT website. All health news articles published online during the 33-week period are included, except for the following: (1) articles from news agencies (e.g., AP), (2) articles in the *Recipes for Health* series, (3) interactive articles (e.g., *Well Quiz*), (4) obituaries, and (5) multimedia-based articles. This exclusion was made to ensure that articles are comparable in their content-type and format.

<sup>3</sup>An article’s abstract is not a part of its full text but an independent summary of the full text.

<sup>4</sup>Throughout this paper, all logarithmic transformations were conducted using natural logarithm.

1.57) and logged social media-sharing (logged summative scale of the Facebook- and Twitter-sharing data;  $M = 4.82$ ,  $SD = 1.46$ ) variables.<sup>5</sup> The NYT API also provided article metadata such as title, abstract, column (category; assigned by NYT), URL, and image URL(s) in each article. The article URL information was then used to extract (via a HTML parser) article full text and online publication timestamp (month and day of the week).

Second, NDT's built-in web crawler visited the main page of the Health section of the NYT website every 15 minutes, concurrently to the data mining via the NYT API. Specifically, NDT collected a list of articles shown in prominent locations at every visit (top six positions in the upper-left-hand corner of the page; editorial cue to news importance; Knobloch-Westerwick, 2015). The total number of hours that articles appeared in the prominent locations was obtained by summing 30 days of data (every 24 hours). The variable showed a lognormal distribution (the  $p$ -value from the Shapiro-Wilk test  $> .84$ ), and thus were log-transformed ( $M = 2.05$ ,  $SD = 1.52$ ).

### Content Analysis

This study content-analyzed *intrinsic* message features of the 760 articles (i.e., message variations that are independent of audience perceptions or responses; O'Keefe, 2003).

**Human coding**—Article teasers were coded in terms of (1) the presence of efficacy information, (2) the mention of diseases or bad health condition, and (3) the mention of expert sources. Content-coding was done separately for titles and abstracts. Each of the title- and abstract-coding tasks was performed by two research assistants. For each task, 90 cases were randomly drawn from the full news sample and used as reliability data (Krippendorff, 2013). Intercoder reliability estimates (Krippendorff's  $\alpha$ s) ranged from .77 to .94 ( $M = .83$ ). A random half of the rest of the full sample was assigned to each coder. Efficacy information was coded to be present if a title (abstract) addressed one or more ways to promote health and wellbeing (or remain healthy) or to overcome (or avoid) a health risk/threat (Moriarty & Stryker, 2008). The coders also judged if there was any mention of one or more diseases (or bad health conditions) such as cancer and flu. The content-coded variations in titles and abstracts were then combined to construct teaser-level variables. Of the 760 articles, 19.7% presented efficacy information and 56.8% mentioned diseases or bad health conditions in their teasers.

Article full texts were also coded by two trained research assistants. The coders assessed efficacy information, exemplification, factual/evaluative statements by experts, topical area, and writing style. Reliability data consisted of 80 cases that were randomly selected from the full news sample. Final intercoder reliability estimates (Krippendorff's  $\alpha$ s) ranged from .77 to 1.00 ( $M = .89$ ). Each coder then assessed a random half of the rest of the full sample. The coders judged the presence of efficacy information in an article full text (present in 24.7% of

---

<sup>5</sup>The NYT API covers Facebook- and Twitter-based news sharing behaviors that take place on the NYT website. As an external validity check for the NYT API measures, this study collected Facebook and Twitter data using social media APIs (Facebook API & Topsy API, respectively) which keep track of a wider range of NYT article-sharing behaviors than the NYT API. The news-sharing data from the NYT API were highly correlated with those from the social media APIs:  $r = .89$  (Facebook);  $.83$  (Twitter);  $.92$  (social media; Facebook & Twitter); all  $p$ -values  $< .001$ . Details about the social media API methods are available upon request from the author.

the 760 articles). The coders recorded the presence of exemplification – a discussion (or mention) of a narrative (personal case/experience) of a person or family that is related to the subject of a given news article (Zillmann & Brosius, 2000) – in an article full text (present in 27.2% of the articles).

**Computerized coding**—LIWC 2007 was used for computerized coding of article teasers and full texts at the word level. LIWC counts words that belong to psychologically meaningful categories as defined by its own dictionary which is developed based on human judgment of word categories (for details about its reliability and validity, see Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007). Computerized coding was conducted separately for article teasers and full texts. The LIWC 2007 lexicon covered a broad range of words used in article teasers and full texts, with high average word-coverage rates (80.4% and 80.7%, respectively). This study focused on the following word categories of LIWC 2007: positive emotion (e.g., good, happy), negative emotion (e.g., bad, fear), word count, words longer than six letters (writing complexity), and words related to death, health, and social processes. The 760 teasers had on average 33.26 words ( $SD = 7.42$ ). Given this small word count, LIWC-measured variables were analyzed in terms of the *number* of words rather than LIWC's default metric, the *percentage* of words (proportion data tend to be unreliable when denominators are small). Based on a previous study (Berger & Milkman, 2012), expressed emotional positivity was measured by the word count difference in positive and negative emotion words ( $M = -.13$ ,  $SD = 1.66$ ), while expressed emotional evocativeness was defined as the sum of positive and negative words (log-transformed because of its distribution;  $M = .88$ ,  $SD = .56$ ). The average word count of the 760 full texts was 796.29 ( $SD = 385.15$ ). Given this substantial word count, the percentage metric was used. The mean of expressed emotional positivity (% positive emotion words – % negative emotion words) was .12 ( $SD = 1.73$ ), and that of expressed emotional evocativeness (% positive emotion words + % negative emotion words) was 3.88 ( $SD = 1.53$ ). This study also employed a HTML parser to count the number of hyperlinks embedded in each article full text using article URL data.

### Message Evaluation Survey

*Perceived* or *effect-based* message features (O'Keefe, 2003) were measured by a message evaluation survey where respondents read and rated article teasers and full texts on the Internet. The goal of this survey was to crowd-source evaluations of perceived content properties for each article (e.g., perceived usefulness) by aggregating assessments from multiple respondents who read the same article. Survey respondents were recruited through Amazon's Mechanical Turk (MTurk; for details about the validity of studies using MTurk samples, see Berinsky, Huber, & Lenz, 2012). A total of 5,092 U.S. adults participated in the survey (aged 18 to 80 years;  $M = 33$ ,  $SD = 11$ ). Of the respondents, 51.0% were female, 76.6% were non-Hispanic White, 69.5% were currently employed, and 51.3% completed some college or more education. During the survey, each participant read and rated six pieces of article texts (three teasers and three full texts) that were randomly selected from the entire sample of NYT health news articles.<sup>6</sup> The survey generated 15,276 ( $= 5,092 \times 3$ ) message assessments for each type of article text. The average number of respondents per article was 20.1 for both teasers ( $SD = 4.5$ ) and full texts ( $SD = 4.4$ ).

Respondents answered a series of questions for each article text. The same questions were asked for full texts and teasers, with the exception of a minor variation in the wording that referred to the type of article text (i.e., “article” vs. “article teaser”). Emotion-related items were assessed on a 5-point scale ranging from *not at all* (= 1) to *extremely* (= 5), while other items were rated on a 5-point scale ranging from *strongly disagree* (= 1) to *strongly agree* (= 5). For each rating item, respondents’ evaluations were averaged across the respondents by article.

Respondents were presented with eight emotion words (pride, amusement, contentment, hope, anger, fear, sadness, surprise; Lazarus, 1991) and asked: “How much does each of the following words describe how you felt while reading the article [article teaser]?” An emotional positivity scale was created by averaging these items, with the exception of the “surprise” item (the anger, fear, and sadness items were reverse-scored):  $\alpha = .87$ ,  $M = 2.78$ ,  $SD = .38$  for teasers;  $\alpha = .87$ ,  $M = 2.80$ ,  $SD = .43$  for full texts. Emotional evocativeness was assessed with a single item. Respondents answered how much the article [article teaser] they read made them feel “aroused”:  $M = 1.46$ ,  $SD = .22$  for teasers;  $M = 1.50$ ,  $SD = .23$  for full texts. To measure novelty, in addition to the “surprise” item mentioned above, respondents were asked to indicate how strongly they agreed that the information in the article [article teaser] was “new” and “unusual” (Turner-McGrievy, Kalyanaraman, & Campbell, 2013). A novelty scale was obtained by averaging the three items:  $\alpha = .85$ ,  $M = 2.77$ ,  $SD = .42$  for teasers;  $\alpha = .84$ ,  $M = 2.91$ ,  $SD = .39$  for full texts. Respondents indicated how strongly they agreed that the information in the article [article teaser] was “controversial” (Chen & Berger, 2013) and “useful” (Berger & Milkman, 2012). The mean of controversiality was 2.93 ( $SD = .57$ ) for teasers and 2.95 ( $SD = .59$ ) for full texts, and that of usefulness was 3.43 ( $SD = .44$ ) for teasers and 3.84 ( $SD = .34$ ) for full texts.

## Analysis

**Overview**—This study analyzes actual online diffusion data of NYT health news articles and associated article content and context data. Given the observational nature of this study, it is crucial to control for potential confounders in order to obtain *unbiased* estimates of message effects on news selection and retransmission. Specifically, this study included selection count as a covariate when predicting sharing count. Since greater exposure to an article can lead to an increase in the frequency of sharing the article (i.e., simply having more opportunity to be shared), the sheer number of times that the article has been shared is confounded by the number of times that it has been viewed. Thus, it is essential for observational studies like this one to disentangle the likelihood of sharing from the likelihood of viewing by statistically controlling for selection count when predicting retransmission count. That is, news virality in this study refers to the extent to which an article gets shared by people who consume it (i.e., retransmission given selection). This study also included as a covariate the total amount of time that articles were shown in

---

<sup>6</sup>The survey was programmed to sample six *different* articles to ensure that no respondent would evaluate a full text and a teaser of the same article. The survey consisted of three sections; in each section, the respondents evaluated one full text and one teaser text. Of the 760 articles, one article was mistakenly excluded from the sampling pool due to an unexpected programming error. Consequently, a total of 759 articles were evaluated by the respondents in this survey.



prominent locations on the main page of the NYT website's Health section to control for effects of an editorial cue to news importance (Knobloch-Westerwick, 2015).

In addition, the current study included the following control variables (see also Table 1): (1) seasonal or periodic variations (online publication month and day of the week); (2) basic linguistic features (word-count and writing complexity); (3) disease-specific expressions in teasers (mention of diseases or bad health conditions) and word-level variations in other broadly health-related dimensions (words related to death, health, and social processes); (4) message variations related to content expertise (mention of expert sources [teaser] and factual or evaluative statements by expert sources [full text]); (5) article category assigned by NYT (e.g., Well) and topical area (full text; e.g., diseases and health conditions); and (6) other format and stylistic features of full texts (writing style [written in a first-person point of view vs. not], number of hyperlinks, and presence of images). For the sake of brevity, further details about the measures and results regarding these six categories of control variables are not reported in this paper, except when necessary. Full information is available upon request from the author.

**Statistical modeling**—Linear regression models were estimated using the ordinary least squares (OLS) method to examine hypotheses related to message effects on news selection and sharing (H1a to H9). Specifically, the logged total selection count was regressed on (1) *teaser-level content factors* ( $df = 18$ ) including central message features of this study (i.e., variables related to informational utility, content valence, novelty) and other control variables; and (2) *context factors* ( $df = 10$ ) including an editorial cue to news importance (i.e., total hours shown in prominent locations), and article publication month and day of the week. For the virality-related hypotheses, the logged total retransmission count was regressed on (1) *full text-level content factors* ( $df = 25$ ) including central message features (i.e., variables related to informational utility, content valence, novelty, exemplification) and other controls; (2) *context factors* ( $df = 10$ ; identical to the news-selection model described above); and (3) the logged total selection count.

Structural equation modeling (SEM) with the full information maximum likelihood estimation method was used to examine RQ1 about how message features differentially affect (1) news retransmissions via e-mail (logged;  $DV_1$ ) and (2) those via social media (logged;  $DV_2$ ). As shown in Figure 1, a structural model was specified as follows: (1)  $DV_1$  and  $DV_2$  are regressed on the same predictors as those for the total news-retransmission model; (2) the predictors are correlated with each other; and (3) the residuals of  $DV_1$  and  $DV_2$  are correlated (i.e., the partial correlation between  $DV_1$  and  $DV_2$ , controlling for the common predictors). This specification makes the structural model just-identified (i.e., model fit is perfect). SEM was preferred over the estimation of separate OLS regression models for  $DV_1$  and  $DV_2$  because SEM uses the full covariance matrix and enables the *statistical* comparison of the effects of message features on the two dependent variables (e.g.,  $b_{11}$  vs.  $b_{21}$  in Figure 1), which is central to answering RQ1.

## Results

Table 2 presents zero-order correlations among the central variables of this study, and Table 3 summarizes the results from bivariate and multiple OLS regression models of total news selections and retransmissions. Consistent with H1a, health news articles that presented efficacy information in their teasers were more frequently selected by NYTimes.com readers than those without such information, unstandardized  $b = .34$ , 95% CI [.09, .59]. Perceived usefulness was unrelated to selection, rejecting H2a. H3a which predicted a positive link between selection and the negativity of emotional responses induced by teasers was rejected. Rather, the relationship was moderated by the mention of diseases or bad health conditions,  $b = -.85$ , 95% CI [-1.34, -.36]. Articles whose teasers evoked more positive emotional responses invited more selections when there was no mention of diseases or bad health conditions in the teasers,  $b = .65$ , 95% CI [.22, 1.07]. When such terms were mentioned, emotional valence was unrelated to selection,  $b = -.20$ , 95% CI [-.54, .14]. Regarding this interaction effect, it is noteworthy that the mention of such terms was negatively related to selection overall:  $b$  for its simple main-effect term (i.e., when the emotional valence variable was held at its mean) was  $-.27$ , 95% CI [-.48, -.06]. Its effect without the interaction term was also significantly negative,  $b = -.30$ , 95% CI [-.51, -.09]. Expressed emotional valence was unrelated to selection, rejecting H4a. A significant negativity bias effect was found for perceived controversiality (which, as expected, was negatively related to the positivity of emotional responses; see Table 2). Articles with more controversial teasers were more frequently selected,  $b = .25$ , 95% CI [.06, .43], supporting H5a. Emotional arousal evoked by teasers was unrelated to selection, rejecting H6a. Consistent with H7a, expressed emotional evocativeness was positively associated with selection,  $b = .16$ , 95% CI [.003, .32]. In contrast to H8a, there was a negative relationship between perceived novelty and selection,  $b = -.23$ , 95% CI [-.47, -.001]. An editorial cue to news importance had a significant effect, such that the longer articles were displayed in prominent locations on the main page of the NYT website's Health section, the more frequently they were selected,  $b = .44$ , 95% CI [.37, .51].

Health news articles presenting efficacy information in their full texts triggered more frequent retransmissions (via e-mail, Facebook, and Twitter) than those without such information,  $b = .13$ , 95% CI [.02, .24], providing support for H1b. Consistent with H2b and H3b, articles were more frequently shared when they provided more useful and positive content,  $b = .50$ , 95% CI [.36, .64],  $b = .19$ , 95% CI [.07, .32], respectively. Virality was unrelated to the following variables: expressed emotional valence, perceived controversiality, induced and expressed emotional evocativeness, perceived novelty, and exemplification. Thus, H4b, H5b, H6b, H7b, H8b, and H9 were rejected. News retransmission was positively related to (1) an editorial cue to news importance (i.e., the duration shown in prominent locations) and (2) news selection,  $b = .04$ , 95% CI [.003, .08],  $b = .84$ , 95% CI [.80, .88], respectively.

Table 3 presents SEM results. The residual correlation between the two dependent variables (i.e.,  $r$  in Figure 1) was .44,  $p < .001$ . An omnibus test of the null hypothesis that all coefficients are identical between the two regression equations (i.e., retransmissions via e-mail and those via social media; see Figure 1) was significant,  $\chi^2(36) = 295.23$ ,  $p < .001$ .

Focusing on the coefficients for the central message features of this study, the omnibus test was also significant,  $\chi^2(9) = 73.62, p < .001$ . Specifically, the presence of efficacy information invited more email-based retransmissions,  $b = .19, 95\% \text{ CI } [.07, .32]$ , but it was unrelated to those via social media,  $b = .002, 95\% \text{ CI } [-.11, .12]$ . The coefficient difference was significant,  $\chi^2(1) = 8.30, p < .01$ . The effect of perceived usefulness was stronger on e-mail-specific virality,  $b = .66, 95\% \text{ CI } [.49, .82]$ , than on social media-specific virality,  $b = .22, 95\% \text{ CI } [.07, .37], \chi^2(1) = 26.59, p < .001$ . Induced emotional evocativeness was positively related to retransmissions via social media,  $b = .34, 95\% \text{ CI } [.13, .55]$ , but it was unrelated to those via email,  $b = -.02, 95\% \text{ CI } [-.24, .21]$ , with the coefficient difference being significant,  $\chi^2(1) = 9.33, p < .01$ . Perceived novelty had an *opposite* effect,  $\chi^2(1) = 19.60, p < .001$ , such that it was positively associated with e-mail-based sharing,  $b = .17, 95\% \text{ CI } [.03, .31]$ , but negatively related to social media-based sharing,  $b = -.16, 95\% \text{ CI } [-.29, -.03]$ . The difference between exemplification effects on the two virality outcomes was marginally significant,  $\chi^2(1) = 3.44, p = .06$ . While the presence of exemplars was unrelated to e-mail-specific virality,  $b = -.005, 95\% \text{ CI } [-.13, .12]$ , it triggered more frequent social media-based retransmissions,  $b = .12, 95\% \text{ CI } [.00002, .23]$ . Effects of the following message features were not different between the two regression equations: induced and expressed emotional positivity, perceived controversiality, expressed emotional evocativeness.

## Discussion

Identifying factors that drive social epidemics of news coverage is essential to our understanding of its impact on audience cognitions, emotions, and behaviors in the new public communication environment because exposure is the *sine qua non* of message effects (Hornik, 2002). By examining behavioral data on the online diffusion of NYT health news articles in relation to the articles' content and context data, this study identifies message-level drivers of audience news selection and sharing behaviors, and sheds light on how online retransmission channels (e-mail vs. social media) shape what news goes viral.

The results indicate support for the notion that informational utility drives what health news people choose to read and retransmit afterwards (Berger & Milkman, 2012; Knobloch-Westerwick, 2015). Health news stories presenting efficacy information – an intrinsic content feature that enhances the persuasiveness of health messages (Witte & Allen, 2000) – were more frequently viewed and shared, and those conveying more useful content were also more viral. The results also showed that articles using more emotion words in their teasers invited more selections, which is largely consistent with previous findings (e.g., Zillmann et al., 2004).

Unlike the prediction of this study that negativity bias operates in news selection and positivity bias drives sharing, the results overall suggest that positivity looms larger in deciding both what to read and what to share. As with previous research (Berger & Milkman, 2012; Kim et al., 2013), articles evoking more positive emotional responses were more viral. But perceived controversiality was unrelated to virality. The nonsignificant link might be explained by a recent study finding (Chen & Berger, 2013), although the study tested conversation likelihood as a final outcome. Chen and Berger (2013) found that

controversial content produces both interest and discomfort simultaneously, especially when personal identity is disclosed, as it was in this study's case (i.e., news sharing via e-mail and social media reveals personal identity). Thus, it is plausible that the two countervailing psychological states evoked by controversial articles led to the observed null effect of perceived controversiality on news sharing. Future research might test psychological factors that mediate or moderate the controversiality-virality relationship.

Articles were more frequently selected when their teasers did not mention diseases or bad health conditions. As teasers without terms related to diseases or unhealthy statuses evoked more positive emotional responses than those using such terms (see Table 2), this finding can be interpreted as showing that positivity bias, rather than negativity bias, operates in news selection. This interpretation is further supported by the interaction effect that positive articles were more frequently selected when their teasers did not include such terms. That is, induced emotional positivity prompted selections for teasers exhibiting positivity in terms of another dimension of content valence (i.e., absence of disease-related terms). The observed positivity bias is at odds with past findings that negativity attracts selection. One reason for this inconsistency might be the difference in topical domains chosen for theory testing. Many message stimuli used in past studies were about politics (Meffert, Chung, Joiner, Waks, & Garst, 2006) and crimes/accidents (Knobloch, Hastall, Zillmann, & Callison, 2003), while this study focused exclusively on health news. Compared to news about politics, crimes, and accidents, health news might be more self-focused and more directly linked to individual well-being. This seems to be particularly true for the articles tested in this study; about 68.8% of them addressed diseases and health conditions. People may avoid negative news stories if they cover such self-oriented health topics. This reasoning is in line with a recent finding that smokers are more likely to choose tobacco control messages evoking positive feelings (Kim et al., 2013). Perceived controversiality was the only negativity-related feature that was positively related to news selection, which is consistent with prior research (Zillmann et al., 2004). In conclusion, the results suggest that it is controversiality (a specific component of negativity), rather than overall negativity, that attracts news selection.

The results revealed a negative association between perceived novelty and audience news selection, which runs counter to previous findings. As with the case of the valence-selection link, topical difference and associated psychological factors might explain the discrepancy between the present and past results. Compared to previous studies which tended to focus on other-oriented news topics such as crime (J. H. Lee, 2008), this study tested messages conveying more self-oriented topics – health news articles more than two thirds of which were about diseases and health conditions. People may choose familiar health information in defense of certainty, rather than unusual or surprising one that is potentially threatening, because the information address self-focused issues such as diseases and health conditions. In contrast, people may still seek out unusual or surprising messages because such messages tend to be remarkable and interesting (Berger, 2013; Silvia, 2008), but only if the messages are about relatively other-focused topics (e.g., crimes). Likewise, a finding that the more often people perform a health behavior, the more likely they seek out messages encouraging, rather than opposing, the behavior (Knobloch-Westerwick, Johnson, & Westerwick, 2013) can also be interpreted as showing the preference for familiar over novel content in health

contexts, assuming that people's behavioral frequency is positively related to their perceived familiarity of the behavior-promoting messages, and vice versa for the behavior-challenging ones. Another possible explanation for the novelty-selection link observed in the present study concerns the persuasiveness of teasers. It may be that articles with novel teasers invite fewer selections as novelty in this context undermines persuasiveness (which is positively related to the likelihood of attracting selections; Kim et al., 2013). People may consider novel health content unpersuasive when it is embedded in short texts like teasers as there is little room to convey supporting reasons or evidence. Future work might examine psychological mechanisms that underlie the negative relationship between the content novelty and audience health news selection, and how they operate differentially across topical domains.

The results suggest that online news-sharing channels such as e-mail (narrowcasting) and social media (broadcasting) significantly affect what news goes viral. Message features related to informational utility were more closely tied to news retransmissions via e-mail than those via social media. On the other hand, induced emotional evocativeness played a larger role in social media-specific virality. The findings are overall consistent with recent theorizing and empirical evidence (Barasch & Berger, 2014) that narrowcasting triggers social sharing of useful content by activating other-focus (i.e., recipients), while broadcasting ignites social propagation of self-enhancing content (e.g., emotionally arousing content) by boosting self-focus (i.e., sharer).

Perceived novelty played an opposite role in e-mail- and social media-based news sharing, which resulted in the pattern that novelty was unrelated to the total number of retransmissions. Consistent with prior research (Berger & Milkman, 2012; Kim et al., 2013), articles presenting more novel content were more frequently shared via email; but novelty was negatively related to social media-based propagations. This might also be due to the difference in the target audience of the two retransmission channels. Sharing health news that is (1) unusual or surprising and (2) closely tied to individual well-being with large and diverse audience members via social media (i.e., broadcasting) might be considered detrimental to enhancing a positive self-view (or at least unclear as to whether it would be helpful to self-enhancement) because doing so could annoy or offend someone in the sharer's social network. On the other hand, e-mail tends to assume a smaller and narrower audience than social media. Moreover, people specify recipients when they use e-mail to forward news, while it is much less common (albeit possible) to do so on social media. People might thus feel safer to share unusual or surprising content (which is remarkable and interesting in general; Berger, 2013; Silvia, 2008) via e-mail because they can narrowcast to particular recipients who they think would like it. That is, when it comes to health news, it appears to be e-mail – rather than social media – that ensures high virality of novel information.

As with novelty, exemplification was not predictive of total retransmissions; instead, its impact differed by news-sharing channels. Articles presenting exemplars were more frequently shared via social media, while exemplification was unrelated to e-mail-based propagations. This retransmission-channel difference might also be explained by the aforementioned psychological tendency for news sharers to focus more on themselves than

recipients when deciding what to share through social media (broadcasting), compared to when using email (narrowcasting). Exemplification might boost social media-based sharing because story-like messages have high social currency when people communicate with a large audience (i.e., self-enhancing content; Berger, 2013, 2014), but not necessarily so when assuming a smaller and narrower audience.

In sum, the results underscore the significant role of news-sharing channels in shaping the relationships between message features and virality. While this study offered some explanations as to why e-mail- and social media-based propagations make a difference in what goes viral, they are speculative rather than empirically grounded, due to the lack of data on the social psychology of such effects. Thus, more research is warranted to test psychological mechanisms that underlie the retransmission-channel effects in health contexts, including the role of narrowcasting- and broadcasting-related news-sharing motivations (Barasch & Berger, 2014; Berger, 2014).

The results regarding content features and an editorial cue to news importance have practical implications for web-based public health communication campaigns where messages and their positions on a webpage are determined a priori. Specifically, the results can be used to quantify what consequences in audience selection and sharing would follow from manipulating the content and editorial factors. Based on multiple OLS regression results (Table 3), this study conducted an ancillary analysis to predict changes in audience selection (sharing via email and social media) in response to changes in focal message features (e.g., efficacy information) and an editorial cue to news importance (i.e., hours shown in prominent locations on the NYT Health section's main webpage) that were found to significantly affect selection (sharing). The analysis predicted articles' selection (sharing) count in the following three cases: (1) when focal message features are *weak* and the editorial cue to news importance is *low*, while other predictors are held constant (*Baseline*); (2) when focal message features are *strong* while everything else remains the same as for the Baseline case (*Message Effects*); (3) when the editorial cue is *high* while other features are identical to the Message Effects case (*Message & Editorial Effects*). When predicting sharing count, the analysis included indirect effects of the message and editorial factors that are mediated through selection count, in addition to their direct effects.<sup>7</sup>

Figure 2 shows the results from this analysis. Articles with strong message features are predicted to invite about 4.3 times more selections and 6.3 times more propagations than

<sup>7</sup>For news selections, article teasers with *strong* (vs. *weak*) message features are those (1) with (vs. without) efficacy information; (2) at one standard deviation above (vs. below) the mean ( $M + SD$  vs.  $M - SD$ ) of induced emotional positivity; (3) without (vs. with) the mention of diseases or bad health conditions; (4) at  $M + SD$  (vs.  $M - SD$ ) of perceived controversiality; (5) at  $M + SD$  (vs.  $M - SD$ ) of logged expressed emotional evocativeness; (6) at  $M - SD$  (vs.  $M + SD$ ) of perceived novelty. For news retransmissions, article full texts with *strong* (vs. *weak*) message features are those (1) with (vs. without) efficacy information; (2) at  $M + SD$  (vs.  $M - SD$ ) of perceived usefulness; (3) at  $M + SD$  (vs.  $M - SD$ ) of induced emotional positivity; (4) at  $M + 1.45$  (vs.  $M$ ) of logged selections (indirect effects; 1.45 = predicted difference in logged selections between teasers with *strong* and *weak* message features). Articles with a *high* (vs. *low*) editorial cue to news importance are those (1) at  $M + SD$  (vs.  $M - SD$ ) of logged hours shown in prominent locations for both selections and retransmissions, and (2) at  $M + 1.34$  (vs.  $M$ ) of logged selections for indirect effects on retransmissions (1.34 = predicted difference in logged selections between *high* and *low* editorial cues; thus, the logged selection score for the *Message & Editorial Effects* case =  $M + 1.45 + 1.34$ ). Regression model-based predicted values (i.e., logged selection [retransmission] count;  $\log \hat{y}$ ) were back-transformed to obtain those in their original form (i.e., selection [retransmission] count;  $y$ ) using the following formula ( $\hat{\sigma}^2$  = mean squared error):  $\hat{y} = \exp(\hat{\sigma}^2/2) \exp(\log \hat{y})$

those with weak ones. Even further increases are predicted when articles with strong message features appear in prominent locations on the webpage for a longer time (about 3.8 times more selections and 3.5 times more propagations). Taken together, the combination of (1) crafting articles with strong message features and (2) displaying the articles in prominent locations is predicted to invite about 16.2 times more selections and 22.0 times more propagations, compared to when no such efforts are made. Figure 2 also reveals that the editorial factor (i.e., message placement) has stronger effects than the message design factor on audience selection and retransmission, which suggests that traditional news outlets and their journalistic judgments about news values still play a central role in the social flow of news in the current public communication environment. With respect to this, it should be noted that while both the message design and placement factors are controllable by, for example, health communication personnel in some circumstances (e.g., creating a health campaign website), the latter is uncontrollable in other contexts (e.g., writing a campaign-related press release). In either case, however, the results of this study as a whole suggest that message features exert significant and independent effects on what news people read and retransmit via email and social media over and above the message placement factor.

While this study reveals how message features and news-sharing channels affect health news diffusion, much more remains to be done to advance this line of research by addressing limitations of the current work. In addition to those already discussed, it should first be noted that this study analyzed NYT health news stories as a study sample. Thus, results reported here may not necessarily generalize to health news articles of other news outlets. While this study focused on NYT data because of their measurement quality (NYT was the only U.S. news outlet that enabled access to selection and sharing *count* data for each article at the time this study was conducted), future research might test the generalizability of the current findings using data from other news outlets. Second, this study did not manipulate key message features with random assignment, but measured them instead. Thus, despite the efforts to control for potential confounders, this study cannot conclusively rule out the possibility that a causal inference from the observed effects is spurious. Unmeasured content features such as open-ended information presentation (Southwell, 2013) might explain the observed message effects. Future research will need to conduct an experiment that manipulates message properties to test their causal impact on audience selection and retransmission in a clearer way. Third, this study treated intrinsic and perceived (or effect-based) message properties as parallel predictors. That is, this study only estimated the *direct* effects of intrinsic features (e.g., the presence of efficacy information) on news selection and sharing behaviors, although they can also have theoretically meaningful *indirect* effects by shaping perceived features (e.g., usefulness; see Table 2). Thus, the reported effects of intrinsic message features are likely underestimates of their total effects (direct + indirect paths). The parallel-predictors approach was nonetheless preferred because this study analyzed aggregate-level data where mediating paths are conceptually less clear than individual-level data and the study opted to conduct rigorous empirical tests for intrinsic message properties. Future research should further test individual-level pathways that flow from intrinsic content features to audience selection and retransmission via perceived features, which can ultimately advance our understanding of message effects on these communication behaviors (O'Keefe, 2003). Similarly, future work

might also examine how message features affect editorial decisions about article placement on websites (i.e., editorial cue to news importance; see Table 2 for the related correlation coefficients), which in turn, as this study shows, drive audience selection and sharing. Identifying message effects on such editorial decisions might illuminate an important mechanism through which content features shape news diffusion. Finally, future studies might examine consequences of message retransmissions (Southwell, 2013). This study focused on what drives sharers to propagate messages. But content features of retransmitted messages may also affect recipients, either independently or in conjunction with the nature (or strength) of sender-receiver relationship, especially in terms of persuasion (Cappella et al., 2014).

In conclusion, the present study advances our understanding of message features and communication channels that shape health news diffusion in the emerging media environment. This study also makes a methodological contribution by estimating message effects on the virality of news articles that are not confounded with the extent to which the articles attract audience selection. It should also be highlighted that the computational social science method developed in this study for automated data collection of behavioral measures of news selections and retransmissions holds promise for future research. It is hoped that future work will advance this line of research by further clarifying social psychological mechanisms through which message features and news-sharing channels drive health news diffusion.

## Acknowledgments

The author acknowledges the funding support of the National Cancer Institute at the National Institutes of Health (R01CA160226 and 5U01CA154254) and the Annenberg (Dissertation Research Fellowship) and Wharton (Russell Ackoff Doctoral Student Fellowship) Schools at the University of Pennsylvania. The content is solely the responsibility of the author and does not necessarily represent the views of the funding agencies. The author thanks Joseph Cappella, Michael Delli Carpini, Robert Hornik, and three anonymous reviewers for their insightful comments on earlier versions of this paper.

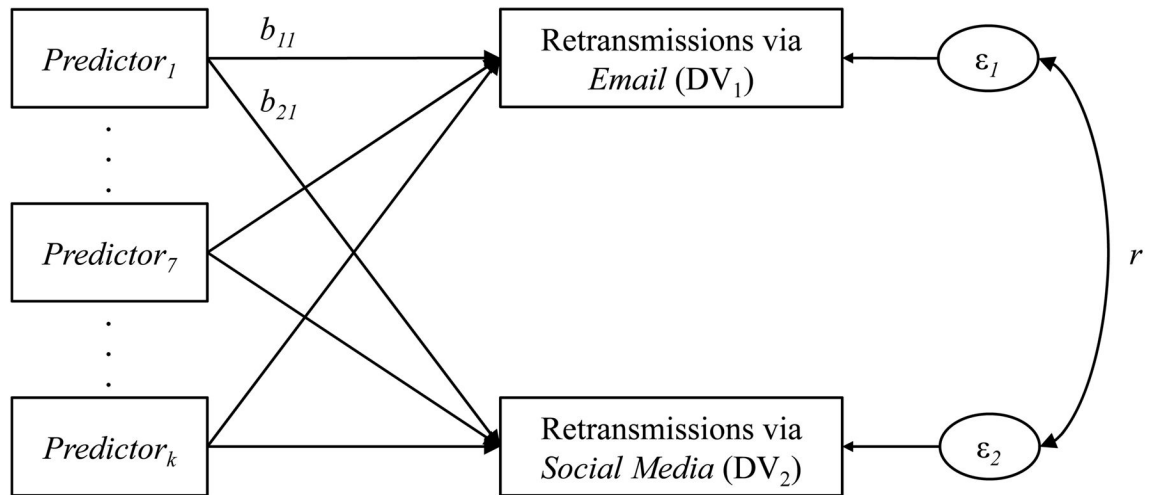
## References

- Bandura A. Health promotion by social cognitive means. *Health Education and Behavior*. 2004; 31:143–164.10.1177/1090198104263660 [PubMed: 15090118]
- Barasch A, Berger J. Broadcasting and narrowcasting: How audience size impacts what people share. *Journal of Marketing Research*. 2014; 51:286–299.10.1509/jmr.13.0238
- Bennett WL, Iyengar S. A new era of minimal effects? The changing foundations of political communication. *Journal of Communication*. 2008; 58:707–731.10.1111/j.1460-2466.2008.00410.x
- Berger, J. *Contagious: Why things catch on*. New York, NY: Simon & Schuster; 2013.
- Berger J. Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology*. 2014; 24:586–607.10.1016/j.jcps.2014.05.002
- Berger J, Milkman KL. What makes online content viral? *Journal of Marketing Research*. 2012; 49:192–205.10.1509/jmr.10.0353
- Berinsky AJ, Huber GA, Lenz GS. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*. 2012; 20:351–368.10.1093/pan/mpr057
- Cappella JN, Kim HS, Albarracín D. Selection and transmission processes for information in the emerging media environment: Psychological motives and message characteristics. *Media Psychology*. 2014 Advance online publication. 10.1080/15213269.2014.941112
- Chen Z, Berger J. When, why, and how controversy causes conversation. *Journal of Consumer Research*. 2013; 40:580–593.10.1086/671465

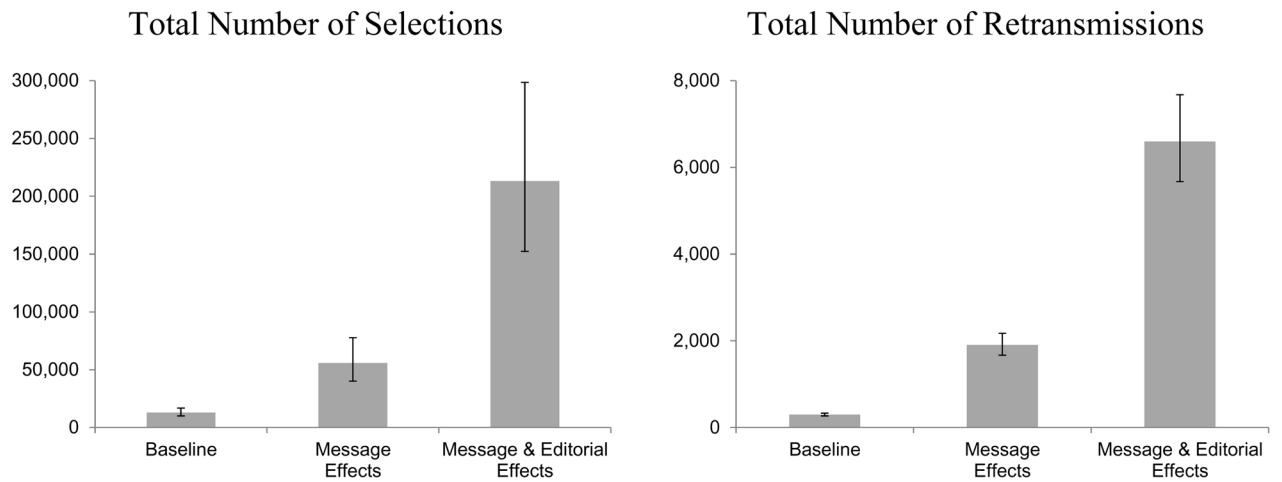


- Hart W, Albarracín D, Eagly AH, Brechan I, Lindberg MJ, Merrill L. Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*. 2009; 135:555–588.10.1037/a0015701 [PubMed: 19586162]
- Hornik RC. Exposure: Theory and evidence about all the ways it matters. *Social Marketing Quarterly*. 2002; 8:31–37.10.1080/15245000214135
- Huang CC, Lin TC, Lin KJ. Factors affecting pass-along email intentions (PAEIs): Integrating the social capital and social cognition theories. *Electronic Commerce Research and Applications*. 2009; 8:160–169.10.1016/j.elerap.2008.11.001
- Kim HS, Bigman CA, Leader AE, Lerman C, Cappella JN. Narrative health communication and behavior change: The influence of exemplars in the news on intention to quit smoking. *Journal of Communication*. 2012; 62:473–492.10.1111/j.1460-2466.2012.01644.x [PubMed: 22736808]
- Kim HS, Lee S, Cappella JN, Vera L, Emery S. Content characteristics driving the diffusion of antismoking messages: Implications for cancer prevention in the emerging public communication environment. *Journal of the National Cancer Institute Monographs*. 2013; 2013:182–187.10.1093/jncimonographs/igt018 [PubMed: 24395989]
- Knobloch S, Hastall MR, Zillmann D, Callison C. Imagery effects on the selective reading of Internet newsmagazines. *Communication Research*. 2003; 30:3–29.10.1177/0093650202239023
- Knobloch-Westerwick S. Choice and preference in media use: Advances in selective exposure theory and research. New York, NY: Routledge; 2015.
- Knobloch-Westerwick S, Johnson BK, Westerwick A. To your health: Self-regulation of health behavior through selective exposure to online health messages. *Journal of Communication*. 2013; 63:807–829.10.1111/jcom.12055
- Knobloch-Westerwick S, Sarge MA. Impacts of exemplification and efficacy as characteristics of an online weight-loss message on selective exposure and subsequent weight-loss behavior. *Communication Research*. 2013 Advance online publication. 10.1177/0093650213478440
- Krippendorff, K. Content analysis: An introduction to its methodology. 3. Thousand Oaks, CA: Sage; 2013.
- Lazarus, RS. Emotion and adaptation. New York, NY: Oxford University Press; 1991.
- Lazer D, Pentland A, Adamic L, Aral S, Barabási AL, Brewer D, Van Alstyne M. Computational social science. *Science*. 2009; 323:721–723.10.1126/science.1167742 [PubMed: 19197046]
- Lee JH. Effects of news deviance and personal involvement on audience story selection: A web-tracking analysis. *Journalism & Mass Communication Quarterly*. 2008; 85:41–60.10.1177/107769900808500104
- Meffert MF, Chung S, Joiner AJ, Waks L, Garst J. The effects of negativity and motivated information processing during a political campaign. *Journal of Communication*. 2006; 56:27–51.10.1111/j.1460-2466.2006.00003.x
- Moriarty CM, Stryker JE. Prevention and screening efficacy messages in newspaper accounts of cancer. *Health Education Research*. 2008; 23:487–498.10.1093/her/cyl163 [PubMed: 17289658]
- Napoli, PM. Audience evolution: New technologies and the transformation of media audiences. New York, NY: Columbia University Press; 2011.
- O’Keefe DJ. Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. *Communication Theory*. 2003; 13:251–274.10.1111/j.1468-2885.2003.tb00292.x
- Pennebaker, JW.; Chung, CK.; Ireland, ME.; Gonzales, A.; Booth, RJ. The development and psychometric properties of LIWC2007. Austin, TX: LIWC.net; 2007.
- Rimé B. Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review*. 2009; 1:60–85.10.1177/1754073908097189
- Rogers, EM. Diffusion of innovations. 5. New York, NY: Free Press; 2003.
- Schank, RC.; Abelson, RP. Knowledge and memory: The real story. In: Wyer, RS., editor. Knowledge and memory: The real story. *Advances in social cognition*. Vol. 8. Hillsdale, NJ: Laurence Erlbaum; 1995. p. 1-85.
- Silvia PJ. Interest – The curious emotion. *Current Directions in Psychological Science*. 2008; 17:57–60.10.1111/j.1467-8721.2008.00548.x

- Southwell, BG. Social networks and popular understanding of science and health: Sharing disparities. Baltimore, MD: Johns Hopkins University Press; 2013.
- Turner-McGrievy G, Kalyanaraman S, Campbell MK. Delivering health information via podcast or web: Media effects on psychosocial and physiological responses. *Health Communication*. 2013; 28:101–109.10.1080/10410236.2011.651709 [PubMed: 22420785]
- van Dijk, TA. *News as discourse*. Hillsdale, NJ: Lawrence Erlbaum; 1988.
- Witte K, Allen M. A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Education and Behavior*. 2000; 27:591–615.10.1177/109019810002700506 [PubMed: 11009129]
- Zillmann, D.; Brosius, H-B. *Exemplification in communication: The influence of case reports on the perception of issues*. Mahwah, NJ: Lawrence Erlbaum; 2000.
- Zillmann D, Chen L, Knobloch S, Callison C. Effects of lead framing on selective exposure to Internet news reports. *Communication Research*. 2004; 31:58–81.10.1177/0093650203260201



**Figure 1.** Message Effects on News Virality by Retransmission Channels Correlations among exogenous variables (i.e.,  $Predictor_1$  to  $Predictor_k$ ) are included in the model but not shown here for brevity.



**Figure 2.**

Combined Effects of Focal Message Features and an Editorial Cue to News Importance

Values in bar graphs represent the predicted total number of news selections (Left) and that of news retransmissions (Right) along with their 95% confidence intervals. An editorial cue to news importance for a given article refers to the logged total hours that the article appeared in prominent locations on the main page of the NYT website's Health section. Further details about the data for the graphs are described in Note 7.

Table 1

List of Article-Related Data by Data Collection Methods

Machine-Based Data Mining	Content Analysis	Message Evaluation Survey
<u>Diffusion Indicators</u> - <i>Real-Time (NYT API)</i> - <i>selection (viewing) count</i> - <i>sharing count</i> - <i>email, Facebook, Twitter</i>	<u>Human Coding</u> - presence of efficacy information - presence of exemplars (exemplification) <i>F</i> - disease-specific expressions <i>T</i> - mention of diseases or bad health conditions - message variations related to content expertise - mention of expert sources <i>T</i> - factual/evaluative statements by expert sources <i>F</i> - topical area <i>F</i> - (a) health policy and health care system, (b) diseases and health conditions, and (c) other - writing style <i>F</i> - written in a first-person point of view or not	<u>Survey Items</u> - <b>emotional valence</b> - pride, amusement, contentment, hope, anger, fear, sadness - <b>emotional evocativeness (arousal)</b> - <b>novelty</b> - newness, unusualness, surprise - <b>controversiality</b> - <b>usefulness</b>
<u>Content Metadata</u> - <i>Real-Time (NYT API)</i> - article category - e.g., Well, The New Old Age, etc. - presence of images <i>F</i> - image URL(s) - <i>Post Hoc(HTML Parser)</i> - seasonal/periodical variations - online publication month and day of the week <u>Computerized Coding</u> - <i>LJWC 2007</i> - <b>positive emotion words</b> - <b>negative emotion words</b> - basic linguistic features - word count - writing complexity (words > six letters) - word-level variations in health-related categories - words related to (a) death, (b) health, and (c) social processes - <i>HTML Parser</i> - number of hyperlinks <i>F</i>		

Note. Data used to create dependent variables are shown in **bold and italicized**. Data used to create focal independent variables are in **bold**. Other content metadata collected by machine-based data mining includes article title, abstract, full text, and URL. Further details about the data and data collection methods are available upon request from the author.

*T* = teaser only.  
*F* = full text only.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Correlation Matrices: Teaser and Full Text Data

<i>Teaser Data</i>	1	2	3	4	5	6	7	8	9	10	11
1. News Selections <sup>a</sup>	1.00										
2. Efficacy Info. Present	.10**	1.00									
3. Usefulness Perceived	-.002	.33***	1.00								
4. Positivity Induced	.07*	.36***	.12**	1.00							
5. Positivity Expressed	.02	-.03	-.09*	.28***	1.00						
6. Controversiality Perceived	.05	-.10**	-.01	-.39***	-.05	1.00					
7. Evocativeness Induced	.11**	.04	.13***	.07*	.03	.16***	1.00				
8. Evocativeness Expressed, <i>a</i>	.09*	.15***	-.001	.001	-.11**	.02	.01	1.00			
9. Novelty Perceived	-.06 <sup>+</sup>	.09*	.24***	.003	-.07*	.29***	.22***	-.09*	1.00		
10. Diseases/BHC Mentioned	-.09*	.16***	.24***	-.21***	-.34***	-.01	-.04	.08*	.24***	1.00	
11. Hours Shown in PL <sup>a</sup>	.51***	-.04	-.01	-.02	.03	.09**	.08*	-.03	.01	.004	1.00

<i>Full Text Data</i>	1	2	3	4	5	6	7	8	9	10	11
1. News Retransmissions <sup>a</sup>	1.00										
2. Efficacy Info. Present	.19***	1.00									
3. Usefulness Perceived	.23***	.28***	1.00								
4. Positivity Induced	.17***	.28***	.08*	1.00							
5. Positivity Expressed	.11**	-.05	-.08*	.30***	1.00						
6. Controversiality Perceived	.04	-.12***	-.06 <sup>+</sup>	-.42***	-.05	1.00					
7. Evocativeness Induced	.15***	.03	.06 <sup>+</sup>	.15***	.10**	.10**	1.00				
8. Evocativeness Expressed	.09*	.02	-.01	-.02	-.03	-.03	-.03	1.00			
9. Novelty Perceived	.09**	.14***	.20***	.09*	-.17***	.18***	.22***	-.15***	1.00		
10. Exemplification	.13***	-.06	-.03	-.08*	.02	.07 <sup>+</sup>	.15***	.12***	.11**	1.00	
11. Hours Shown in PL <sup>a</sup>	.49***	-.03	.01	.003	.09*	.07 <sup>+</sup>	.13***	-.09**	.04	.07 <sup>+</sup>	1.00
12. News Selections <sup>a</sup>	.89***	.13***	.10**	.11**	.09*	.03	.11**	.08*	.03	.11**	.51***

Note. Cell entries are Pearson's zero-order correlation coefficients – equivalent to (1) point-biserial correlation coefficients for the relationships between dichotomous and continuous variables (e.g., exemplification & retransmission) and (2) phi coefficients for those between dichotomous variables (e.g., exemplification & the presence of efficacy information). Other variables are not shown here for the sake of brevity. Full correlation matrices are available upon request from the author. "News Selections" and "Hours shown in PL"<sup>†</sup> are identical between the teaser and full-text matrices.

<sup>†</sup>Log-transformed. BHC = bad health conditions. PL = prominent locations.

+  $p < .10$ ,

\*  $p < .05$ ,

\*\*  $p < .01$ ,

\*\*\*  $p < .001$ .



**Table 3**

Predicting News Selections and Retransmissions

	Ordinary Least Squares Regression				Structural Equation Modeling		
	News Selections		News Retransmissions		Retransmissions by Channels		
	Bivariate	Multiple	Bivariate	Multiple	Email	Social Media	
Efficacy Information Present	.35** (.13)	.34** (.13)	.63*** (.12)	.13* (.06)	.19** (.06)	.002 (.06)	
Usefulness (Perceived)	-.01 (.12)	.02 (.11)	.99*** (.15)	.50*** (.07)	.66*** (.08)	.22** (.08)	
Emotional Positivity (Induced)	.29* (.14)	.65** (.22)	.57*** (.12)	.19** (.06)	.17* (.07)	.26*** (.07)	
Emotional Positivity (Expressed)	.02 (.03)	-.01 (.03)	.09** (.03)	.01 (.01)	.01 (.02)	.01 (.02)	
Controversiality (Perceived)	.14 (.09)	.25** (.09)	.11 (.09)	-.01 (.05)	-.03 (.06)	.06 (.05)	
Emotional Evocativeness (Induced)	.74** (.23)	.31 (.20)	.95*** (.23)	.10 (.10)	-.02 (.11)	.34** (.11)	
Emotional Evocativeness (Expressed)	.23* (.09)	.16* (.08)	.08* (.03)	.02 (.01)	.03 <sup>+</sup> (.02)	.02 (.02)	
Novelty (Perceived) Exemplification	-.22 <sup>+</sup> (.12)	-.23* (.12)	.35** (.14)	.05 (.06)	.17* (.07)	-.16* (.07)	
Diseases (or BHC) Mentioned	-.26* (.11)	-.27* (.11)	.44*** (.12)	.03 (.06)	-.005 (.06)	.12* (.06)	
Positivity (Induced) × Diseases		-.85*** (.25)					
Hours Shown in Prominent Locations	.48*** (.03)	.44*** (.03)	.48*** (.03)	.04* (.02)	.04* (.02)	.03 (.02)	
News Selections			.92*** (.02)	.84*** (.02)	.87*** (.02)	.79*** (.02)	
<i>R</i> <sup>2</sup>		.37***		.86***	.83***	.83***	

Note. *N* = 758 for the multiple regression and structural equation models. Cell entries are unstandardized regression coefficients with standard errors in parentheses. The following variables were log-transformed: news selections, news retransmissions, email-based and social media-based news retransmissions, expressed emotional evocativeness (teaser only), and hours displayed in prominent locations (i.e., editorial cue to news importance). Other predictor variables are not shown here for brevity. Full results are available upon request from the author. Emotional positivity (induced) was mean-centered (the “News Selections” model). All variance inflation factors (VIFs) were smaller than (1) 2.35 for the model predicting news selections (1.81 when the mention of diseases or BHC was effect-coded) and (2) 3.30 (2.08 for the predictors shown in this table) for the models predicting news retransmissions. Missing data were handled with listwise deletion. BHC = bad health conditions.

<sup>+</sup> *p* < .10.  
<sup>\*</sup> *p* < .05.  
<sup>\*\*</sup> *p* < .01.

.100) < *d*  
\*\*\*

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript