Virality: The Science of Sharing and the Sharing of Science

Jonah Berger[1] and Katy Milkman[2]
[1] *The University of Pennsylvania, 700 JMHH, 3730 Walnut St. Philadelphia, PA 19104*
[2] *Harvard Business School, Soldiers Field Road, Boston, MA 02163*

Summary: Awe-inspiring and emotionally rich content is more viral.

Interpersonal transmission plays an important role in social influence and social epidemics, but little is known about why certain cultural items are shared more than others. We explore characteristics of cultural items that influence their virality, examining the relationship between content characteristics and whether an article makes the *New York Times'* most emailed list. Analysis of over 7,500 articles suggests that the fit between cultural items and shared human psychology plays an important role in cultural transmission. Mind-opening and emotionally rich articles are more likely to make the most emailed list, as are more practically useful, surprising, and positive articles. These relationships hold controlling for external drivers of attention and support the notion that shared psychology acts as a selection mechanism on culture.

What makes certain cultural items more viral than others? Cultural transmission occurs in both humans and animals (de Waal 1999; Horner et al 2006), and it plays an important role in a diverse set of phenomena including the contagion of health behaviors (Couzin 2009; Christakis and Fowler 2008), diffusion of innovations (Rogers 1995), spread of ideas (Goffman and Newill 1964; Heath, Bell, and Sternberg 2001), persistence of stereotypes (Schaller, Conway, and Tanchuk, 2002), and emergence and prominence of culture (Kashima 2008; Schaller and Crandall 2004). However, while biologists and epidemiologists have studied why certain diseases have greater transmission probability (Anderson and May 1991; Giesecke 1994), less is known about characteristics of cultural items that influence their virality. Some have argued that cultural prominence is highly stochastic and driven by external factors such as advertising or social influence dynamics (Bentley and Shennan 2005; Salganik, Dodds, and Watts 2006). An alternate explanation, built on biological notions of selection (Dawkins 1976), suggests that cultural success is driven by the fit between characteristics of cultural items and shared

human psychology (Heath 1996; Heath, Bell, and Sternberg 2001; Kashima, 2000a; 2000b; Schaller and Crandall 2004). This perspective suggests that cultural items are selected and retained in the social environment based on their ability to fit with emotions, motivations, and cognitive constraints that are shared across people. Unfortunately, however, empirical tests of such cultural selection have been limited by the ability to examine the actual cultural transmission of a broad set of items in the field.

We study cultural transmission by examining the virality of over 7,500 articles from one of the world's most popular newspapers: *The New York Times*. Because the *Times* covers a wide range of topics (from world news and politics to science and travel) and has a large and diverse readership (supporting materials), it is an ideal venue for examining what types of cultural items are frequently shared. The *Times* continually reports which articles have been most emailed in the last 24 hours, and we examine the relationship between psychological aspects of content and whether articles make the most-emailed list.

One intuitive reason content may become viral is because it contains useful information (Rosnow and Fine 1976). Stories about healthy foods or websites that offer bargains contain material that is practically useful for the receiver. Because useful information has social exchange value, sharing it may encourage reciprocity (Fehr, Kirchsteiger, Riedl, 1998) and generate esteem (Engel, Blackwell, and Miniard 1993).

Though we study whether practically useful content is more likely to make the *Times'* most emailed list, we focus our attention on less intuitive content characteristics that may influence virality. We examine the hypothesis that articles that open the mind and inspire awe are more viral. People often share articles about cosmology or quantum

physics, and almost a third of *New York Times* articles about science make the paper's most emailed list (compared with a baseline rate of 20%). While it is clear how sharing health news or technology reviews provides helpful information to others, it is less obvious how articles about astronomy or brain structures provide information that is practically useful in everyday life. Instead, because such articles open the mind to vast and often unconsidered possibilities, they expand thought and cause people to look beyond themselves (Keltner and Haidt, 2003). By leading people to feel connected to the broader social world (Shiota, Keltner, and Mossman 2007), such awe-inspiring articles may induce sharing.

Emotional content more generally may also be particularly viral (Heath, Bell and Sternberg 2001, Heath 1996). One intriguing explanation for the evolution of language is that it evolved as a mechanism to help bond large social groups (Dunbar 1996). If this is the case, then people should be particularly likely to share articles that service that function. Sharing affectively rich content can reinforce shared views and deepen social bonds (Peters and Kashima 2007).

We contrast these suggestions about the virality of mind-opening and affect-laden content with a mere entertainment hypothesis: people may be more likely to share positive or surprising stories because they entertain the receiver or boost her mood. Consequently, we control for how positive and surprising each article is to examine whether articles that are mind-opening or emotionally rich are more viral above and beyond any practical or entertainment value they provide.

We quantify the extent to which each article in our sample is practically useful, mind-opening, affect-laden, positive, and surprising (supporting materials) and then run a

logistic regression to determine how these psychologically-relevant characteristics relate to an article's likelihood of making the most e-mailed list. Further, we do so while controlling for a number of external factors (separate from the content itself) that may influence virality by functioning like advertising. Appearing earlier in the physical paper, spending more time in a prominent position on the homepage, being written by a famous author, or being released online at a time when readership is greater should generate attention for an article and increase its chances of making the most emailed list. Consequently, our analyses control for where articles appeared in the physical paper, how prominently and for how long they were featured on the *New York Times* homepage, author fame, and the time of day the article was released (supporting materials). We also control for writing style (complexity and author gender) and article length, which might both influence transmission and affect the likelihood that an article contains emotionally rich or awe-inspiring content.

Our results demonstrate that shared human psychology plays an important role in virality and cultural success. Mind-opening and emotionally rich articles are significantly more likely to make the *New York Times*' most e-mailed list (Figure 1, Table S6, Model 2). As expected, articles that are more practically useful, surprising, and positive also tend to be more viral. These effects hold even controlling for external factors that might drive attention, writing style, and article length, and are robust to a host of alternative specifications (supporting materials). Our main results are also robust to controlling for an article's general topic (e.g., Opinion or Health; see Table S8, Model 3). This demonstrates that our findings are not merely driven by certain topic areas (e.g., Opinion) that having both the characteristics of interests and being more viral. Rather, this

particularly conservative test of our hypothesis demonstrates that our effects occur not only across topics but also within them; even among articles about Science or Health, for example, mind-opening and affect-laden articles are significantly more likely to be viral.

Though external drivers of attention are undoubtedly important to cultural success, our results indicate that the fit between cultural items and shared human psychology is of similar importance. Parameter estimates imply that an increase of 10% in the number of emotional words in an article increases the fitted odds that it will make the most e-mailed list by a factor of 3.4. This is equivalent to spending an additional 12 hours as the lead story on the *New York Times* website. Similarly, a one standard deviation increase in a story's "mind opening" score is equivalent to spending an additional 4.6 hours as the *Times'* lead online story. These comparisons underscore the importance of cultural selection in cultural success.

Our results also illuminate how the valence of cultural items influences transmission. Word of mouth tends to be positive (Rosen 2009), but it is unclear whether this is because of preferential transmission (i.e., people prefer to pass along positive information) or the distribution of events (i.e., there are more positive than negative things to talk about). The fact that the *New York Times* publishes a reasonably representative set of stories (ranging from extremely positive to extremely negative) allows us to distinguish between these two potential drivers. Consistent with preferential sharing, our results show that more positive articles are more likely to be viral. This does not mean people never transmit negative content (Heath, Bell, and Sternberg 2001; anxiety-provoking articles may also be more viral, supporting materials), but in general, our findings suggest that there is a bias towards sharing positive over negative news.

The findings also underscore the notion that transmission is not just about the exchange of value, but also about deepening social connections. Even controlling for their practical or entertainment value, emotional and awe-inspiring content is more likely to be viral. In fact, comparison of the relative importance of each of these variables shows that a single standard deviation increase in an article's emotionality or ability to inspire awe actually increase an article's virality considerably more than a similar increase of the other dimensions (Figure 1).

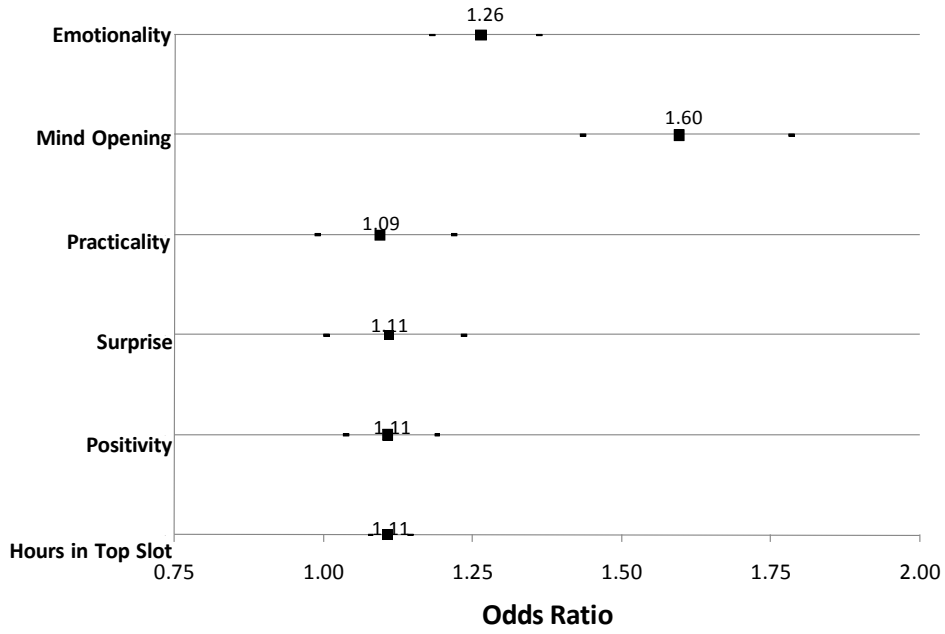
In considering how aspects of human psychology may drive cultural transmission, our findings dovetail with recent theorizing by Schaller and Crandall (2004) on the psychological foundations of culture. While individual research streams have separately discussed the diffusion of stereotypes, technological innovations, norms, and rumors, some of the psychological processes underlying transmission in each area may actually be quite similar. Though the science of memetics has become tangled in definitional questions, the general notion of cultural units remains useful because it encourages cross-disciplinary connections. Just as network researchers have benefitted from studying similarities across various domains (Watts and Strogatz 1998), researchers interested in cultural transmission may benefit from considering psychological process that may act as a selection mechanism on all types of culture.

More broadly, social epidemics are complex, multiply-determined phenomena. Though research has shown how social contagion influences behavior (Christakis and Fowler 2008) and how network structure contributes to collective outcomes (Borgatti, Mehra, Brass, and Labianca 2009; Watts 2002), less is known about how psychological factors influence *which* cultural items are more likely to be transmitted over social ties.

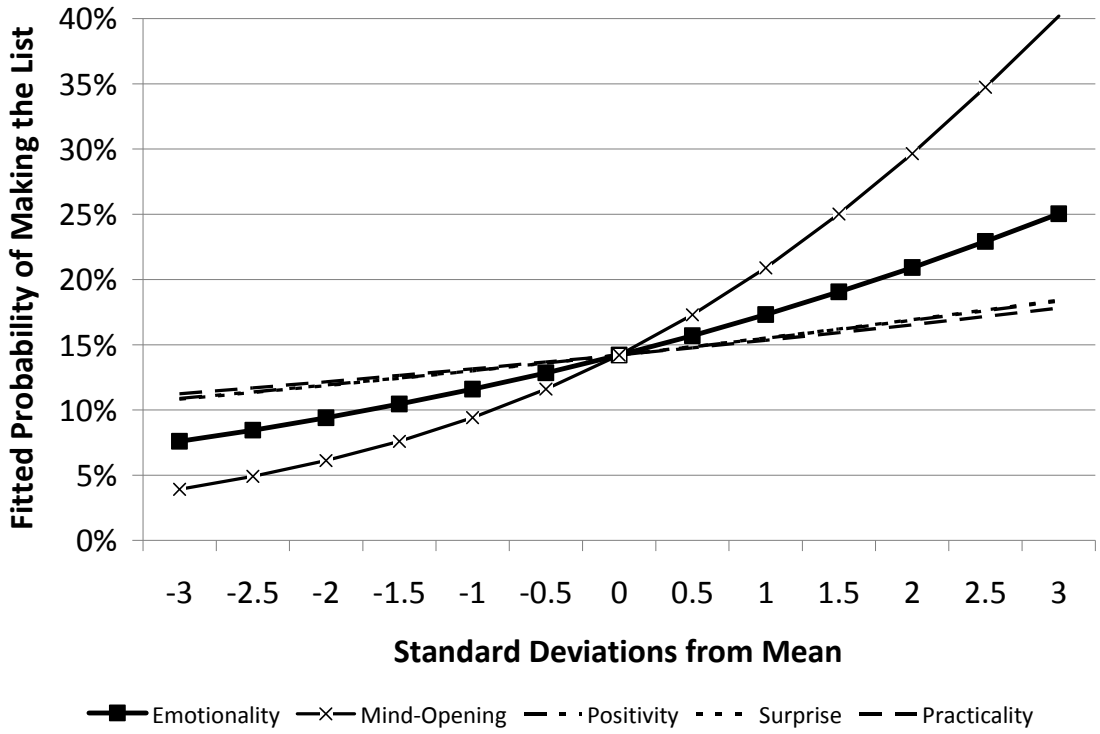
Truly gaining insight into social epidemics requires not only examining these individual factors (i.e., social network or psychological drivers of transmission) in isolation, but understanding how they combine to drive aggregate social phenomena.

Figure 1. Relationship between psychological aspects of content and virality.

Option A



Option B



References and Notes

1. F.B.M. de Waal, *Nature.* **399**, 635 (1999).
2. V. Horner, A. Whiten, E. Flynn, F.B.M. de Waal, *Proc. Natl. Acad. Sci.* **103**, 13878 (2006).
3. J. Couzin, *Science.* **323**, 454 (2009).
4. N. Christakis, J. Fowler, *New Eng. J. Med.* **358(21),** 2249 (2008).
5. E. Rogers, *Diffusion of Innovations* (Free Press, New York, 1995).
6. W. Goffman, V. Newill, *Nature.* **204**, 225 (1964).
7. C. Heath , C. Bell, E. Sternberg, *J. Pers. Soc. Psychol*. **81**, 1028 (2001).
8. M. Schaller, L. Conway, T. Tanchuk, *J. Pers. Soc. Psychol.* **82**, 861 (2002).
9. Y. Kashima, *Soc. Pers. Psychol. Comp.* **2**, 107 (2008).
10. M. Schaller, C. Crandall, *The Psychological Foundations of Culture* (Lawrence Erlbaum Associates, Mahwah, 2004).
11. R. Anderson, R. May, *Infectious Diseases of Humans: Dynamics and Control.* (Oxford University Press, Oxford, 1991)
12. J. Giesecke, *Modern Infectious Disease Epidemiology* (Arnold Publication, London, 1994)
13. R.A. Bentley, S. Shennan, *Science.* **309**, 877 (2005).
14. M. Salganik, P. Dodds, D. Watts, *Science.* **311**, 854 (2006).
15. R. Dawkins, *The Selfish Gene* (Oxford University Press, Oxford, 1976)
16. C. Heath, *Org. Behav. Hum Dec. Proc.* **68**, 79 (1996).
17. Y. Kashima, *Pers. Soc. Psychol. Bull.* **26**, 594 (2000).
18. Y. Kashima, *Eur. J. Soc. Psychol.* **30**, 383 (2000).
19. Supporting Materials
20. R. L. Rosnow, G.A. Fine, *Rumor and Gossip: The Social Psychology of Hearsay* (Elsevier, New York, 1976).
21. E. Fehr, G. Kirchsteiger, A. Riedl, *Eur. Econ. Rev.* **42**, 1 (1998).
22. B. Engel, R.D. Blackwell, P.W. Miniard, *Consumer Behavior* (The Dryden Press, Fort Worth, TX, 1993).
23. D. Keltner, J. Haidt, *Cog. Emot.* **17**, 297 (2003).
24. M. Shiota, D. Keltner, A. Mossman, *Cog. Emot.* **21**, 944 (2007).
25. R. Dunbar, *Grooming, Gossip, and the Evolution of Language* (Harvard University Press, Cambridge, MA, 1996).
26. K. Peters, Y. Kashima, *J. Pers. Soc. Psychol*. **93**, 780 (2007).
27. E. Rosen, *The Anatomy of Buzz Revisited.* (Broadway Business, New York, 2009).
28. D. J. Watts, S. H. Strogatz, *Nature.* **393**, 440 (1998).
29. S. Borgatti, A. Mehra, D. Brass, G. Labianca, *Science.* **323**, 892 (2009).
30. D. Watts, *Proc. Natl. Acad. Sci.* **99**, 5766 (2002).

Supporting Online Materials

Materials and Methods

Data

The data set contains information about every article that appeared on the *New York Times* homepage (www.nytimes.com) between August 30th, 2008 and February 15th, 2009. The *Times* was the number one newspaper website at the time of our study (*S1*) and reached approximately 22.4 million U.S. readers (*S2*). The paper's online audience is approximately 56% male with a median age of 48 years and a median income of $84,660 (*S3*). The data was captured by a webcrawler that took snapshots of the *Times* homepage every 15 minutes. We focus our attention on articles written specifically for the *New York Times* and our final data set contains 7,710 articles.[1]

We captured each article's title, full text, author name(s), topic area (e.g., Opinion or Sports), and one sentence summary created by the *Times*. We also captured the section, page, and publication date when an article appeared in the print paper, and the dates, times, locations and durations of all appearances the article made on the *Times* homepage.

Finally, we collected data every 15 minutes on which articles appeared on the *New York Times* website's most emailed list. This list is updated every fifteen minutes and reports the 25 articles that were e-mailed from the *Times* website most frequently over the previous 24 hours. 77% of articles that make the most emailed list are featured on the homepage at some point, and 20% of the articles in our final data set (N =1,545) eventually earned a position on the most e-mailed list.

Independent Variables

Primary Predictor Variables

We used the data described above, which was captured for each *New York Times* article, to generate our independent variables. Automated sentiment analysis was used to quantify the positivity (i.e., valence) and emotionality of each article. A computer program counted the number of positive and negative words in each *New York Times* article using a list of 7,630 words classified as positive or negative by human readers (following *S4* we rely on *OpinionFinder's Subjectivity Lexicon*). *Positivity* was quantified as the difference between the percentage of positive and negative words in an article. Examples of highly positive articles include: "Wide-Eyed New Arrivals Falling in Love With the City" and "Free Films on Imdb," while examples of extremely negative articles include: "Web Rumors Tied to Korean Actress's Suicide" and "Stampede in India Kills at Least 147." *Emotionality* was quantified as the percentage of words in an article that were classified as either positive or negative. Examples of highly emotional articles include titles such as: "Redefining Depression as Mere Sadness" and "Brad Pitt Supports Same-Sex Marriage."

We relied on human coders to generate our three other primary predictor variables. Using the title and summary of each article, independent coders (two per

[1] 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 had no text and were also not included.

dimension) rated the same random sample of articles (N = 3,000) on the extent to which they were practically useful, mind-opening, and surprising. Articles that earned a position on the most e-mailed list were oversampled to comprise 27% of the coded subsample. Raters received detailed coding instructions (see Table S1) and were given feedback on their coding of a test set of 50 articles until it was clear they understood the relevant dimension. All variables were coded on 1-5 likert scales and instances of extreme disagreement (more than three units) were resolved through discussion. Inter-rater reliability was good for all three measures (practically-useful r = .80; mind-opening r = .56; surprise r = .57), and scores were averaged across the two coders for each variable (see Table S2 for summary statistics describing our primary predictor variables). Examples of articles with high practical value include: "Vacation Plans That Avoid the Pump" and "2008 Holiday Gift Guide". Examples of highly mind-opening articles include titles such as: "From Old Vials, New Hints on Origin of Life" and "Now in Sight: Far-Off Planets". Examples of articles rated as extremely surprising include: "Iverson Adds Scoring and Who-Knows-What to Pistons" and "Day Care May Offer Asthma Protection." Standardized versions of the primary predictor variables were used to improve the interpretability of our regression results (see Table S5 for correlations between our standardized primary predictor variables).

Control Variables

External Drivers of Attention. We control for external drivers of attention to each article in a number of ways. First, we control for where, when, and for how long articles were featured on the *Times* homepage. Articles could appear in several dozen positions on the homepage, so we aggregated positions into seven general regions based on locations that should receive similar amounts of attention (see Figure S1).[2] The average article in our data set appeared somewhere on the homepage for a total of 27 hours (standard deviation = 30 hours), and variables indicating the amount of time an article spent in each region were included as controls in our primary analyses after winsorization of the top 1% of outliers (see Table S3 for summary statistics). We also control for the day of the week and the time of day (6 am – 6 pm EST or 6 pm – 6 pm EST) when an article first appeared online.

Articles may also receive more attention if they appear in certain sections of the print edition of the *Times* or on earlier pages of a given section. Consequently, our analyses include dummy variables to control for the section where an article appeared in print (A, B, C, D, E, F or "Other") as well as an indicator variable for the page on which

[2] These include: (1) the "top feature" area, comprised of the two articles whose headlines appeared in the largest font at the top of the page, one of which was always associated with a lead photograph; (2) the "near top feature" area, comprised of the several articles whose headlines appeared below the top feature, which typically were followed by one paragraph summaries; (3) the "bulleted sub-feature" area, comprised of the small bullet point stories that were associated with "top feature" and "near top feature" stories; (4) the "right column" area on the top right of the homepage, where featured opinion and special interest stories appeared; (5) the "more news" area, in which headlines were provided below the main news stories in bullet form without descriptions; (6) the "middle feature bar" area, in which six stories selected by the editors for their appeal were always featured with associated images or summaries; and (7) the "bottom list" area, in which three stories from each section of the newspaper (e.g., Sports, World, Business, etc.) were listed in bulleted form below their section header and where most stories that appeared in other positions eventually migrated. The homepage was displayed in a standardized format throughout our data collection period.

it appeared (e.g., 1, 17, etc.). See Table S4 for summary statistics. We also interact each section dummy with our page indicator variable to control for the possibility that appearing earlier in some sections has a different effect on an article's likelihood of making the most e-mailed list than appearing earlier in other sections.

Finally, articles written by prominent authors may attract more attention. We quantify first author fame by counting the number of hits returned by a Google search of the author's full name as of February 15, 2009. Due to its skew, we use the logarithm of this variable in our analyses (see Table S5 for summary statistics).

Other Controls. We also control for writing style and article length. One important measure of writing style is complexity, or how easy a piece of text is for people to read. We control for this in our analyses using the *SMOG Complexity Index* (*S5*). Prior research has also demonstrated that male and female authors tend to have different writing styles (*S6*). Consequently, 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-gender mapping list used in prior research (*S7*). For names that were classified as gender neutral or did not appear on this list, research assistants found the author's biography online and manually determined his or her gender.

Finally, we control for article length. Longer articles may be more likely to have some of the characteristics of interest (e.g., ability to inspire awe, etc.) but may simply be more viral because they contain more information. Consequently we include the number of words in a given article as a control variable in our primary analysis. See Table S2 for summary statistics.

Regression Specification

To analyze the relationship between content characteristics and the likelihood an article will make the *New York Times'* most e-mailed list, we use the following logistic regression specification:

$$(2) \qquad makes\_it_a = \frac{1}{1+\exp\left\{-\left(\begin{array}{l} \alpha + \beta_1 * z\text{-}practicality_a + \beta_2 * z\text{-}surprise\_score_a + \\ \beta_3 * z\text{-}positivity_a + \beta_4 * z\text{-}emotionality_a + \\ \beta_5 * z\text{-}mind\text{-}opening\_score_a + \theta' * X_a \end{array}\right)\right\}}$$

where $makes\_it_a$ is a variable that takes on a value of one when an article a earns a position on the most e-mailed list and zero otherwise. Our primary predictor variables quantify the extent to which an article a was deemed practically useful, surprising, positive, emotional or mind-opening. X_a is a vector of the other control variables described above, including: indicators of the number of hours an article spent in each of seven online locations, dummies indicating the day of the week when the article first appeared online (Monday omitted), a dummy indicating whether the article first appeared online at night (6 pm – 6 am EST), a dummy indicating which section in the physical paper the article appeared in, an indicator of the page number in that section in the physical paper that the article appeared in, interactions between physical paper section dummies and the page number indicator, the first author's fame score, the article's standardized complexity score, dummies indicating whether the first author is female or of unknown gender (due to a missing byline), the article's wordcount, and a dummy

indicating whether the article in question was one of the 3,000 coded manually on the characteristics: *practicality*, *surprise*, and *mind-opening*.[3]

Results

Table S6 displays our main regression results. Model 1 presents a baseline logistic regression model including only our primary predictor variables (*emotionality*, *surprise*, *practicality*, *positivity*, and *mind-opening*). As predicted, our results demonstrate that the higher an article's rating on most of these characteristics, the more likely it is to make the most e-mailed list (the effect of *practicality* is not statistically significant).

These results are robust to the inclusion of a host of control variables. Model 2 – the primary specification we discuss – includes controls for external drivers of attention (e.g., prominent placement on the *Times* homepage or in the physical paper), writing style, and article length. Not surprisingly, we find that articles that received more external "advertising" (i.e., appeared on earlier pages of the physical paper, spent longer on more visible positions on the homepage, or were written by more famous authors) were more likely to make the most emailed list. Longer articles and articles written by women are also significantly more viral. There is no significant relationship between article complexity and virality. Even accounting for these factors, however, the relationship between our content characteristics and virality persists, and in fact, the effect of *practicality* becomes marginally significant.

Our primary findings also hold when we control for the general topic of each article. *The New York Times* classifies each article into one of 20 general topic areas: Arts, Books, Business, Dining, Education, Fashion, Health, Home, Domestic, International, Magazine, Movies, New York, Opinion, Science, Sports, Technology, Travel, Politics, Giving, Real Estate, or Automobiles.[4] Certain general topics may attract more attention from readers (e.g., Opinion), and if our results about characteristics of viral content are robust, they should persist whether or not an article's topic area is held constant. Table S6 (Model 3) shows that this is the case. Even when indicator variables for each topic area are included in our model, more surprising, mind-opening, emotional, and positive articles are all significantly more likely than others to make the *New York Times*' most e-mailed list. More practically useful articles are also more likely to make the list on average, though this trend is not statistically significant.

These results are robust to changes in the specification of our control variables. Our primary results remain meaningfully unchanged in terms of magnitude and statistical significance if we: add squared and/or cubed terms quantifying how long an article spent in each of seven homepage regions; add dummies indicating whether an article ever appeared in a given homepage region; split the region variables into time spent in each

[3] This dummy accounts for the oversampling of articles that made the most e-mailed list in the coded subset of variables. The three manually coded variables are set to zero for all articles besides the 3,000 that were coded, and the dummy variable indicating that an article was not coded accounts for the average likelihood that these articles make the most e-mailed list.

[4] Because articles from the Giving, Real Estate and Automobiles sections appear with extreme infrequency on the *New York Times* homepage (less than 60 times each), we group these topical sections into a single "other" category and this is the omitted dummy variable in our analyses that control for topical section.

region during the day (6 am – 6 pm EST) and night (6 am – 6 pm EST); control for the day of the week when an article was published in the physical paper (instead of online); or winsorize the top and bottom 1% of outliers for each control variable in our regression. These robustness checks increase our confidence in the validity of our findings.

Additional analyses suggest that anxiety-producing content may also be more viral. While our results show that, in general, positive articles are more likely to make the most emailed list than negative articles, certain negative emotions may be linked to transmission (e.g., Heath, Bell, and Sternberg 2001). We used textual analysis software (Linguistic Inquiry and Word Count) to calculate the percentage of anxiety-related words in each article. When this standardized variable is included in our primary regression, we see that more anxiety-inducing articles are more likely to make the most emailed list (our other effects of interest remain the same). A qualitative examination of the set of articles rated as highly surprising also supports this finding. While a number of highly surprising articles appear entertaining, many others seem to be about anxiety-inducing topics (e.g., a plane crash or shooting). Such articles are certainly about unexpected events, but they may be shared not only because they are surprising and entertaining, but also because they contain negative shock value, and people want to share such stories with others to reduce their own anxiety. Thus while overall, positivity may lead to sharing, content that induces certain negative emotions may also be shared.

Supporting References

S1. Nielson, *NetRatings Netview* (2008).
S2. Nielson Online, *NetView*. **11** (2008).
S3. Nielson, *NetRatings*. **3** (2009).
S4. T. W. J. Wilson, P. Hoffmann, *Hum*. *Lang*. *Tech*. *Con*. (2005).
S5. G. H. McLaughlin, *J. Read*. **12**, 639 (1969).
S6. M. Koppel, S. Argamon, A.R. Shimoni, *Lit*. *Ling*. *Comp*. **17**, 401 (2002).
S7. F. S. Morton, F. Zettelmeyer, *Quant*. *Mktg*. *Econ*. **1**, 65 (2003).

Figure S1: Homepage location classifications. *The New York Times* homepage layout remained in a standard format throughout the period of data collection. Portions with "X's" through them always featured AP and Reuters news stories, videos, blogs, or advertisements rather than articles by *Times* reporters.

Table S1: Descriptions provided to coders.

Practically Useful

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. An article about travel might influence where the reader goes on vacation. Importantly, practical utility focuses on the here and soon. An article about cosmology, for example, might be very eye-opening, but if it doesn't affect what the reader does soon, that should not be coded as having high practical utility. Also focus on utility for the reader. If it won't potentially lead them to behave differently, it shouldn't be coded as having practical utility.

Code these articles on a 1-5 scale where:
1 = not at all practically useful and
5 = extremely practically useful

Importantly, code based on whether the article would have practical utility to the reader. So an article about how to be a better rower would have practical utility because it would change rowers' technique who read it, even if it would not affect how you behave. Similarly, if the article provides useful practical info on traveling to Monaco, it would be practically useful to people traveling there even if you are not planning on going.

Mind-Opening/Awe-Inspiring

Some articles enlighten the reader. Such articles teach people things or share a perspective that opens people's minds. This knowledge may not change how people live their everyday life, and thus may have little practical value, but it is more about knowledge for knowledge's sake. For example, reading an article about how the human mind works, or the structure of the universe is unlikely to change what people eat or what they do this weekend, but it will open up their mind to new possibilities and insights. Articles that are high on this dimension may evoke feelings of awe and wonder, and even articles that are mildly high on this dimension may just shift how the readers sees things, even if it doesn't really directly have practical utility.

Code these articles on a 1-5 scale where:
1 = not at all mind-opening/mind opening and
5 = extremely mind-opening/mind opening

Surprising/Unexpected

When thinking about sending an article to someone else, people may consider whether the information in the article will be "surprising" to the person receiving it. This can happen in two ways. One possibility is that the recipient will find it shocking. If an article suggested that milk killed brain cells, that would be very surprising. Another possibility is that the recipient will have a very low likelihood of being aware of the information contained in the article already. If milk killed brain cells, but everyone had already heard that news, then it would not be very surprising. Similarly, if the Yankees win the world series that would be interesting, but given that most people would already know about it, it would not be very surprising.

Code articles using a 1-5 scale, where:
1 = very expected and
5 = really unexpected

Table S2: Predictor variable summary statistics.

		Mean	Std. Dev.
Primary Predictor Variables	**Emotionality***	7.43%	1.91%
	Positivity*	0.99%	1.85%
	Surprising*	2.56	0.95
	Mind-Opening*	1.73	0.74
	Practicality*	0.47	0.76
Other Control Variables	**Wordcount**	1,033.36	740.58
	Complexity*	11.10	1.53
	Author Fame	9.14	2.53
	Author Female	0.29	0.45
	Author Male	0.66	0.47

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

Table S3: 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	26%	32%	2.63	2.97
Near Top Feature	31%	31%	5.00	5.09
Right Column	21%	31%	3.80	5.07
Middle Feature Bar	23%	32%	11.52	11.62
Bulleted Sub-Feature	34%	25%	3.62	4.11
More News	28%	24%	3.65	4.16
Bottom List	81%	20%	23.00	28.25

Table S4: 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	40%	25%	15.52	11.96
Section B	15%	11%	6.35	3.71
Section C	9%	16%	4.05	3.78
Section D	7%	19%	3.01	1.65
Section E	3%	22%	4.74	4.72
Section F	2%	42%	3.28	2.02
Other Section	13%	24%	9.47	13.64
Never in Paper	11%	10%	-	-

Table S5: Correlations between predictor variables.

	Word Count x 10^{-3}	Complexity	Author Fame	Author Female	Emotionality	Positivity	Surprise	Mind-Opening
Word Count x 10^{-3}	1.00							
Complexity	-0.06*	1.00						
Author Fame	-0.05*	0.02	1.00					
Author Female	-0.01	-0.03*	0.00	1.00				
Emotionality	0.07*	0.05*	-0.10*	-0.07*	1.00			
Positivity	0.06*	-0.05*	-0.04*	0.06*	0.05*	1.00		
Surprise	0.01	0.06*	0.04*	0.11*	-0.15*	-0.08*	1.00	
Mind-Opening	0.12*	-0.03	0.01	0.08*	0.11*	0.13*	0.31*	1.00
Practical	0.05*	-0.05*	-0.03*	0.05*	0.06*	0.11*	0.10*	0.44*

*Significant at 5% level.

Table S6: Logistic regression predicting the probability that a given article will earn a position on the *New York Times'* most e-mailed list based on characteristics of its content, authors, and appearance in the physical paper and *New York Times* homepage.

		(1)	(2)	(3)
Primary Predictor Variables	**Mind-Opening**	0.51***	0.47***	0.38***
		(0.05)	(0.06)	(0.06)
	Emotionality	0.26***	0.23***	0.11***
		(0.03)	(0.04)	(0.04)
	Practicality	-0.02	0.09*	0.08
		(0.05)	(0.05)	(0.06)
	Surprise	0.11**	0.10*	0.20***
		(0.05)	(0.05)	(0.06)
	Positivity	0.08***	0.10***	0.09**
		(0.03)	(0.04)	(0.04)
Homepage Location Control Variables	**Top Feature**	-	0.10***	0.09***
		-	(0.02)	(0.02)
	Near Top Feature	-	0.09***	0.08***
		-	(0.01)	(0.01)
	Right Column	-	0.14***	0.09***
		-	(0.01)	(0.01)
	Middle Feature Bar	-	0.05***	0.05***
		-	(0.00)	(0.00)
	Bulleted Sub-Feature	-	0.04***	0.05***
		-	(0.01)	(0.01)
	More News	-	-0.01	0.05***
		-	(0.01)	(0.01)
	Bottom List x 10	-	0.05***	0.07***
		-	(0.01)	(0.02)
Other Control Variables	**Word Count x 10^{-3}**	-	0.48***	0.66***
		-	(0.06)	(0.06)
	Complexity	-	0.02	0.03
		-	(0.04)	(0.04)
	First Author Fame	-	0.16***	0.15***
		-	(0.02)	(0.02)
	Female First Author	-	0.30***	0.28***
		-	(0.07)	(0.08)
	Uncredited	-	0.40*	-0.64**
		-	(0.23)	(0.28)
Newspaper Location & Web Timing Controls		No	Yes	Yes
Article Section Dummies (arts, books, etc.)		No	No	Yes
Observations		7,710	7,710	7,710
McFadden's R^2		0.06	0.24	0.31
Log likelihood		-3635.25	-2940.01	-2653.35
Likelihood-ratio chi^2 (A nested in B)			1390.48***	573.33***

*Significant at 10% level. **Significant at 5% level. ***Significant at the 1% level.