Author Manuscript

Curiosity in News Consumption

Journal:	*Applied Cognitive Psychology*
Manuscript ID	ACP-23-0309.R2
Wiley - Manuscript type:	Research Article
Recommended Editor:	Charles B. Stone
Keywords:	Belief-based Utility, Curiosity, Information Gap, Natural Language Processing, News Consumption

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

appendix.tex
descriptive_stats_facebook.tex
Headline Curiosity_ACP.zip
manuscript.tex
measure_facebook.tex
nb_exclude_duplicate.tex
nb_facebook.tex
nb_impute.tex
ols_alternative_sentiment.tex
ols_exclude_duplicate.tex
ols_facebook.tex
ols_impute.tex
quadratic_kl_ols_nb.tex
Refs.bib
reg_topics.tex
reg_validation.tex

SCHOLARONE™
Manuscripts

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acp.4195

This article is protected by copyright. All rights reserved.

Author Manuscript

Curiosity in News Consumption

Abstract

We analyze how curiosity drives news consumption. We test predictions of the information-gap theory of curiosity using over 100,000 WeChat news articles, applying NLP methods to construct measures of salience, importance, and surprisingness associated with news headlines, experimentally validating these measures, and using them to predict clicks. Our findings confirm that people tend to consume news when: the headline sparks a salient question; the content appears more important (e.g., emphasized by the headline's position on the webpage or an exclamation mark); the headline refers to more surprising topics (measured as the KL-divergence from a baseline topic distribution); and the headline has lower valence. Information-gap theory helps predict aggregate news consumption. Yet our data also reveal a small negative correlation between the number of clicks and the ratio of likes to clicks, suggesting that while inducing curiosity can drive short-term news consumption, it doesn't necessarily enhance long-term reader engagement.

For Peer Review

Author Manuscript

1 Introduction

In the modern age of information, people have more choice than ever before about what news to consume, and yet people remain poorly informed, sometimes even misinformed, about current events (Watts, Rothschild, & Mobius, 2021). How do people decide what news articles to read? Rational choice theory suggests that people will read the news that may prove useful to know, for example, seeking business news that will help them make investments or political news that will help them vote rationally. Political psychologists have recognized that confirmation bias (Nickerson, 1998) and identity-threat avoidance (Golman, Loewenstein, Moene, & Zarri, 2016; Kahan et al., 2012; Van Bavel & Pereira, 2018) may also cause people to demand slanted political news from sources that share their political leanings and to avoid news from sources they tend to disagree with (Bakshy, Messing, & Adamic, 2015; Gentzkow & Shapiro, 2010; Mullainathan & Shleifer, 2005). Nowadays people often get their news from echo chambers on social media (Cinelli, Morales, Galeazzi, Quattrociocchi, & Starnini, 2021; Hosseinmardi et al., 2021; Sunstein, 2017). When the news is consistent with their pre-existing views, they may rely on intuitive reasoning and accept it, whereas when it conflicts with their worldview, they may engage in more deliberative reasoning and possibly reject it (Pennycook, 2023; Taber & Lodge, 2006). Yet, while many people demand slanted political news, many others avoid political news altogether (Bode, Vraga, & Troller-Renfree, 2017; Skovsgaard, Shehata, & Strömbäck, 2016), while at the same time seeking out human-interest stories, such as celebrity news (Hamilton, 2004). This suggests that curiosity may be a critical determinant of news consumption.

Curiosity is a powerful motivational force (Hilchey, Rondina, & Soman, 2022; Hsee & Ruan, 2016; Van Dijk & Zeelenberg, 2007). The information-gap theory of curiosity suggests that people become curious when their attention is drawn to a question (or information gap). The theory thus predicts that curiosity depends on a variety of factors that influence attention (Golman & Loewenstein, 2018; Golman, Loewenstein, Molnar, & Saccardo, 2021; Loewenstein, 1994). The theory accounts for the effectiveness of some forms of clickbait, which make an information gap salient by using forward-referencing, adopting particular linguistic strate-

Author Manuscript

gies, or deliberately hiding a piece of information (Blom & Hansen, 2015; Scott, 2021; Venneti & Alam, 2018). In this paper we test a number of predictions of information-gap theory for the consumption of news.

We analyze a large dataset consisting of more than 100,000 news articles posted to WeChat, the most popular social media platform in China. (In the Appendix we also present an analysis of news articles liked on Facebook, which yields similar results.) We observe how many users read each article. We identify proxy variables for the existence of a salient information gap and for the importance of the information gaps presented by the headlines, and by applying natural language processing (NLP) methods to these headlines, we construct a measure of the surprisingness of the topics covered in the headlines. We perform validation analysis showing that these variables align with perceptions of the salience, importance, and surprisingness of information gaps in five preregistered experiments. We then estimate a model of news consumption driven by these variables.

We find that people are more likely to consume news when: 1) the headline raises a salient question, indicated by the presence of a question mark or ellipsis in the headline; 2) the news appears more important, either because the headline includes an exclamation mark for emphasis or because the headline is displayed in a higher position on the webpage; 3) the topics mentioned in the headline are more surprising, as measured by the relative entropy (KL-divergence) of the distribution over topics referenced in the headline relative to a baseline distribution over topics; and 4) the headline has lower (more negative) valence in sentiment analysis. These empirical results confirm the predictions of information-gap theory, as well as earlier findings of negativity bias for news. News consumption depends not only on the potential usefulness of the news, but also on how well the headlines grab attention and stimulate curiosity. Yet while opening a salient information gap or teasing the audience with a surprising tidbit can stimulate curiosity to read a news article, it does not necessarily make the article more interesting to read. We find a slightly negative relationship between the number of readings of an article and the fraction of readers that like the article. Simply grabbing an audience's attention and sparking their curiosity about particular articles may be insufficient to keep readers

For Peer Review

Author Manuscript

engaged over the long run.

Our finding that headlines that highlight information gaps stimulate news consumption has a number of implications for psychology, marketing, politics, economics, and public policy. Broadly, it contributes to the accumulating evidence that belief-based utility is often as, or more, important than material utility when people are deciding what they want to know (Golman, Hagmann, & Loewenstein, 2017; Sharot & Sunstein, 2020). Rational choice theory cannot fully account for demand for news. Specifically, our study provides naturalistic evidence from the field in support of information-gap theory. Golman et al. (2021) show that the importance, salience, and valence of an information gap affect curiosity in a highly controlled experimental task, but no prior work has tested these predictions in the field, where a variety of other stimuli may compete for people's attention.

News consumption is not only a test bed for information-gap theory, but an important topic in its own right. Our findings that news consumption depends not only on the potential usefulness of the news, but also on how well the headlines grab attention and stimulate curiosity, connect well to the practice of curiosity marketing, where people can read part of an online article for free but need to pay to read the whole article. Additionally, our result provides an account of why politicians may compete for our attention, instead of advancing the best, most informative arguments for their positions, since voters may only consume news about them when it first grabs their attention. Politicians who attract attention may thus utilize news coverage as free publicity to raise their own popularity. Our work also provides guidance to governments and non-profit organizations aiming to disseminate useful information through the news media. Simply making information available is far from enough, since increased exposure to information does not necessarily produce more accurate beliefs about current events (Gentzkow & Shapiro, 2004). Beyond just making information available, public-serving organizations need to attract attention to this information, present it clearly and concisely, and trigger behavioral responses, if they want it to make a real impact.

The remainder of the paper is organized as follows. Section 2 describes our data sources and computational methodology. Section 3 describes our measures

For Peer Review

Author Manuscript

of information-gap salience, importance, surprisingness, and valence. Section 4 presents our regression specifications. Section 5 presents our results. Section 6 concludes.

2 Data

2.1 WeChat Dataset

We obtain a large WeChat dataset collected by newrank.cn, which is also available from WeChat through their open-access API. This WeChat dataset contains 120,000 original articles published from about 37,000 accounts from March 18 to March 20 in 2021.

This dataset is appealing because it offers comprehensive data. It contains all articles posted on these three days, so there is no selection based on article popularity or topic. Moreover, in contrast to other news platforms, WeChat records the specific number of readings for each published article and makes this data publicly accessible. This gives us a clear picture of news consumption on the platform.

We restrict and clean the data to avoid a number of confounds. First, we wish to focus on the text of an article's headline, rather than the image presented alongside it. When browsing WeChat articles, users not only see each article's headline but also an associated image beside the headline. (See Figure 1 in Appendix.) Sometimes a headline is displayed with a prominent image that occupies most of the layout, whereas other headlines are each displayed with a small image that occupies only approximately 20 percent of the layout. To avoid the confounding influence of these images, we did not collect data for the articles displayed with prominent images but instead collected data for the articles displayed with small images. (We also keep track of the layout position of each article, which is a predictor of news consumption described in Section 2.2.6.)

Second, we wish to focus our analysis on headlines that are all in the same language. While Mandarin is the most prevalent language for articles posted to

For Peer Review

Author Manuscript

WeChat, there are also some articles whose headlines are in other languages, such as Tibetan, Korean, or English. We exclude articles whose headlines do not contain Mandarin, comprising 1,809 articles, 1.5 percent of the original WeChat dataset.

Third, we wish to focus on articles that are accessible to at least a minimal audience. Articles posted on WeChat are generally accessible to an account's subscribers (much like Twitter or Facebook). While we do not directly observe the number of subscribers for each account, we keep track of the maximum number of readings of any of the articles an account has posted in the three days. If this maximum is 0, i.e., if nobody has read any of the articles posted from a particular account, we exclude all of these articles from the analysis, effectively attributing the lack of interest in these articles to the absence of an audience for the account rather than to features of the headlines. (We retain many articles that are not read at all in our dataset, as long as the account that posted these articles also posted another article read by at least one person.) In total, we exclude 5,836 articles, 4.9 percent of the WeChat dataset, for this reason. (Furthermore, we also control for the source of each article in our primary regression specifications, as described in Section 4.)

There are also 4,388 duplicate articles published several times by the same accounts, comprising 3.9 percent of the cleaned WeChat dataset.[1] We includeduplicate headlines in our main empirical analysis and then exclude duplicate headlines in a robustness check (presented in Tables 5 and 6 in Appendix).

One drawback of this dataset is that WeChat does not report the exact number of readings of the most popular articles. If an article is read more than 100,000 times, WeChat records the number of readings as 100,000+ instead of reporting an exact number. This is the case for 135 articles, 0.11 percent of the dataset. Our primary analysis imputes a value of 100,001 readings, effectively rounding down, for these articles. As a robustness check, in the Appendix, we also fit a lognormal distribution to the number of readings and impute the expected value of this distribution conditional on exceeding 100,000 for the missing number of readings of these articles. We thus impute approximately 171,000 readings for these 135 articles for the robustness check. (See Tables 7 and 8 in Appendix.) After the

[1]No articles were published by multiple accounts in our data

Author Manuscript

data cleaning described above, the resulting dataset includes 112,901 news articles published by more than 36,000 accounts on three consecutive days.

2.2 Computational Methodology

2.2.1 Pre-processing for Text Analysis

To implement text analysis and topic modeling, we first tokenize each headline into a list of words. We then remove common *stop-words* like "of" or "and" to focus on informational words. Next, as some words occur together frequently, we identify these word pairs (also called as N-grams, where N is the number of words in a sequence). After processing headlines into these lists of terms, we apply doc2bow to create a dictionary, i.e, a mapping between word terms and their integer IDs, by computing document frequency (i.e., the number of documents containing a term) and term frequency (i.e., the number of occurrences of a term in each document) using the gensim.corpora module in the Gensim Python library. In our context, a "document" is a single headline.

2.2.2 Topic Modeling

With the dictionary derived from text pre-processing, we apply the Latent Dirichlet Allocation (LDA) topic model in natural language processing (Blei, Ng, & Jordan, 2003), to identify a certain number of topics in a collection of headlines. In our LDA model, each topic is represented as a probability distribution over words within our constructed dictionary. Correspondingly, in the LDA model, each headline is typically assigned multiple topics with different probabilities, representing the different themes or concepts that the headline touches upon.[2] Then we use our LDA model to automatically classify each headline according to its relevance to each identified topic, where the number of topics is an integer parameter we must specify.

[2]The probabilities indicate the extent to which each topic is relevant to the headline, and the sum of the probabilities across all topics is equal to 1. This allows for a more nuanced and flexible representation of the content of the headline than a simple categorical assignment of a single topic label.

Author Manuscript

To improve the quality of the identified topics from the LDA model, we take topic coherence (Röder, Both, & Hinneburg, 2015) as a proxy for topic quality, and we determine the number of topics for our LDA model by maximizing topic coherence. By iterating the LDA model through different numbers of topics, we identify 55 topics in our LDA model, which gives the highest topic coherence score, 0.575, in the series of iterations (see Figure 2 in Appendix).

2.2.3 Punctuation Mapping

The use of punctuation is an important way for individuals to express (Provine, Spencer, & Mandell, 2007) and perceive emotions (Hancock, Landrigan, & Silver, 2007) in text-based communication. To detect the existence of multiple punctuation marks in headlines, we use regular expression mapping to respectively identify question marks, exclamation marks and ellipses in headlines.

2.2.4 Regular Expression Mapping

In addition to punctuations, we further employ regular expression mapping to identify negation headlines. We identify negations based on whether a headline includes negation words, such as “no”, “not” etc. The completelist of negation words are listed in Figure 3 in Appendix. Of course, the presence of negation words does not guarantee that the headline rejects an accepted premise or expectation, but these simple indicator variables are useful proxies for higher-level features of headlines that we cannot perfectly extract with computational methods.

2.2.5 Sentiment Analysis

We use sentiment analysis to determine the valence of headlines, and include this as a predictor in our model. While there are multiple methods of measuring sentiment and other dimensions of emotions in English text, such as using the Evaluative Lexicon (Rocklage, Rucker, & Nordgren, 2018) (see Kardes, Herr, and Schwarz (2019) for a review), most of these methods are not directly applicable to Chinese text

For Peer Review

Author Manuscript

because they are primarily developed with only English language training inputs. For our primary analysis, we make use of SnowNLP, an off-the-shelf approach especially for analyzing Chinese text, to get sentiment values for headlines with a range of $(0, 1)$. We also identify a Python package, cntext, for evaluating valence and arousal in Chinese language texts. We redo our analysis with these sentiment measures as a robustness check in the Appendix.

2.2.6 Position Tag

A WeChat account can post articles only once per day. When an account posts multiple news articles at the same time, they are grouped together. The account holder effectively makes an editorial decision about the order, or position, of the headlines. Articles posted on WeChat without prominent images have position tags marked from one to seven from top to bottom.[3]

2.2.7 Publication Time Stamp

Based on the published date and time stamps in the WeChat dataset, we identify published days of week for all news items. We also categorize publication time by breaking a day into eight consecutive 3-hour intervals (i.e., 8-11am, 11am-2pm, 2-5pm, etc.).

2.2.8 Topic Tag

WeChat account holders may add topic tags (e.g., celebrity, education, science, technology, politics, health, fashion, parenting, finance and business, etc.) or even more specific tags (e.g., finance on company and finance on fund) when they publish articles. We merge similar specific topic tags into the same general topic category, such as merging finance on company and finance on fund together into the finance category, and add these topic tags as controls in our regression analysis.

[3]Articles with prominent images would have position tag zero, but are excluded from our analysis.

For Peer Review

Author Manuscript

3 Operationalization

To explain preference for acquiring information based not only on material concerns but also on curiosity, Golman et al. (2021) propose that three fundamental determinants of attention contribute to curiosity and thus predict demand for information when people are exposed to an information gap. First, curiosity depends on the salience of the information gap, i.e., the degree to which an unanswered question (or questions) is highlighted by exogenous contextual factors. Second, curiosity depends on the importance of the information gap, i.e., the extent to which one's utility depends on the missing information. Third, curiosity depends on the surprisingness of any information already received,i.e., the degree to which one's existing beliefs have recently been changed by preliminary information. The basic conceptual premise is that any question that attracts attention induces curiosity. Moving beyond information-gap theory, Berger (2013) suggests that emotional information tends to attract more attention and spread more quickly through social media. A separate literature on valence (e.g., Ito, Larsen, Smith, & Cacioppo, 1998; Rozin & Royzman, 2001) also suggests that negative information tends to attract more attention. While people tend to avoid negative information about themselves (Golman et al., 2017), they may be attracted to negative information in the news (Robertson et al., 2023; Soroka, Fournier, & Nir, 2019; Trussler & Soroka, 2014). We now operationalize these concepts for our specific setting where people choose whether or not to read news articles based on headlines, aiming to test whether variables capturing information-gap salience, importance, surprisingness, and negativity predict news consumption.[4]

[4]Berger and Milkman (2012) and Berger, Moe, and Schweidel (2023) suggest that content that evokes high emotional arousal is more likely to spread virally and to hold attention and sustain engagement. Following this line of work, we also conduct an additional analysis adding a variable measuring the use of high arousal words in the headlines as an additional predictor, and we present these results in the Appendix.

Author Manuscript

3.1 Indicators of Salient Information Gaps

A news article offers readers information they presumably do not already know, but a headline only creates a salient information gap if it makes them aware of what they do not know. Explicitly asking a question makes that question more salient, because it brings the question to mind,i.e., opens an information gap. Therefore, we identify headlines with a question mark as a proxy to indicate whether a headline creates a salient information gap. Similarly, an ellipsis may raise a question about the rest of the headline, which is hidden from the reader. Thus, this indicator variable may also indicate the presence of a salient information gap.

3.2 Indicators of Importance

Readers typically rely on editorial choices as cues of importance. An exclamation mark is usually used to express strong feelings and can thus make a headline feel more important. This is our first proxy for importance. Additionally, when posting multiple news articles, an account holder makes an editorial decision about the position of the headlines. Typically, more important headlines are placed above less important headlines. This is analogous to the demand for higher positions on consumer search websites. We use the published position of each headline as a second proxy for the importance of an article (although we do recognize that headlines displayed in highest positions may be more salient as well as more important).[5]

3.3 Measuring Surprisingness

When people read a news headline, they may obtain some initial information, which may be more or less surprising, given their existing beliefs and the kinds of headlines they typically encounter. A headline could be surprising for (at least) two different reasons. The first reason a headline could be surprising is that it could reject

[5]There are also individual differences in perceptions of news importance, which have been shown to affect news consumption (Bolsen & Leeper, 2013; Iyengar, Hahn,Krosnick, & Walker, 2008), but which we cannot capture in our data.

For Peer Review

Author Manuscript

a reader's existing belief. Headlines classified as "negation" statements by regular expression mapping may reveal information that deviates from expectations.

Second, it might offer news about a topic that readers did not expect to encounter. We use our LDA topic model to create a measure of the surprisingness of a news topic in a particular headline, following Vosoughi, Roy, and Aral (2018). Specifically, we compute the Kullback-Leibler divergence (KL-divergence) of the distribution over topics in the headline relative to the baseline distribution of topics across all headlines:

$$D_{KL}(P|Q) = \sum_{x \in X} P(x) \log \left(\frac{P(x)}{Q(x)} \right),$$

where $P(x)$ is the LDA topic value for the particular news headline with respect to a given topic x (i.e., a measure of how much the headline aligns with that topic), and $Q(x)$ is the mean LDA topic value for topic x across all news headlines. This KL-divergence describeshow a particular headline's topic distribution is different from the general distribution. The underlying premise here is that the mean topic distribution across headlines captures the topics people have come to expect to encounter in the news. A larger KL-divergence means that the headline covers topics that are less commonly in the news, and thus more surprising to encounter. This gives us a second proxy variable for surprisingness.

For Peer Review

3.4 Measuring Valence

We include the sentiment value of headlines as another predictor of news consumption. Negatively valenced headlines, at least when they are not self-relevant, may attract more attention and thus induce more curiosity. Note that in the context of news consumption, a negatively valenced headline does not necessarily correspond to a negatively valenced belief upon reading the news article. People may experience schadenfreude when learning of bad outcomes for others (Myrick & Chen, 2021), and negativity in a headline could induce positively or negatively valenced beliefs, depending on the specific information in the article. Still, as long as head-

Author Manuscript

lines are not typically self-relevant, negative valence should attract more attention.

3.5 Validation of Our Measures

To validate our text-based measures of the existence of salient information gaps (indicated by the presence of a question mark and by the presence of an ellipsis), importance (indicated by the presence of an exclamation mark) and surprisingness (measured by classification as a negation and by KL-divergence), we conduct five separate pre-registered experiments, each with 100 subjects pre-screened to be fluent in Mandarin, on Prolific. For each experiment, we identify ten pairs of Mandarin headlines from WeChat that have similar semantic meanings[6] but differ on the variable of interest. Participants see all twenty headlines in a randomized order within each pair, and answer questions about their perceptions of the articles. To validate the two text-based indicators of the existence of salient information gaps, participants read the headlines and report in free response the questions that they hope the corresponding news article would answer. We count the number of reported questions as the perceived number of salient information gaps. We then check whether headlines with question marks or with ellipses generate a greater number of perceived salient information gaps. To validate our text-based indicator of information-gap importance, participants read the headlines and rate the importance of the information the article would offer on a five-point Likert scale (i.e., from not important at all to extremely important), and we check whether headlines with exclamation marks are perceived as more important. To validate the text-based measures of information-gap surprisingness, participants rate how surprising the headlines are on a five-point Likert scale (i.e., from not surprising at all to extremely surprising), and we check whether headlines classified as negations or higher in KL-divergence are perceived as more surprising.

We analyze each experiment by running a fixed effects regression for all twenty headlines, specifying the variable of interest as the primary predictor and including individual fixed effects and paired headline fixed effects, and clustering standard er-

[6]We compute semantic similarity using the SentenceTransformers package from SBERT.net.

For Peer Review

Author Manuscript

rors by individual. Table 10 in Appendix presents the regression results of the five experiments. Validating our measures of information-gap salience, importance, and surprisingness, our text-based measures are significant predictors of the corresponding perceptual measures, with $p < 0.05$ for the ellipsis, exclamation mark, negation, and KL-divergence, and $p = 0.056$ for the question mark.

3.6 Summary of Measures and Descriptive Statistics

Table 1 summarizes the measures and indicates how we obtain them. Note that we represent our second measure of importance, the position of the headline, as a vector of 7 indicator variables for each of the 7 possible positions rather than as a single value between 1 and 7, so we can obtain estimates of the distinct effect of each position. Table 2 reports descriptive statistics. [7] The average number of readings of a news article on any given day is 900, with a large standard deviation (above 5000) indicating that there is a fat tail produced by a small numberofarticles that many, many people read.

Table 1: Summary of Measures

Construct	Measure	Source
Salience	1: Question mark	Punctuation mapping
	2: Ellipsis	Punctuation mapping
Importance	1: Exclamation mark	Punctuation mapping
	2: Position	WeChat dataset
Surprisingness	1: Negation	Regular expression mapping
	2: KL-divergence	LDA
Valence	Sentiment of headline	SnowNLP

4 Estimation Strategy

We estimate a model of news consumption in which people's curiosity about a news article is stimulated by the existence of a salient information gap and its importance,

[7] There are fewer articles in lower positions because some accounts publish fewer than 7 articles.

Author Manuscript

Table 2: Descriptive Statistics

Statistics	Mean	St. Dev.	Min	Max
Binary variables			0	1
Question mark	0.138	0.345		
Ellipsis	0.045	0.208		
Exclamation mark	0.267	0.443		
Position: 1	0.458	0.498		
Position: 2	0.189	0.391		
Position: 3	0.117	0.321		
Position: 4	0.083	0.275		
Position: 5	0.063	0.243		
Position: 6	0.048	0.214		
Position: 7	0.043	0.203		
Negation	0.225	0.418		
Continuous variables				
KL-divergence	1.640	0.612	0.076	4.940
Sentiment of headline	0.673	0.346	0	1
Reading	901	5153	0	100,001
				N = 112,901

as well as the surprisingness and valence of the news article's headline. We fit consumption C_i of news article i (measured as the number of readings of the article) as a log-linear function of each of our measures of information-gap salience S_i, importance I_i, surprisingness R_i, valence V_i, a vector of control variables W_i, and an error term ε which captures individuals' specific tastes for reading news articles:

$$\ln(C_i + 1) = \sum_{j=1}^{2} \alpha_{1j} S_{ij} + \sum_{j=1}^{2} \alpha_{2j} I_{ij} + \sum_{j=1}^{2} \alpha_{3j} R_{ij} + \alpha_4 V_i + \beta W_i + \varepsilon.$$

(Note that the multiple measures of information-gap salience, importance, and surprisingness respectively are indexed with the subscript j. Also note that we represent our second measure of importance, position (i.e., I_{i2}), as a vector of seven indicator variables for the seven different positions rather than as an integer.) We always control for the source of each article. In our second specification, we also introduce controls for day of week and the three-hour interval in which each article

Author Manuscript

was published. In our preferred third specification, we additionally control for the explicit topic tags posted on WeChat and the topics identified through LDA. We always cluster standard errors by source.

We specify a log-linear function for the number of readings because the distribution of the number of readings has a fat tail,i.e., there are a few articles that are read by a large number of people. Because our dependent variable is a count, we also estimate a negative binomial model, which accommodates zeros and extreme values in the fat tail (Cameron & Trivedi, 2013).

Controlling for the source of each article is necessary because different accounts have different numbers of subscribers, so the potential audience for any article depends on its source. Moreover, controlling for sources also absorbs any variation that might be due to expected differences in the quality of the articles based on their sources. Controlling for publication time (i.e., published day of week, and published hour interval) also helps control for different potential audience size at different times. Controlling for topics (i.e., explicit topic tags as well as inferred LDA topics) allows us to capture differential interest in different news topics.

Because alternative theories of curiosity suggest that curiosity might be concave in surprisingness (Berlyne, 1966; Dubey & Griffiths, 2020), in the Appendix we also estimate a regression model with a quadratic term of KL-divergence (i.e., R_{i2}^2):

$$\ln(C_i+1) = \sum_{j=1}^{2} \alpha_{1j} S_{ij} + \sum_{j=1}^{2} \alpha_{2j} I_{ij} + \sum_{j=1}^{2} \alpha_{3j} R_{ij} + \alpha_{42} R_{i2}^2 + \alpha_5 V_i + \beta W_i + \varepsilon .$$

We cannot include a quadratic term of our other measure of surprisingness, i.e., negation, because it is a binary variable.

We do not think of our specific measures, such as the presence or absence of a question mark or exclamation mark, as directly causing greater or lesser news consumption, but as markers that are associated with the salience and importance of the information gaps created by the headlines, which we believe are causal drivers of curiosity for news. That is, if an editor were to simply add an unnecessary exclamation mark into a headline where it would not serve to emphasize the importance of

Author Manuscript

the news, it would not have the same effect on news consumption. Indeed, the large number of headlines with exclamation marks (more than a quarter of the sample) suggests that some editors may have recognized the correlation between exclamation mark usage and readership, and may now use exclamation marks liberally. To the extent that this weakens the association between importance and the presence of an exclamation mark, or even creates an association for readers between exclamation marks and clickbait, this would diminish the effect we observe. More generally, features of headlines are not randomly assigned, but are written by editors, so we have to be concerned about possible confounding explanations for any relationship between news consumption and our measures of information-gap salience, importance, surprisingness, and valence. Our control variables in our regression should take care of most plausible confounds. For example, while it may be possible that some particular sources are more likely to use exclamation marks than others, and these sources may differ systematically in the quality of articles they post and the readership they generate, controlling for sources means that we are identifying effects based on the variation between articles from the same sources.

5 Empirical Results

Table 3 presents the results of three OLS regression models and three negative binomial regression models that explore the relationship between news consumption and our measures of information-gap salience, importance, surprisingness, and valence of the headlines. Broadly, we obtain highly significant, positive effects of these measures of information-gap salience, importance, and surprisingness, as well as a significant negative effect of valence. In column (1) and (4) we control only for published sources. Column (2) and (5) include controls for published hour interval and published day of the week. In column (3) and (6), we add controls for topic tags and LDA topic values.

We interpret the regression results focusing on column (3), where we estimate the log-linear model including all the controls. Supporting our hypothesis that a salient information gap stimulates curiosity, we find that articles whose headlines

For Peer Review

Author Manuscript

Table 3: OLS and Negative Binomial Regression Results

Model:	OLS			Negative Binomial		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	*ln (Reading+1)*			*Reading*		
Question mark	0.029***	0.029***	0.034***	0.031**	0.031**	0.036***
	(0.010)	(0.010)	(0.010)	(0.014)	(0.014)	(0.014)
Ellipsis	0.087***	0.087***	0.097***	0.089***	0.089***	0.097***
	(0.020)	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)
Exclamation	0.067***	0.067***	0.070***	0.076***	0.076***	0.079***
	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
Position 1	0.691***	0.691***	0.688***	0.918***	0.918***	0.913***
	(0.016)	(0.016)	(0.016)	(0.028)	(0.028)	(0.028)
Position 2	0.075***	0.075***	0.075***	0.175***	0.176***	0.174***
	(0.014)	(0.014)	(0.014)	(0.025)	(0.025)	(0.025)
Position 3	-0.010	-0.009	-0.009	0.028	0.028	0.027
	(0.014)	(0.014)	(0.014)	(0.025)	(0.025)	(0.025)
Position 4	-0.034**	-0.034**	-0.033**	-0.038	-0.038	-0.038
	(0.014)	(0.014)	(0.014)	(0.026)	(0.026)	(0.026)
Position 5	-0.062***	-0.062***	-0.062***	-0.080***	-0.081***	-0.081***
	(0.014)	(0.014)	(0.014)	(0.027)	(0.027)	(0.027)
Position 6	-0.049***	-0.049***	-0.051***	-0.060**	-0.060**	-0.064**
	(0.014)	(0.014)	(0.014)	(0.027)	(0.027)	(0.027)
Negation	0.020***	0.020***	0.019***	0.027***	0.027***	0.024**
	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
KL-divergence	0.015***	0.015***	0.036***	0.018***	0.018***	0.038***
	(0.005)	(0.005)	(0.009)	(0.006)	(0.006)	(0.012)
Sentiment	-0.074***	-0.074***	-0.065***	-0.086***	-0.086***	-0.073***
	(0.010)	(0.010)	(0.010)	(0.013)	(0.013)	(0.013)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

For Peer Review

Author Manuscript

contain question marks are read 3.4 percent more often. Supporting our hypothesis that importance stimulates curiosity, we find that articles in the highest position (i.e., position 1) are read almost twice as often (99 percent more often) than articles in the lowest position (i.e., position 7), and articles are read less often as the position decreases from position 1 to position 6. Interestingly, articles in positions 4, 5, and 6 are actually read less often than articles in position 7, suggesting that the lowest position may actually attract some attention, perhaps due to the absence of subsequent articles that could serve as distractors. Supporting our hypothesis that surprising headlines stimulate curiosity, we find that an article with a one-standard-deviation increase in KL-divergence is read 3.6 percent more often. (Looking at Table 4 column (3) in Appendix, we do find that news consumption is concave in this measure of surprisingness, so an increase in KL-divergence may have a bigger effect when KL-divergence is low and a smaller effect when it is already high.) Consistent with earlier findings of negativity bias, we find that an article with a one-standard-deviation decrease in the sentiment value of the headline is read 6.8 percent more often.

Our findings survive with or without controls for topics and publication time, indicating that our results are robust to a variety of specifications. The regression results using the negative binomial model (shown in Table 3 columns 4-6) are also highly similar to the OLS regression results.

As a robustness check, we redo our analysis using alternative valence and arousal measures obtained using cntext, as discussed in Section 2.2.5, in place of the SnowNLP sentiment measure. This additional analysis, presented in Table 9 in Appendix, produces a similar pattern with respect to the effect of valence on news consumption. We find no significant effect of the arousal measure on news consumption. While existing literature, such as Berger and Milkman (2012) and Berger et al. (2023), suggests that emotionally arousing text contributes to content virality, we do not detect such an effect in our context.

We also explore which particular topics are read more or less often by looking at the topic tag controls in the regression results. (See Table 11 in Appendix.) The coefficient of the topic tag "politics" is significantly positive, suggesting that some

For Peer Review

Author Manuscript

news consumption may be partially driven by instrumental value, as people find it useful to know what their government is doing. At the same time, the topic tag "celebrity" also has a significantly positive effect on news consumption, suggesting that people also sometimes read news solely to satisfy curiosity. To explore the relationship between the usefulness of a news topic and consumption of news articles on that topic, we collected additional data from 50 subjects on Prolific, asking them to rate the usefulness of 36 different news topics on a Likert scale from 0 to 10. We test whether the perceived usefulness of news topics is positively correlated with the coefficients of these topic tags in the regression results from the observational WeChat data. The Spearman correlation coefficient of 0.34 (with $p = 0.043$) confirms a positive relationship between the perceived usefulness of news topics and how often articles on each particular topic tend to be read. This finding aligns with existing research indicating that the practical value of information is a key factor in how quickly it spreads through social media (Berger, 2013; Berger & Milkman, 2012).

We also have data on the number of "likes" for each article, which allows us to test whether an article that gets more readings also generates a larger fraction of likes from its readers. To avoid the confounding influence that some accounts may also give a like to their own articles, we only consider articles with more than 100 readings here (so that the effect of liking one's own article is negligible). The Spearman correlation coefficient between the number of readings and the fraction of readers that like the article is -0.08 (with $p < 0.001$), indicating a weak relationship between attracting readers and satisfying readers. Headlines that draw in more readers are not necessarily associated with content that readers actually like. There are many potential explanations for the absence of a relationship between reading and liking, including individual differences in breadth vs depth of interest in news topics as well as the possibility that clickbait headlines attract readers but are ultimately unsatisfying. It is also consistent with the finding that satisfying curiosity is often disappointing (Loewenstein, 1994), which may be a consequence of the belief-resolution effect in information-gap theory (Golman & Loewenstein, 2018).

In Appendix we also report results from an analysis of news popularity for ap-

For Peer Review

Author Manuscript

proximately 81,000 news articles concerning four topics (i.e., Economy, Microsoft, Obama and Palestine) derived from more than 5,000 sources and published on Facebook from November 2015 to July 2016 (Moniz & Torgo, 2018). (See Tables 12, 13, 14 and 15 in Appendix.) Unfortunately, this dataset does not include the specific number of readings for each published article, so this analysis relies on the popularity of the article on Facebook (number of likes) as the dependent variable measure.

The results are largely consistent with the results for the WeChat dataset presented here with one notable difference: question marks are actually associated with lower popularity in the Facebook data. However, we also make use of the NPS Chat Corpus Model (Forsythand & Martell, 2007) to identify whether a headline poses a question (as well as to identify negations along with headlines that add emphasis). We do find that headlines that pose a wh-question are more popular. While we cannot be sure why question marks have different effects in the different analyses, our conjecture is that question marks are primarily used to indicate questions in Mandarin (a language in which the same word can be used for a question word such as "who" or an existential classifier such as "anybody" (Cheng, 1994)), but question marks may be used to indicate rhetorical questions in English (when question words alone already reliably indicate questions).

For Peer Review

6 Discussion

By applying NLP methods to a unique dataset on news consumption, we test information-gap theory's prediction that curiosity depends on the degree to which a question (or information gap) seems salient, important, and surprising. We additionally explore the effect of a headline's valence,in light of earlier results indicating a negativity bias in news consumption. Our large dataset allows us to control for a number of other factors that influence news consumption and still have power to detect subtle effects of our variables of interest. We find that people are more likely to consume news when: 1) the headline raises a salient question, indicated by the presence of a question mark or ellipsis in the headline; 2) the news appears more important, either

Author Manuscript

because the headline is displayed in a higher position on the webpage or because the headline includes an exclamation mark for emphasis; 3) the topics mentioned in the headline are more surprising, as measured by the relative entropy (KL-divergence) of the distribution over topics referenced in the headline relative to a baseline distribution over topics; and 4) the headline has lower (more negative) valence. These findings suggest that information-gap theory, along with negativity bias, generally successfully predicts aggregate news consumption patterns.

A few other recent papers have explored the effects of headlines on news consumption, with mixed results on the effects of questions in headlines. Banerjee and Urminsky (2021) analyze thousands of A/B experiments conducted on Upworthy and find that questions in headlines decrease the number of people reading the article, whereas Lai and Farbrot (2014) and She and Zhang (2019), like us, find that headlines with questions are read more often.[8] There may be a tradeoff in that explicit questions may open information gaps, but may also make headlines appear less informative (Scacco & Muddiman, 2020). It is possible that the presence of a question in a headline could interact with other features of a headline such that it stimulates news consumption in some contexts but decreases engagement in other contexts.

Our finding that the surprisingness of news headlines contributes to people reading, or at least clicking on, those articles is consistent with a few different lines of research on the effects of surprisal. One line of research highlights the role of surprisingness in attracting attention, including in perceptual salience (Horstmann & Herwig, 2015; Itti & Baldi, 2009; Taylor & Fiske, 1978). Another line of research connects surprisal to learning and memory (Foster & Keane, 2019; Kang et al., 2009; Rescorla & Wagner, 1972; Vogl, Pekrun, Murayama, & Loderer, 2020). Indeed, we believe that surprising information stimulates curiosity, learning, and memory precisely because it grabs attention.

Our results further contribute to our understanding of curiosity. Whereas most psychological research on curiosity relies on laboratory studies, and often self-

[8]Berger (2023) also highlights the efficacy of posing questions in engaging audiences in interpersonal communication.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Author Manuscript

reports of feeling curious, our analysis of observed field behavior, i.e., the number of readings of articles on a social media platform, relies on a behavioral measure of information-seeking in a naturalistic environment. This naturalistic field evidence complements experimental findings (e.g., Chen & He, 2021; Falk & Zimmermann, 2016; Golman et al., 2021; Kang et al., 2009; Litman, Hutchins, & Russon, 2005; Singh & Manjaly, 2021; Wright, Clarkson, & Kardes, 2018) in providing support for the information-gap theory of curiosity. The predictions of information-gap theory are transferable across domains. For example, teachers can try to open information gaps that seem more salient, important, and surprising, to improve children's learning incentives and help children stay more focused in class, before providing the information they want to impart.

Our results also contribute to our understanding of news consumption, providing a clear demonstration that curiosity plays a role in determining what news people choose to read. In a paper on social media and fake news, Allcott and Gentzkow (2017) highlight the need to better understand the process by which "consumers see 'headlines' and then decide whether to 'click' to learn more detail." Curiosity is clearly part of this process, but there are still open questions about how curiosity interacts with confirmation bias and identity-threat avoidance. Additionally, we should expect strategic editors to exploit curiosity-inducing tactics to reach larger audiences, and it is unclear whether and how sophisticated readers might react to such tactics, especially if the headlines are perceived as clickbait. We have only uncovered part of the story about how people decide what news to consume.

For Peer Review

Author Manuscript

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, *31*(2), 211–36.

Bakshy, E., Messing, S., & Adamic, L. A. (2015, June). Exposure to ideologically diverse news and opinion on Facebook. *Science (New York,N.Y.)*, *348*(6239), 1130–1132. doi: 10.1126/science.aaa1160

Banerjee, A., & Urminsky, O. (2021). The language that drives engagement: A systematic large-scale analysis of headline experiments. *Available at SSRN 3770366*.

Berger, J. (2013). *Contagious: How to build word of mouth in the digital age*. Simon and Schuster.

Berger, J. (2023). *Magic words*. Bentang Pustaka.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of marketing research*, *49*(2), 192–205.

Berger, J., Moe, W. W., & Schweidel, D. A. (2023). What holds attention? linguistic drivers of engagement. *Journal of Marketing*, 00222429231152880.

Berlyne, D. E. (1966, July). Curiosity and Exploration. *Science (New York, N.Y.)*, *153*(3731), 25–33. doi: 10.1126/science.153.3731.25

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003, March). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*(null), 993–1022.

Blom, J. N., & Hansen, K. R. (2015, January). Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, *76*, 87–100. doi: 10.1016/j.pragma.2014.11.010

Bode, L., Vraga, E. K., & Troller-Renfree, S. (2017, April). Skipping politics: Measuring avoidance of political content in social media. *Research & Politics*, *4*(2), 2053168017702990. doi: 10.1177/2053168017702990

Bolsen, T., & Leeper, T. J. (2013, July). Self-Interest and Attention to News Among Issue Publics. *Political Communication*, *30*(3), 329–348. doi: 10.1080/10584609.2012.737428

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge university press.

For Peer Review

Author Manuscript

Chen, Y., & He, Y. (2021). Information acquisition and provision in school choice: an experimental study. *Journal of Economic Theory*, *197*, 105345.

Cheng, L. L.-s. (1994). Wh-words as polarity items. *Chinese Languages and Linguistics*, 615–640.

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021, March). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9). doi: 10.1073/pnas.2023301118

Dubey, R., & Griffiths, T. L. (2020). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological Review*, *127*(3), 455–476. doi: 10.1037/rev0000175

Falk, A., & Zimmermann, F. (2016). Beliefs and utility: Experimental evidence on preferences for information.

Forsythand,E. N., & Martell,C. H. (2007). Lexical and discourse analysis of online chat dialog. In *International conference on semantic computing (icsc 2007)* (pp. 19–26).

Foster, M. I., & Keane, M. T. (2019). The role of surprise in learning: Different surprising outcomes affect memorability differentially. *Topics in cognitive science*, *11*(1), 75–87.

Gentzkow, M., & Shapiro, J. M. (2004). Media,education and anti-americanism in the muslim world. *Journal of Economic perspectives*, *18*(3), 117–133.

Gentzkow, M., & Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, *78*(1), 35–71. doi: 10.3982/ECTA7195

Golman, R., Hagmann, D., & Loewenstein, G. (2017, March). Information Avoidance. *Journal of Economic Literature*, *55*(1), 96–135. doi: 10.1257/jel.20151245

Golman, R., & Loewenstein, G. (2018). Information gaps: A theory of preferences regarding the presence and absence of information. *Decision*, *5*(3), 143–164. doi: 10.1037/dec0000068

Golman, R., Loewenstein, G., Moene, K. O., & Zarri, L. (2016, September). The Preference for Belief Consonance. *Journal of Economic Perspectives*, *30*(3), 165–188. doi: 10.1257/jep.30.3.165

For Peer Review

Author Manuscript

Golman, R., Loewenstein, G., Molnar, A., & Saccardo, S. (2021, December). The Demand for, and Avoidance of, Information. *Management Science*. doi: 10.1287/mnsc.2021.4244

Hamilton, J. T. (2004). *All the News That's Fit to Sell: How the Market Transforms Information into News*. Princeton University Press. doi: 10.2307/j.ctt7smgs

Hancock, J. T., Landrigan, C., & Silver, C. (2007, April). Expressing emotion in text-based communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 929–932). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/1240624.1240764

Hilchey, M. D., Rondina, R., & Soman, D. (2022). Information-seeking when information doesn't matter. *Journal of Behavioral Decision Making*.

Horstmann, G., & Herwig, A. (2015). Surprise attracts the eyes and binds the gaze. *Psychonomic Bulletin & Review*, *22*, 743–749.

Hosseinmardi, H., Ghasemian, A., Clauset, A., Mobius, M., Rothschild, D. M., & Watts, D. J. (2021, August). Examining the consumption of radical content on YouTube. *Proceedings of the National Academy of Sciences*, *118*(32). doi: 10.1073/pnas.2101967118

Hsee, C. K., & Ruan, B. (2016). The pandora effect: The power and peril of curiosity. *Psychological science*, *27*(5), 659–666.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887–900. doi: 10.1037/0022-3514.75.4.887

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306.

Iyengar, S., Hahn, K. S., Krosnick, J. A., & Walker, J. (2008). Selective exposure to campaign communication: The role of anticipated agreement and issue public membership. *The Journal of Politics*, *70*(1), 186–200. doi: 10.1017/S0022381607080139

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012, October). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10),

For Peer Review

732–735. doi: 10.1038/nclimate1547

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological science*, *20*(8), 963–973.

Kardes, F. R., Herr, P. M., & Schwarz, N. (2019). *Handbook of research methods in consumer psychology*. Routledge.

Lai, L., & Farbrot, A. (2014). What makes you click? the effect of question headlines on readership in computer-mediated communication. *Social Influence*, *9*(4), 289–299.

Litman, J., Hutchins, T., & Russon, R. (2005, June). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition and Emotion*, *19*(4), 559–582. doi: 10.1080/02699930441000427

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98. doi: 10.1037/0033-2909.116.1.75

Moniz, N., & Torgo, L. (2018, January). Multi-Source Social Feedback of Online News Feeds. *arXiv:1801.07055 [cs]*.

Mullainathan, S., & Shleifer, A. (2005, September). The Market for News. *American Economic Review*, *95*(4), 1031–1053. doi: 10.1257/0002828054825619

Myrick, J. G., & Chen, J. (2021, April). Schadenfreude After Watching the News: How Audiences Respond to Media Coverage of Partisans Disclosing Illnesses. *Journalism & Mass Communication Quarterly*, 10776990211008534. doi: 10.1177/10776990211008534

Nickerson, R. S. (1998, June). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, *2*(2), 175–220. doi: 10.1037/1089-2680.2.2.175

Pennycook, G. (2023). A framework for understanding reasoning errors: From fake news to climate change and beyond. *Advances in experimental social psychology*, *67*(1), 1–85.

Provine, R. R., Spencer, R. J., & Mandell, D. L. (2007, September). Emotional Expression Online: Emoticons Punctuate Website Text Messages.

Author Manuscript For Peer Review

Journal of Language and Social Psychology, *26*(3), 299–307. doi: 10.1177/0261927X06303481

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. *Classical Conditioning, Current Research and Theory*, *2*, 64–69.

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, 1–11.

Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). The evaluative lexicon 2.0: The measurement of emotionality, extremity, and valence in language. *Behavior research methods*, *50*, 1327–1344.

Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399–408). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/2684822.2685324

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. doi: 10.1207/S15327957PSPR05042

Scacco, J. M., & Muddiman, A. (2020). The curiosity effect: Information seeking in the contemporary news environment. *New Media & Society*, *22*(3), 429–448.

Scott, K. (2021, April). You won't believe what's in this paper! Clickbait, relevance and the curiosity gap. *Journal of Pragmatics*, *175*, 53–66. doi: 10.1016/j.pragma.2020.12.023

Sharot, T., & Sunstein, C. R. (2020, January). How people decide what they want to know. *Nature Human Behaviour*, *4*(1), 14–19. doi: 10.1038/s41562-019-0793-1

She, J., & Zhang, T. (2019). The impact of headline features on the attraction of online financial articles. *International Journal of Web Information Systems*, *15*(5), 510–534.

Singh, A., & Manjaly, J. A. (2021). The effect of information gap and uncertainty

Author Manuscript
For Peer Review

on curiosity and its resolution. *Journal of Cognitive Psychology*, *33*(4), 403–423. doi: 10.1080/20445911.2021.1908311

Skovsgaard, M., Shehata, A., & Strömbäck, J. (2016, October). Opportunity Structures for Selective Exposure: Investigating Selective Exposure and Learning in Swedish Election Campaigns Using Panel Survey Data. *The International Journal of Press/Politics*, *21*(4), 527–546. doi: 10.1177/1940161216658157

Soroka, S., Fournier, P., & Nir, L. (2019, September). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*, *116*(38), 18888–18892. doi: 10.1073/pnas.1908369116

Sunstein, C. R. (2017). A prison of our own design: divided democracy in the age of social media. *Democratic Audit UK*.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, *50*(3), 755–769.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In *Advances in experimental social psychology* (Vol. 11, pp. 249–288). Elsevier.

Trussler, M., & Soroka, S. (2014, July). Consumer Demand for Cynical and Negative News Frames. *The International Journal of Press/Politics*, *19*(3), 360–379. doi: 10.1177/1940161214524832

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in cognitive sciences*, *22*(3), 213–224.

Van Dijk, E., & Zeelenberg, M. (2007). When curiosity killed regret: Avoiding or seeking the unknown in decision-making under uncertainty. *Journal of Experimental Social Psychology*, *43*(4), 656–662.

Venneti,L., & Alam,A. (2018, June). How Curiosity can be modeled for a Clickbait Detector. *arXiv:1806.04212 [cs]*.

Vogl, E., Pekrun, R., Murayama, K., & Loderer, K. (2020). Surprised–curious–confused: Epistemic emotions and knowledge exploration. *Emotion*, *20*(4), 625.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151.

Author Manuscript For Peer Review

Watts, D. J., Rothschild, D. M., & Mobius, M. (2021, April). Measuring the news and its impact on democracy. *Proceedings of the National Academy of Sciences*, *118*(15). doi: 10.1073/pnas.1912443118

Wright, S. A., Clarkson, J. J., & Kardes, F. R. (2018). Circumventing resistance to novel information: Piquing curiosity through strategic information revelation. *Journal of Experimental Social Psychology*, *76*,

Author Manuscript
For Peer Review

Author Manuscript

For Peer Review

Author Manuscript

Appendix (For Online Publication Only)



Figure 1: WeChat Webpage

Author Manuscript

Table 1: OLS Regression Results (Excluding Duplicate Headlines)

Dependent Variable:	*ln* (Reading+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.0289***	0.0287***	0.0328***	0.0292***	0.0290***	0.0331***
	(0.0100)	(0.0100)	(0.0101)	(0.0100)	(0.0100)	(0.0101)
Ellipsis	0.0936***	0.0938***	0.1027***	0.0944***	0.0945***	0.1038***
	(0.0193)	(0.0193)	(0.0194)	(0.0193)	(0.0193)	(0.0194)
Exclamation mark	0.0731***	0.0730***	0.0752***	0.0727***	0.0726***	0.0754***
	(0.0092)	(0.0092)	(0.0092)	(0.0092)	(0.0092)	(0.0092)
Position 1	0.6775***	0.6776***	0.6751***	0.6775***	0.6776***	0.6750***
	(0.0161)	(0.0161)	(0.0160)	(0.0161)	(0.0161)	(0.0160)
Position 2	0.0660***	0.0659***	0.0661***	0.0659***	0.0658***	0.0660***
	(0.0142)	(0.0142)	(0.0142)	(0.0142)	(0.0142)	(0.0142)
Position 3	-0.0139	-0.0140	-0.0135	-0.0140	-0.0141	-0.0135
	(0.0137)	(0.0137)	(0.0137)	(0.0137)	(0.0137)	(0.0137)
Position 4	-0.0359**	-0.0359**	-0.0349**	-0.0360***	-0.0360***	-0.0350**
	(0.0140)	(0.0140)	(0.0140)	(0.0140)	(0.0140)	(0.0140)
Position 5	-0.0565***	-0.0566***	-0.0566***	-0.0565***	-0.0566***	-0.0566***
	(0.0143)	(0.0143)	(0.0143)	(0.0143)	(0.0143)	(0.0143)
Position 6	-0.0493***	-0.0494***	-0.0512***	-0.0491***	-0.0492***	-0.0510***
	(0.0146)	(0.0146)	(0.0146)	(0.0146)	(0.0146)	(0.0146)
Negation	0.0187***	0.0187***	0.0178**	0.0187***	0.0187***	0.0182**
	(0.0072)	(0.0072)	(0.0073)	(0.0072)	(0.0072)	(0.0073)
KL-divergence	0.0165***	0.0165***	0.0386***	0.0642***	0.0637***	0.1088***
	(0.0051)	(0.0051)	(0.0089)	(0.0191)	(0.0191)	(0.0260)
Sentiment	-0.0770***	-0.0769***	-0.0683***	-0.0770***	-0.0769***	-0.0674***
	(0.0096)	(0.0095)	(0.0096)	(0.0096)	(0.0095)	(0.0096)
Quadratic KL-divergence				-0.0128***	-0.0126**	-0.0179***
				(0.0049)	(0.0049)	(0.0061)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	108,513	108,513	108,513	108,513	108,513	108,513

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

For Peer Review

Author Manuscript

Table 2: Negative Binomial Regression Results (Excluding Duplicate Headlines)

Dependent Variable:	*R*eading					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.0292**	0.0290**	0.0340**	0.0295**	0.0293**	0.0345**
	(0.0137)	(0.0137)	(0.0139)	(0.0137)	(0.0137)	(0.0139)
Ellipsis	0.0957***	0.0957***	0.1040***	0.0966***	0.0965***	0.1051***
	(0.0201)	(0.0201)	(0.0205)	(0.0201)	(0.0201)	(0.0205)
Exclamation mark	0.0830***	0.0828***	0.0848***	0.0827***	0.0825***	0.0852***
	(0.0113)	(0.0113)	(0.0113)	(0.0113)	(0.0113)	(0.0113)
Position 1	0.8979***	0.8981***	0.8929***	0.8980***	0.8982***	0.8932***
	(0.0275)	(0.0275)	(0.0273)	(0.0275)	(0.0275)	(0.0273)
Position 2	0.1658***	0.1659***	0.1641***	0.1658***	0.1659***	0.1643***
	(0.0251)	(0.0251)	(0.0251)	(0.0251)	(0.0251)	(0.0251)
Position 3	0.0241	0.0240	0.0234	0.0242	0.0242	0.0236
	(0.0249)	(0.0250)	(0.0249)	(0.0249)	(0.0249)	(0.0249)
Position 4	-0.0391	-0.0391	-0.0394	-0.0390	-0.0391	-0.0394
	(0.0254)	(0.0254)	(0.0254)	(0.0254)	(0.0254)	(0.0254)
Position 5	-0.0716***	-0.0722***	-0.0725***	-0.0713***	-0.0719***	-0.0722***
	(0.0267)	(0.0267)	(0.0266)	(0.0267)	(0.0267)	(0.0266)
Position 6	-0.0584**	-0.0583**	-0.0623**	-0.0579**	-0.0578**	-0.0618**
	(0.0270)	(0.0270)	(0.0270)	(0.0270)	(0.0270)	(0.0269)
Negation	0.0256***	0.0254***	0.0219**	0.0254***	0.0253***	0.0222**
	(0.0090)	(0.0090)	(0.0092)	(0.0090)	(0.0090)	(0.0092)
KL-divergence	0.0192***	0.0192***	0.0403***	0.0730***	0.0730***	0.1105***
	(0.0063)	(0.0063)	(0.0115)	(0.0235)	(0.0235)	(0.0333)
Sentiment	-0.0879***	-0.0878***	-0.0745***	-0.0879***	-0.0878***	-0.0734***
	(0.0122)	(0.0122)	(0.0123)	(0.0122)	(0.0122)	(0.0123)
Quadratic				-0.0144**	-0.0144**	-0.0179**
KL-divergence				(0.0061)	(0.0061)	(0.0077)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	108,363	108,363	108,363	108,363	108,363	108,363
Over-dispersion	3.6978	3.6988	3.7282	3.6982	3.6991	3.7286

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. Note: 42 fixed-effects (150 observations) removed because of only 0 outcomes.

For Peer Review

Author Manuscript

Table 3: OLS Regression Results (Imputing 171,000 as the Maximum Readings)

Dependent Variable:	*l*n (Reading+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.0293***	0.0291***	0.0337***	0.0296***	0.0294***	0.0341***
	(0.0101)	(0.0101)	(0.0102)	(0.0101)	(0.0101)	(0.0102)
Ellipsis	0.0867***	0.0872***	0.0965***	0.0877***	0.0881***	0.0979***
	(0.0200)	(0.0200)	(0.0201)	(0.0200)	(0.0199)	(0.0201)
Exclamation mark	0.0673***	0.0672***	0.0698***	0.0669***	0.0668***	0.0701***
	(0.0094)	(0.0094)	(0.0094)	(0.0094)	(0.0094)	(0.0094)
Position 1	0.6909***	0.6911***	0.6888***	0.6909***	0.6911***	0.6888***
	(0.0162)	(0.0162)	(0.0161)	(0.0162)	(0.0162)	(0.0161)
Position 2	0.0755***	0.0756***	0.0758***	0.0754***	0.0755***	0.0757***
	(0.0144)	(0.0144)	(0.0144)	(0.0144)	(0.0144)	(0.0143)
Position 3	-0.0081	-0.0081	-0.0077	-0.0082	-0.0082	-0.0078
	(0.0139)	(0.0139)	(0.0138)	(0.0139)	(0.0139)	(0.0138)
Position 4	-0.0333**	-0.0332**	-0.0325**	-0.0335**	-0.0333**	-0.0327**
	(0.0140)	(0.0140)	(0.0140)	(0.0140)	(0.0140)	(0.0140)
Position 5	-0.0614***	-0.0614***	-0.0613***	-0.0615***	-0.0614***	-0.0615***
	(0.0144)	(0.0144)	(0.0143)	(0.0144)	(0.0144)	(0.0143)
Position 6	-0.0490***	-0.0489***	-0.0508***	-0.0488***	-0.0487***	-0.0506***
	(0.0144)	(0.0144)	(0.0144)	(0.0144)	(0.0144)	(0.0144)
Negation	0.0208***	0.0207***	0.0202***	0.0207***	0.0207***	0.0206***
	(0.0073)	(0.0073)	(0.0074)	(0.0073)	(0.0073)	(0.0073)
KL-divergence	0.0154***	0.0154***	0.0357***	0.0662***	0.0655***	0.1115***
	(0.0052)	(0.0052)	(0.0092)	(0.0196)	(0.0196)	(0.0269)
Sentiment	-0.0732***	-0.0731***	-0.0649***	-0.0731***	-0.0731***	-0.0639***
	(0.0097)	(0.0097)	(0.0098)	(0.0097)	(0.0097)	(0.0098)
Quadratic KL-divergence				-0.0136***	-0.0134***	-0.0193***
				(0.0051)	(0.0051)	(0.0063)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

For Peer Review

Author Manuscript

Table 4: Negative Binomial Regression Results (Imputing 171,000 as the Maximum Readings)

Dependent Variable:	*Reading*					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.0310**	0.0307**	0.0364***	0.0313**	0.0311**	0.0369***
	(0.0138)	(0.0138)	(0.0140)	(0.0138)	(0.0138)	(0.0140)
Ellipsis	0.0891***	0.0892***	0.0972***	0.0901***	0.0901***	0.0985***
	(0.0205)	(0.0205)	(0.0209)	(0.0205)	(0.0205)	(0.0209)
Exclamation mark	0.0760***	0.0759***	0.0784***	0.0757***	0.0755***	0.0788***
	(0.0115)	(0.0115)	(0.0115)	(0.0115)	(0.0115)	(0.0115)
Position 1	0.9195***	0.9198***	0.9147***	0.9196***	0.9200***	0.9150***
	(0.0278)	(0.0278)	(0.0276)	(0.0278)	(0.0278)	(0.0276)
Position 2	0.1769***	0.1771***	0.1752***	0.1770***	0.1772***	0.1755***
	(0.0255)	(0.0255)	(0.0255)	(0.0255)	(0.0255)	(0.0255)
Position 3	0.0322	0.0323	0.0312	0.0323	0.0324	0.0314
	(0.0255)	(0.0256)	(0.0255)	(0.0255)	(0.0256)	(0.0255)
Position 4	-0.0366	-0.0365	-0.0373	-0.0366	-0.0365	-0.0373
	(0.0256)	(0.0257)	(0.0256)	(0.0256)	(0.0257)	(0.0256)
Position 5	-0.0791***	-0.0794***	-0.0799***	-0.0789***	-0.0792***	-0.0796***
	(0.0268)	(0.0268)	(0.0268)	(0.0268)	(0.0268)	(0.0268)
Position 6	-0.0600**	-0.0598**	-0.0639**	-0.0595**	-0.0593**	-0.0635**
	(0.0268)	(0.0268)	(0.0268)	(0.0268)	(0.0268)	(0.0268)
Negation	0.0289***	0.0287***	0.0255***	0.0286***	0.0285***	0.0259***
	(0.0092)	(0.0092)	(0.0094)	(0.0092)	(0.0092)	(0.0094)
KL-divergence	0.0181***	0.0181***	0.0381***	0.0775***	0.0774***	0.1158***
	(0.0064)	(0.0064)	(0.0117)	(0.0241)	(0.0240)	(0.0344)
Sentiment	-0.0846***	-0.0846***	-0.0714***	-0.0847***	-0.0846***	-0.0703***
	(0.0125)	(0.0126)	(0.0126)	(0.0126)	(0.0126)	(0.0126)
Quadratic KL-divergence				-0.0159**	-0.0158**	-0.0198**
				(0.0062)	(0.0062)	(0.0080)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901
Over-dispersion	3.5635	3.5645	3.5917	3.5639	3.5649	3.5922

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

For Peer Review

Table 5: Regression Results of the Five Validation Experiments

Dependent Variables:	Salience		Importance	Surprisingness	
	(1)	(2)	(3)	(4)	(5)
Question mark	0.057*				
	(0.029)				
Ellipsis		0.058**			
		(0.028)			
Exclamation mark			0.233***		
			(0.033)		
Negation				0.071**	
				(0.035)	
KL-divergence					0.110***
					(0.023)
Individual FE	Yes	Yes	Yes	Yes	Yes
Paired headline FE	Yes	Yes	Yes	Yes	Yes
Observations	2,000	2,000	2,000	2,000	2,000

Standard errors clustered at the individual level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript
For Peer Review

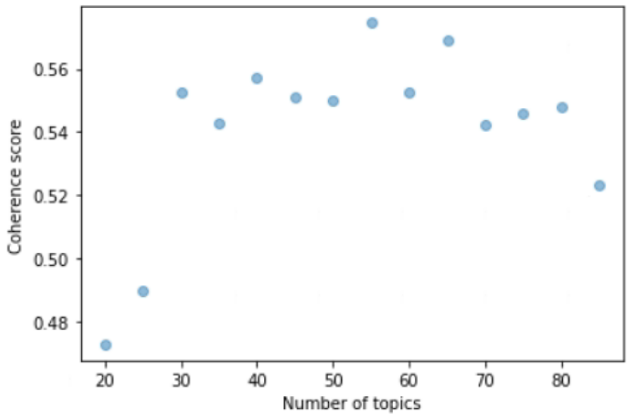


Figure 2: Scatterplot of coherence and number of topics (WeChat)

Author Manuscript
For Peer Review

Mandarin	非	别	不	没	无	勿	阻止	制止
English	not	don't	no	no	no	no	stop	stop
Mandarin	差	未	难	跌停	止	开除	莫	愁
English	poor	not	difficult	fall	stop	expel	no	sorrow
Mandarin	否	难	缺	乏	苦	艰	危	险
English	no	difficult	short of	short of	no	difficult	dangerous	risky

Figure 3: Negation words

Author Manuscript

For Peer Review

Author Manuscript

Table 6: Topic Tags' Regression Results of Specification (3)

Topic tag		Topic Tag	
Affection	0.0753	Agriculture and industry	0.0201
	(0.0487)		(0.1219)
Art and culture	0.1723*	Business and finance	0.0395
	(0.0883)		(0.0502)
Car	0.0239	Cartoon	-0.2921
	(0.1021)		(0.2275)
Celebrity	0.4049***	Documentary	0.3693*
	(0.1403)		(0.2201)
Education and training	0.0288	Fashion and skincare	0.1094*
	(0.0483)		(0.0625)
Fate and constellation	-0.1972	Fiction	-0.1197
	(0.3016)		(0.1331)
Food and catering	-0.0271	Game	0.0533
	(0.0688)		(0.1007)
Health	0.1067*	History	0.1320*
	(0.0601)		(0.0785)
Joke	-0.0519	Literature	0.0596
	(0.0713)		(0.0520)
Military	0.2410***	Marketing	0.0175
	(0.0915)		(0.4832)
Motivation	-0.0181	Movie	0.1104
	(0.1061)		(0.1060)
Music	0.0137	Parenting	-0.0559
	(0.0816)		(0.0777)
Pet	-0.1081	Photography	0.2969
	(0.1655)		(0.2942)
Politics	0.2269***	Real estate	0.0979
	(0.0616)		(0.0756)
Religion	-0.1386	Science	0.2200
	(0.1674)		(0.1346)
Society	0.1880***	Sports	0.1877
	(0.0423)		(0.1174)
Technology	0.0041	Travel	0.0634
	(0.0785)		(0.0710)
TV series	-0.0874	Workplace	0.1491
	(0.1315)		(0.0984)
Miscellaneous	0.0315		
	(0.0684)		

Standard errors clustered at the individual level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

For Peer Review

Table 7: Summary of Measures for Facebook Dataset

Construct	Measure	Source
Salience	1: Question mark	Punctuation
	2: Ellipsis	Punctuation
	3: Wh-question sentence	NPS Chat Corpus
Importance	1: Exclamation mark	Punctuation
	2: Emphasis sentence	NPS Chat Corpus
Surprisingness	1: Reject sentence	NPS Chat Corpus
	2: KL-divergence	LDA
Valence	Sentiment of headline	qdap R package

Author Manuscript
For Peer Review

Table 8: Descriptive Statistics for Facebook Dataset

Statistics	*Mean*	*St. Dev.*	*Min*	*Max*
Binary variables			0	1
Question mark	0.035	0.183		
Ellipsis	0.294	0.455		
Wh-question	0.054	0.226		
Exclamation mark	0.004	0.061		
Emphasis sentence	0.040	0.195		
Reject sentence	0.014	0.116		
Continuous variables				
KL-divergence	1.956	0.303	0.071	3.900
Sentiment of title	-0.006	0.137	-0.951	0.962
Facebook popularity	129.363	661.182	0	49211
				N = 81,637

Author Manuscript
For Peer Review

Table 9: OLS Regression Results for Facebook Data

Dependent Variable:	*ln* (Popularity+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	-0.0392	-0.0386	-0.1810***	-0.0395	-0.0389	-0.1812***
	(0.0362)	(0.0361)	(0.0638)	(0.0362)	(0.0361)	(0.0639)
Ellipsis	0.5930***	0.5689***	0.2146***	0.5902***	0.5660***	0.2144***
	(0.0143)	(0.0143)	(0.0323)	(0.0143)	(0.0143)	(0.0323)
Wh-question	0.5285***	0.5152***	0.1476***	0.5284***	0.5151***	0.1475***
	(0.0293)	(0.0292)	(0.0284)	(0.0293)	(0.0292)	(0.0284)
Exclamation mark	0.5322***	0.5114***	0.1805**	0.5350***	0.5141***	0.1813**
	(0.1070)	(0.1066)	(0.0851)	(0.1070)	(0.1065)	(0.0850)
Emphasis	0.1202***	0.1168***	0.0765***	0.1175***	0.1142***	0.0760***
	(0.0334)	(0.0332)	(0.0281)	(0.0334)	(0.0332)	(0.0280)
Reject	0.1885***	0.1935***	0.1294***	0.1911***	0.1962***	0.1291***
	(0.0558)	(0.0555)	(0.0450)	(0.0558)	(0.0555)	(0.0450)
KL-divergence	-0.1601***	-0.1701***	-0.0340	0.7053***	0.7060***	0.1806
	(0.0270)	(0.0269)	(0.0411)	(0.1442)	(0.1436)	(0.2050)
Sentiment	-0.4059***	-0.3978***	-0.2073***	-0.4086***	-0.4006***	-0.2078***
	(0.0475)	(0.0473)	(0.0348)	(0.0475)	(0.0473)	(0.0348)
Quadratic KL-divergence				-0.2127***	-0.2153***	-0.0510
				(0.0348)	(0.0347)	(0.0449)
LDA topic FE	No	No	Yes	No	No	Yes
Source FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
News topic FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	81,637	81,637	81,350	81,637	81,637	81,350

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. Note: In column (3) and column (6), where we add source fixed effects and LDA topic fixed effects, some observations are dropped because of the lack of sources of some articles in the Facebook dataset.

Author Manuscript For Peer Review

Author Manuscript

Table 10: Negative Binomial Regression Results for Facebook Data

Dependent Variable:	*Popularity*					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	-0.3544***	-0.3531***	-0.3503***	-0.3543***	-0.3529***	-0.3507***
	(0.0405)	(0.0405)	(0.0810)	(0.0405)	(0.0405)	(0.0810)
Ellipsis	0.3827***	0.3626***	0.2136***	0.3827***	0.3626***	0.2125***
	(0.0160)	(0.0161)	(0.0387)	(0.0160)	(0.0161)	(0.0386)
Wh-question	0.3078***	0.3123***	0.1656***	0.3061***	0.3114***	0.1668***
	(0.0327)	(0.0327)	(0.0446)	(0.0328)	(0.0328)	(0.0446)
Exclamation_mark	0.1639	0.1456	0.2227*	0.1646	0.1461	0.2234*
	(0.1196)	(0.1195)	(0.1211)	(0.1196)	(0.1195)	(0.1208)
Emphasis	0.2222***	0.2067***	0.1607***	0.2224***	0.2068***	0.1596***
	(0.0373)	(0.0373)	(0.0562)	(0.0373)	(0.0373)	(0.0561)
Reject	0.2811***	0.3136***	0.3496*	0.2816***	0.3139***	0.3480*
	(0.0625)	(0.0625)	(0.1889)	(0.0625)	(0.0626)	(0.1884)
KL-divergence	-0.0364	-0.0538*	0.0628	-0.1523	-0.1253	0.5839**
	(0.0289)	(0.0291)	(0.0530)	(0.1733)	(0.1744)	(0.2887)
Sentiment	-0.3876***	-0.3494***	-0.3301***	-0.3874***	-0.3493***	-0.3321***
	(0.0544)	(0.0544)	(0.0744)	(0.0544)	(0.0544)	(0.0745)
Quadratic				0.0271	0.0167	-0.1249*
KL-divergence				(0.0399)	(0.0402)	(0.0655)
LDA topic FE	No	No	Yes	No	No	Yes
Source FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
News topic FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	81,637	81,637	78,023	81,637	81,637	78,023
Over-dispersion	0.23467	0.23543	0.48534	0.23467	0.23543	0.48539

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. Note: In column (3) and column (6), where we add source fixed effects and LDA topic fixed effects, some observations are dropped because of the lack of sources of some articles in the Facebook dataset.

For Peer Review

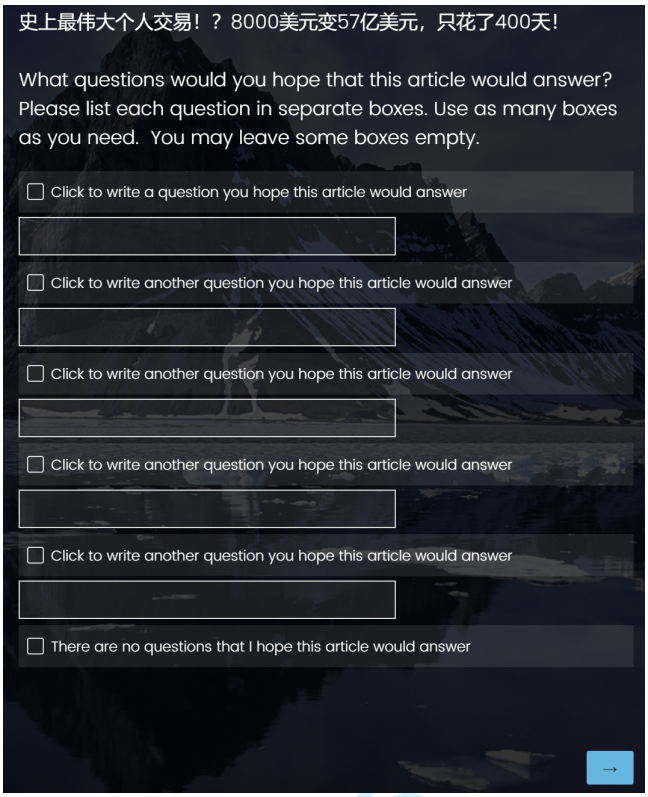


Figure 4: Screenshot of the information-gap salience validation experiment (measured by the presence of a question mark)

Author Manuscript
Under Review

Originally with question mark	Originally with question mark	Originally without question mark	Originally without question mark	Semantic similarity
史上最作大个人交易！？8000美元变57亿美元，只花了400天！	Greatest personal deal ever! ? $8,000 became $5.7 billion in just 400 days!	收益超330%！靠这个股票，巴菲特5年狂赚1200亿美元！	The income is over 330%! With this stock, Buffett earned $120 billion in 5 years!	0.452171296
取消SAT成绩要求？！加州大学录取现在看什么？	Dropping SAT score requirements? ! What does the University of California admission look like now?	加州：不许给差生打低分！为了教育公平！	California: Don't give poor students a low score! For education fairness!	0.522243023
全美首个或执行死刑的中国留学生，犯了什么罪？	What crime did the first Chinese student in the United States execute or execute?	江歌案宣判！美国留学生情杀案，也如此离奇！	The Jiang Ge case is sentenced! The murder case of American students is also so bizarre!	0.515769899
在美国这个职业，居然没有work life balance吗？	Is there no work life balance in this profession in the United States?	在美国工作没"大包"，我该如何开心的活着	How can I live happily if I don't have a "big bag" to work in the United States	0.483206153
美国将拿发200万张绿卡！排期还能交钱插队？！	The United States will issue 2 million green cards! Can I still pay to cut the queue? !	我酸了！特赦16万移民直接拿绿卡！	I'm sour! Amnesty for 160,000 immigrants to get green cards directly!	0.516782403
交了5个月房租，发现房东竟是我老婆！？	After paying 5 months rent, I found out that the landlord is my wife! ?	湾区condo抄底经历，1个月内买到自住房！	Bay Area condo bargain hunting experience, buy a house within 1 month!	0.481140703
什么副业既优雅，又有"钱"景？	What sideline is both elegant and "money"?	推荐一个很适合女性的副业	Recommend a side business that is very suitable for women	0.468031824
实习生重写大厂代码是什么体验？	What is the experience of the intern rewriting the code of the big factory?	分享analytics跳槽到大厂的经历和感想	Share the experience and impressions of analytics jumping to Dachang	0.458224714
为什么跳槽，肯定要比升上去的拿得多？	Why jump jobs, must get more than the promotion?	今年2次跳槽，工资从$4.5万涨到$13万！	Twice this year, the salary has increased from $45,000 to $130,000!	0.469226003
被面试官用十招坑死面试？反击攻略来啦！	Was interviewed by the interviewer with ten tricks? Counterattack strategy is here!	加拿大找工求职请避开这些坑！	Please avoid these pits when looking for a job in Canada!	0.439164698

Figure 5: Headlines in the information-gap salience validation experiment (measured by the presence of a question mark). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

For Peer Review

Author Manuscript

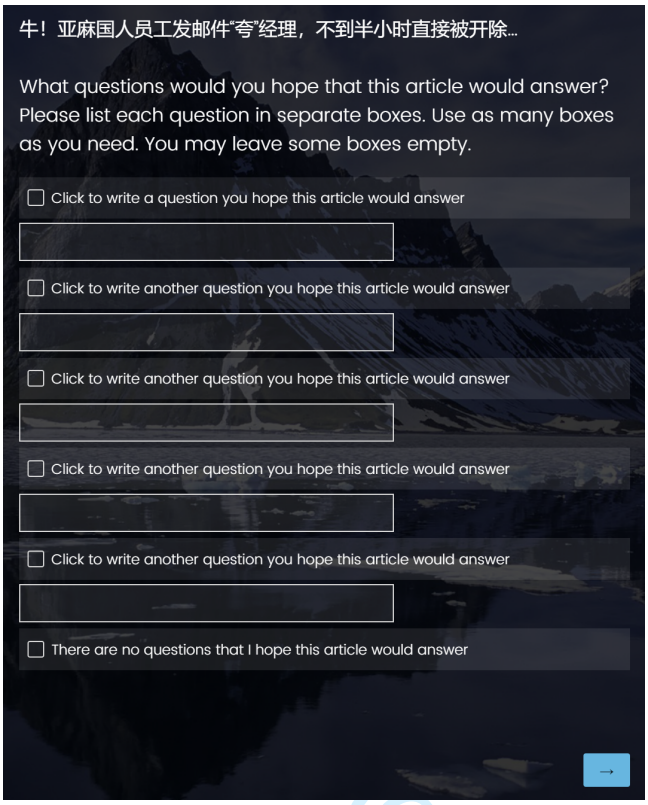


Figure 6: Screenshot of the information-gap salience validation experiment (measured by the presence of an ellipsis).

Author Manuscript
For Review

Originally with ellipsis	Originally with ellipsis	Originally without ellipsis	Originally without ellipsis	Semantic similarity
牛！亚麻国人员工发邮件"安"经理，不到半小时直接被开除…	Surprising! Flax workers sent emails to "praise" the manager and were fired directly in less than half an hour…	定了！老板下班后发信息给我，违法！	Deal! It's illegal for my boss to text me after getting off work!	0.407233924
2022美国雇主百强榜，Zoom跌至垫底，第一是…	In the 2022 Top 100 U.S. Employers List, Zoom fell to the bottom, the first is…	2021，科技公司最全工资范围！	In 2021, the most complete salary range of technology companies!	0.461193502
再见！超过50%的湾区人正打算永远离开…	Goodbye! Over 50% of Bay Area residents are planning to leave forever…	从湾区到 Google 苏黎世，离开湾区这两年的新感悟！	From the Bay Area to Google Zurich, the new insights of leaving the Bay Area in the past two years!	0.462497652
笑话变神话！22岁大学生靠卖自拍，20天赚了100万美元…	Jokes become myths! 22-year-old college student made $1 million in 20 days by selling selfies…	一天稳赚七八百，安利一个超适合留学生的副业，时间地点自由！	Earn seven or eight hundred RMB a day, recommend a side business that is super suitable for international students, and the time and place are flexible!	0.41891697
突发！1人遇难4人受伤，中国留学生惨遭严重车祸…	Sudden! 1 killed, 4 injured, Chinese students suffered a serious car accident…	突发！纽约高楼发生火灾！一群青少年跳窗逃脱，多人伤亡！	Sudden! New York high-rise building fire! A group of teenagers jumped out of the window to escape, many casualties!	0.430882156
最新！全球最具吸引力公司排名，华为上榜！FB第34，第1名是…	Up to date! The world's most attractive company ranking, Huawei is on the list! FB No. 34, No. 1 is…	全球程序员收入排行榜！最高工资120万美元！	Global Programmer Income Ranking! $1.2 million maximum salary!	0.402175754
内部员工爆料：谷歌、脸书面试大放水！科技公司已在跪着招人了……	Internal staff broke the news: Google, Facebook interviews are full of water! Tech companies are already on their knees to recruit…	Google已经不是最难进的大厂了！	Google is not the most difficult big company to enter!	0.407426566
3个孩子和9岁弟弟尸体同住一年，遗弃他们的父母竟被释放了…	3 children live with 9-year-old brother's body for a year, parents who abandoned them were released…	离谱！美国女网红离奇死亡，警察追捕未婚夫期间发现了9具尸体！	Outrageous! The mysterious death of a female Internet celebrity in the United States, the police found 9 bodies during the pursuit of her fiance!	0.416379511
强势斩获9枚offer！我的申请经验是这样的…	Strongly won 9 offers! My application experience is like this…	来分享科技公司面试官经验，9个注意点！	Come to share the experience of interviewers in technology companies, 9 points to pay attention to!	0.44659543
微软竟斥资687亿美元买下它！太魔幻了…	Microsoft paid $68.7 billion to buy it! so magical…	工作三年后终于离开了微软！	Finally left Microsoft after working for three years!	0.52946502

Figure 7: Headlines in the information-gap salience validation experiment (measured by the presence of an ellipsis). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Manuscript Under Review

For Peer Review

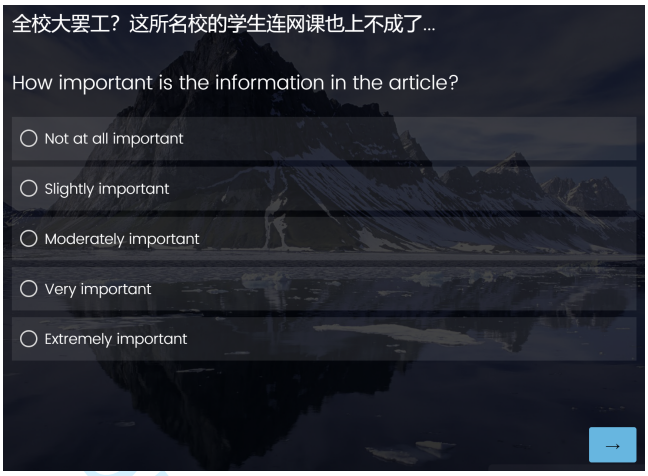


Figure 8: Screenshot of the information-gap importance validation experiment (measured by the presence of an exclamation mark).

Author Manuscript
For Peer Review

Originally with exclamation mark	Originally with exclamation mark	Originally without exclamation mark	Originally without exclamation mark	Semantic similarity
就离谱！哥大罢工了，可是留学生们惨了	Outrageous! Columbia University strikes, but foreign students are miserable...	全校大罢工？这所名校的学生连网课也上不成了…	A school-wide strike? Students in this famous university can't even take online courses...	0.671105981
英国工资也太低了吧！	Wages in the UK are too low!	法国工资比德国还低？	Wages are lower in France than in Germany?	0.626296401
想去大厂？转码太难！不如换个方向？	Want to go to a big factory? Transcoding is too difficult! Why don't we go the other way?	大厂8年，码农还能转什么？	Eight years in a big factory, what else can a code farmer do?	0.587154448
毕业3年留美买房拿身份？！我能重新留学吗！	Graduate 3 years to stay in the United States to buy a house to take identity? ! Can I study abroad again?	留学生毕业，究竟要不要回国？我劝你冷静想想这些事	Should international students return home after graduation? I urge you to think about these things calmly	0.564003944
买房丢了几十万首付，又失而复得的危险经历！	Buy a house lost hundreds of thousands of down payment, and lost the dangerous experience!	差5万没买到dream house，真的好难受...	I am so sad that I failed to buy a dream house for 50,000 dollars...	0.536043465
10分钟，加州一居民被骗走1160万美元巨款！	California resident cheated out of $11.6 million in 10 minutes!	10秒抢走劳力士？加州成"0元购"天堂？	Rolex in 10 seconds? California a $0 paradise?	0.515210629
防不胜防！聚会时突然昏迷、失去意识，醒来发现被扎了…	Impossible to guard against! Party suddenly coma, lose consciousness, wake up to find stabbed...	半夜思念老人，疫情之下不知什么时候能团聚...	Missing the elderly in the middle of the night, I do not know when to reunite under the epidemic...	0.512128651
推荐一个很适合女性的小众高薪副业：600–800元/天！	Recommend a niche, high-paying sideline for women: 600–800 RMB per day!	推荐一个很适合女性的副业	Recommend a side hustle for women	0.667689204
湾区condo抄底经历，1个月内买到自住房！	Bay Area condo fishing experience, buy a home in 1 month!	湾区艰难自建房实录（开始设计）	Hard To Build a House in Bay Area (Start designing)	0.514964283
难道现在是房价下跌的第1个阶段！？	Is this the first stage of falling housing prices!?	都这个时候了，你还要冲进去买房子？	You want to rush in and buy a house at this hour?	0.510557771

Figure 9: Headlines in the information-gap importance validation experiment (measured by the presence of an exclamation mark). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review

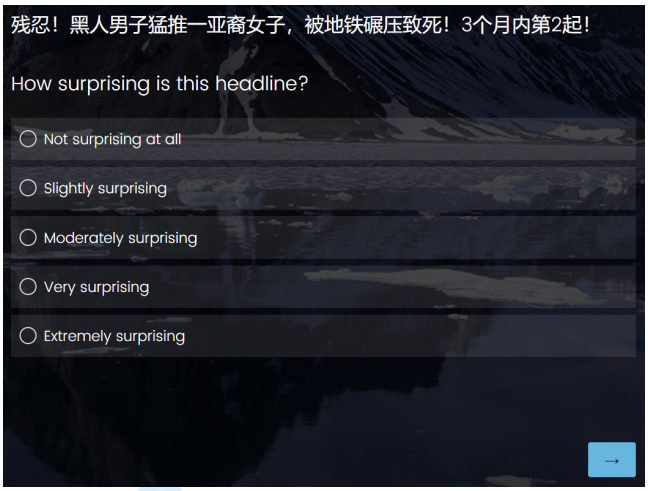


Figure 10: Screenshot of the information-gap surprisingness validation experiment (measured by classification as a negation)

For Peer Review

Author Manuscript

with negation	with negation	without negation	without negation	Semantic Similarity
亚裔女生为救同胞，却惨遭黑人群殴！现场无一人施救！	Asian girls were beaten by black people to save their compatriots! No one was there to help!	残忍！黑人男子猛推一亚裔女子，被地铁碾压致死！3个月内第2起！	Cruel! A black man shoved an Asian woman and was crushed to death by the subway! 2nd in 3 months!	0.723154008
全校大哭工？这所名校的学生连网课也上不成了...	A school-wide strike? Students at this prestigious school can't even take online classes...	就离谱！哥大哭工了，可是留学生们惨了...	Outrageous! Columbia University is on strike, but the international students are miserable...	0.671105981
没绿卡？！强烈建议别去Facebook！	No permanent residency? ! Strongly recommend not going to Facebook!	活久见！花钱直接拿绿卡？！合法敛财来了？	Live longer see more! Spend money directly to get permanent residency? ! Is it legal to make money that way?	0.578808665
差5万没买到dream house，真的好难受...	It's really sad that I didn't buy a dream house for $50,000...	100w美金，在中部能买什么样的房...	100 million US dollars, what kind of house can I buy in the middle...	0.54601264
谷歌面试官求你们了！面试的时候别这么整！	Google interviewers beg you! Don't do this during the interview!	想跳槽？你可能被谷歌安全部盯上了！	Want to jump ship? You may be targeted by Google Security!	0.545233369
骗子项目！留学花费几十万，回国不能通过学历认证！	Liar Project! Studying abroad costs hundreds of thousands, and this cannot pass the academic certification when returning to China!	留学生毕业回国后多久能赚够一百万？	How long does it take for international students to earn one million after graduation?	0.523976445
“我留了两年学，就再也没法回国了”	"I stayed for two years to study abroad, and I will never be able to return to China"	“这一届留学生，活得可太容易了。”	"Life is too easy for this class of international students."	0.523976445
炒币炒股不如纯打工？！湾区人实现财富自由的方式竟是……	Investing in stocks or currency cannot do better than just working for others? ! The way people in the Bay Area achieve financial freedom is actually...	那些早早迈入富人圈的湾区人，都做对了同一件事...	Those Bay Area people who became rich early did the same thing right...	0.527217567
接到大使馆电话，摁了7后，父母积蓄全没了...	After receiving the call from the embassy and pressing 7, my parents' savings were all gone...	留学4年，最怕给客服打电话，哭...	After studying abroad for 4 years, I am most afraid of calling customer service. Crying...	0.490797371
想去大厂？转码太难！不如换个方向？	Want to go to a big company? Transiting to a new career in software engineering is too hard! Why not change direction?	中年转行成功！拿到大厂职位！如何平衡本职工作、面试和家庭？	Successful middle-aged career change! Get a job in a big company! How do you balance work, interviews, and family?	0.486133635

Figure 11: Headlines in the information-gap surprisingness validation experiment (measured by classification as a negation). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

For Peer Review

Author Manuscript

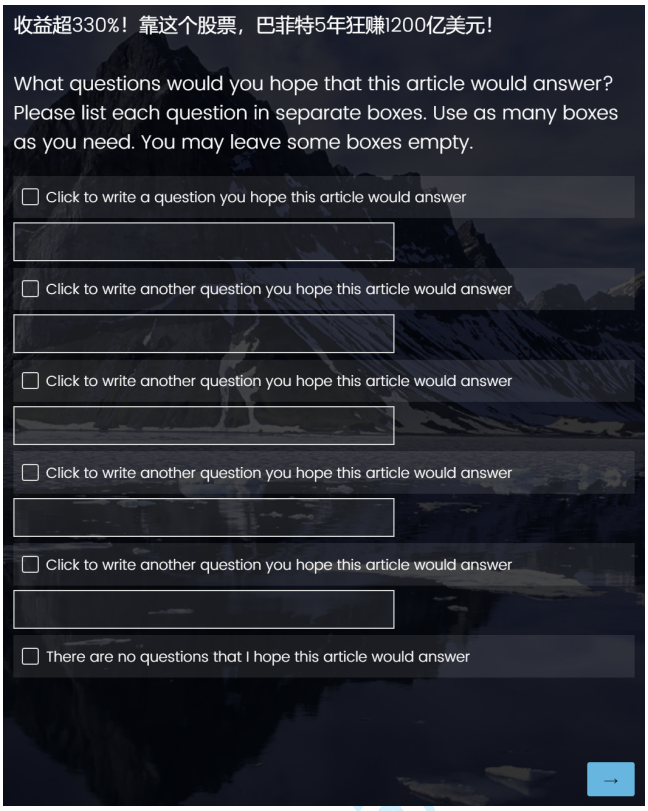


Figure 12: Screenshot of the information-gap surprisingness validation experiment (measured by KL-divergence)

Author Manuscript
For Review

High KL-divergence	High KL-divergence	KL-divergence	Low KL-divergence	Low KL-divergence	KL-divergence	Semantic Similarity
气愤！多名中国留学生遭重物击打头	Angry! Several Chinese students were hit in the head with heavy objects!	3.316964093	就离谱！哥大罢工了，可是留学生们惨了...	This is ridiculous! Columbia's on strike, but the foreign students are in trouble...	0.669410288	0.4732
来美7年间踩过的坑！	7 years in the United States to step on the pit!	1.935551708	转码7年，面试7把，最终上岸！	Seven Years of transcoding, seven rejected interviews, and finally went on the shore!	0.566100069	0.4756
没绿卡？！强烈建议别去Facebook！	Without a green card? ! I strongly recommend not going to Facebook!	1.976028918	好消息！Facebook 的H1b终于恢复办理	Good news! Facebook H1b is finally back to green card!	1.083041886	0.5891
突发！华人男子遭同族当街射杀！全美最大华人区还安全吗？	Sudden! Chinese man shot dead in the street! Is America's largest Chinese enclave safe?	1.99468195	华人大哥被3个黑人抢劫，拔枪反杀，歹徒1死1伤！	The Chinese brother was mugged by 3 black men. He drew his gun and killed 1 dead and 1 wounded!	1.014321091	0.5681
恐怖！自杀式爆炸再现，出租车司机神操作救了一整座城市！	Terrifying! Suicide bombing, taxi driver saved an entire city!	1.940101668	突发！纽约高楼发生火灾！一群青少年跳窗逃脱，多人伤亡！	Sudden! New York high-rises on fire! A bunch of teenagers jumped out of a window. Multiple casualties!	1.015168744	0.4467
华裔妻子惨死家中！2天前刚购买巨额保险...	Chinese wife dies at home! Just bought a huge insurance policy 2 days ago...	2.8323984	痛心！湾区突发枪战，2岁华裔男童中弹身亡...	Ouch! Two-year-old Chinese boy shot dead in a shootout in the Bay Area...	0.850930501	0.4662
湾区和西雅图房子一点都不划算，为什么那么多人买？	Bay Area and Seattle houses are not a good deal. Why do so many people buy them?	1.592110566	谁说西雅图比湾区便宜的...	Who says Seattle's cheaper than the Bay Area...	0.82202861	0.5974
现在社招码农进谷歌工作是不是很难	Not sure whether it is very difficult to obtain technical upgrade via social recruitment into Google as programmers now...	1.835407559	说说谷歌的庸俗化...	Talk about the vulgarization of Google...	0.684593194	0.3896
难道现在是房价下跌的第1个阶段！？	Is this the first stage of falling housing prices!?	2.882783493	都这个时候了，你还要冲进去买房子？	At this point, you're gonna go in there and buy a house?	0.734891767	0.4326
大厂的高层领导，每天到底干啥？	What do the senior leaders of big companies do every day?	2.149929156	向上管理，搞定老板...	Managing up, managing the boss...	1.192484231	0.4723

Figure 13: Headlines in the information-gap surprisingness validation experiment (measured by KL-divergence). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review

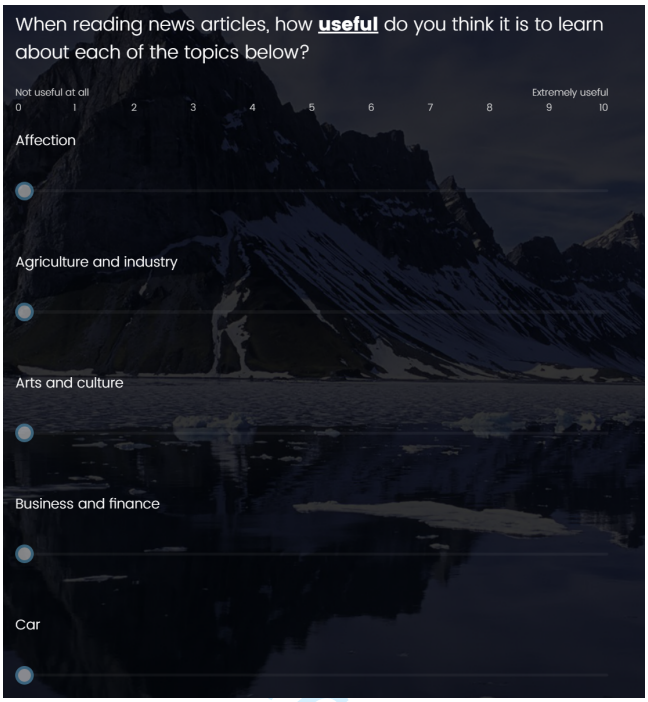


Figure 14: Screenshot of rating the usefulness of 36 different news topics on a Likert scale from 0 to 10

Author Manuscript
For Review

Author Manuscript
For Peer Review

Appendix (For Online Publication Only)

Supplemental Tables and Figures

Author Manuscript

For Peer Review

Author Manuscript

Table 4: OLS and Negative Binomial Regression Results

Model:	OLS			Negative Binomial		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	*ln (Reading+1)*			*Reading*		
Question mark	0.030***	0.029***	0.034***	0.031**	0.031**	0.037***
	(0.010)	(0.010)	(0.010)	(0.014)	(0.014)	(0.014)
Ellipsis	0.088***	0.088***	0.098***	0.090***	0.090***	0.099***
	(0.020)	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)
Exclamation	0.067***	0.067***	0.070***	0.076***	0.076***	0.079***
	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
Position 1	0.691***	0.691***	0.688***	0.918***	0.918***	0.913***
	(0.016)	(0.016)	(0.016)	(0.028)	(0.028)	(0.028)
Position 2	0.075***	0.075***	0.075***	0.175***	0.176***	0.174***
	(0.014)	(0.014)	(0.014)	(0.025)	(0.025)	(0.025)
Position 3	-0.010	-0.010	-0.009	0.028	0.028	0.027
	(0.014)	(0.014)	(0.014)	(0.025)	(0.025)	(0.025)
Position 4	-0.034**	-0.034**	-0.033**	-0.038	-0.038	-0.038
	(0.014)	(0.014)	(0.014)	(0.026)	(0.026)	(0.026)
Position 5	-0.062***	-0.062***	-0.062***	-0.080***	-0.080***	-0.081***
	(0.014)	(0.014)	(0.014)	(0.027)	(0.027)	(0.027)
Position 6	-0.049***	-0.049***	-0.051***	-0.060**	-0.060**	-0.064**
	(0.014)	(0.014)	(0.014)	(0.027)	(0.027)	(0.027)
Negation	0.020***	0.020***	0.020***	0.027***	0.027***	0.024***
	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
KL-divergence	0.066***	0.066***	0.111***	0.078***	0.078***	0.115***
	(0.020)	(0.020)	(0.027)	(0.024)	(0.024)	(0.034)
Quadratic KL-divergence	-0.014***	-0.013***	-0.019***	-0.016**	-0.016**	-0.020**
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.008)
Sentiment	-0.074***	-0.074***	-0.064***	-0.086***	-0.086***	-0.072***
	(0.010)	(0.010)	(0.010)	(0.013)	(0.013)	(0.013)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

For Peer Review

Table 5: OLS Regression Results (Excluding Duplicate Headlines)

Dependent Variable:	*ln* (Reading+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.029***	0.029***	0.033***	0.029***	0.029***	0.033***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Ellipsis	0.094***	0.094***	0.103***	0.094***	0.095***	0.104***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Exclamation mark	0.073***	0.073***	0.075***	0.073***	0.073***	0.075***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Position 1	0.677***	0.678***	0.675***	0.678***	0.678***	0.675***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Position 2	0.066***	0.066***	0.066***	0.066***	0.066***	0.066***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 3	-0.014	-0.014	-0.014	-0.014	-0.014	-0.014
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 4	-0.036**	-0.036**	-0.035**	-0.036***	-0.036***	-0.035**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 5	-0.057***	-0.057***	-0.057***	-0.057***	-0.057***	-0.057***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 6	-0.049***	-0.049***	-0.051***	-0.049***	-0.049***	-0.051***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Negation	0.019***	0.019***	0.018**	0.019***	0.019***	0.018**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
KL-divergence	0.016***	0.016***	0.039***	0.064***	0.064***	0.109***
	(0.005)	(0.005)	(0.009)	(0.019)	(0.019)	(0.026)
Sentiment	-0.077***	-0.077***	-0.068***	-0.077***	-0.077***	-0.067***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Quadratic KL-divergence				-0.013***	-0.013**	-0.018***
				(0.005)	(0.005)	(0.006)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	108,513	108,513	108,513	108,513	108,513	108,513

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript For Peer Review

Author Manuscript

Table 6: Negative Binomial Regression Results (Excluding Duplicate Headlines)

Dependent Variable:	*Reading*					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.029**	0.029**	0.034**	0.029**	0.029**	0.034**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Ellipsis	0.096***	0.096***	0.104***	0.097***	0.097***	0.105***
	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)	(0.021)
Exclamation mark	0.083***	0.083***	0.085***	0.083***	0.083***	0.085***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Position 1	0.898***	0.898***	0.893***	0.898***	0.898***	0.893***
	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)	(0.027)
Position 2	0.166***	0.166***	0.164***	0.166***	0.166***	0.164***
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Position 3	0.024	0.024	0.023	0.024	0.024	0.024
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Position 4	-0.039	-0.039	-0.039	-0.039	-0.039	-0.039
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Position 5	-0.072***	-0.072***	-0.073***	-0.071***	-0.072***	-0.072***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Position 6	-0.058**	-0.058**	-0.062**	-0.058**	-0.058**	-0.062**
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Negation	0.026***	0.025***	0.022**	0.025***	0.025***	0.022**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
KL-divergence	0.019***	0.019***	0.040***	0.073***	0.073***	0.111***
	(0.006)	(0.006)	(0.012)	(0.024)	(0.024)	(0.033)
Sentiment	-0.088***	-0.088***	-0.075***	-0.088***	-0.088***	-0.073***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Quadratic KL-divergence				-0.014**	-0.014**	-0.018**
				(0.006)	(0.006)	(0.008)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	108,363	108,363	108,363	108,363	108,363	108,363
Over-dispersion	3.6978	3.6988	3.7282	3.6982	3.6991	3.7286

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. Note: 42 fixed-effects (150 observations) removed because of only 0 outcomes.

For Peer Review

Table 7: OLS Regression Results (Imputing 171,000 as the Maximum Readings)

Dependent Variable:	*ln* (Reading+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.029***	0.029***	0.034***	0.030***	0.029***	0.034***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Ellipsis	0.087***	0.087***	0.097***	0.088***	0.088***	0.098***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Exclamation mark	0.067***	0.067***	0.070***	0.067***	0.067***	0.070***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Position 1	0.691***	0.691***	0.689***	0.691***	0.691***	0.689***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Position 2	0.076***	0.076***	0.076***	0.075***	0.076***	0.076***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 3	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 4	-0.033**	-0.033**	-0.033**	-0.034**	-0.033**	-0.033**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 5	-0.061***	-0.061***	-0.061***	-0.062***	-0.061***	-0.062***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 6	-0.049***	-0.049***	-0.051***	-0.049***	-0.049***	-0.051***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Negation	0.021***	0.021***	0.020***	0.021***	0.021***	0.021***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
KL-divergence	0.015***	0.015***	0.036***	0.066***	0.066***	0.112***
	(0.005)	(0.005)	(0.009)	(0.020)	(0.020)	(0.027)
Sentiment	-0.073***	-0.073***	-0.065***	-0.073***	-0.073***	-0.064***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Quadratic KL-divergence				-0.014***	-0.013***	-0.019***
				(0.005)	(0.005)	(0.006)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript For Peer Review

Author Manuscript

Table 8: Negative Binomial Regression Results (Imputing 171,000 as the Maximum Readings)

Dependent Variable:	*Reading*					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.031**	0.031**	0.036***	0.031**	0.031**	0.037***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Ellipsis	0.089***	0.089***	0.097***	0.090***	0.090***	0.098***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Exclamation mark	0.076***	0.076***	0.078***	0.076***	0.076***	0.079***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Position 1	0.920***	0.920***	0.915***	0.920***	0.920***	0.915***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Position 2	0.177***	0.177***	0.175***	0.177***	0.177***	0.176***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Position 3	0.032	0.032	0.031	0.032	0.032	0.031
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Position 4	-0.037	-0.037	-0.037	-0.037	-0.037	-0.037
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Position 5	-0.079***	-0.079***	-0.080***	-0.079***	-0.079***	-0.080***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Position 6	-0.060**	-0.060**	-0.064**	-0.060**	-0.059**	-0.064**
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Negation	0.029***	0.029***	0.026***	0.029***	0.029***	0.026***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
KL-divergence	0.018***	0.018***	0.038***	0.078***	0.077***	0.116***
	(0.006)	(0.006)	(0.012)	(0.024)	(0.024)	(0.034)
Sentiment	-0.085***	-0.085***	-0.071***	-0.085***	-0.085***	-0.070***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Quadratic				-0.016**	-0.016**	-0.020**
KL-divergence				(0.006)	(0.006)	(0.008)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901
Over-dispersion	3.5635	3.5645	3.5917	3.5639	3.5649	3.5922

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

For Peer Review

Table 9: OLS Regression Results with Alternative Sentiment Analysis

Dependent Variable:	*ln* (Reading+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.032***	0.032***	0.035***	0.032***	0.032***	0.036***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Ellipsis	0.090***	0.090***	0.098***	0.091***	0.091***	0.100***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Exclamation	0.072***	0.071***	0.073***	0.071***	0.071***	0.073***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Position 1	0.690***	0.690***	0.688***	0.690***	0.690***	0.688***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Position 2	0.075***	0.075***	0.075***	0.075***	0.075***	0.075***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 3	-0.010	-0.010	-0.009	-0.010	-0.010	-0.009
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 4	-0.034**	-0.034**	-0.033**	-0.034**	-0.034**	-0.033**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 5	-0.062***	-0.062***	-0.062***	-0.062***	-0.062***	-0.062***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 6	-0.050***	-0.050***	-0.051***	-0.049***	-0.049***	-0.051***
	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)
Negation	0.027***	0.027***	0.025***	0.027***	0.027***	0.025***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
KL-divergence	0.013**	0.013**	0.035***	0.066***	0.066***	0.113***
	(0.005)	(0.005)	(0.009)	(0.020)	(0.020)	(0.027)
Valence	-0.008***	-0.008***	-0.007***	-0.008***	-0.008***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Arousal	0.002	0.002	0.003	0.002	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Quadratic KL-divergence				-0.014***	-0.014***	-0.020***
				(0.005)	(0.005)	(0.006)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript For Peer Review

Table 10: Regression Results of the Five Validation Experiments

Dependent Variables:	Salience		Importance	Surprisingness	
	(1)	(2)	(3)	(4)	(5)
Question mark	0.057*				
	(0.029)				
Ellipsis		0.058**			
		(0.028)			
Exclamation mark			0.233***		
			(0.033)		
Negation				0.071**	
				(0.035)	
KL-divergence					0.110***
					(0.023)
Individual FE	Yes	Yes	Yes	Yes	Yes
Paired headline FE	Yes	Yes	Yes	Yes	Yes
Observations	2,000	2,000	2,000	2,000	2,000

Standard errors clustered at the individual level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript

For Peer Review

Table 11: Topic Tags' Regression Results from Primary OLS Specification

Affection	0.073	Agriculture and industry	0.018
	(0.047)		(0.121)
Art and culture	0.172*	Business and finance	0.039
	(0.088)		(0.050)
Car	0.024	Cartoon	-0.292
	(0.102)		(0.227)
Celebrity	0.404***	Documentary	0.369*
	(0.140)		(0.220)
Education and training	0.028	Fashion and skincare	0.109*
	(0.048)		(0.062)
Fate and constellation	-0.198	Fiction	-0.121
	(0.301)		(0.133)
Food and catering	-0.028	Game	0.053
	(0.069)		(0.101)
Health	0.106*	History	0.131*
	(0.060)		(0.078)
Joke	-0.056	Literature	0.059
	(0.070)		(0.052)
Military	0.241***	Marketing	0.017
	(0.091)		(0.483)
Motivation	-0.019	Movie	0.110
	(0.106)		(0.106)
Music	0.013	Parenting	-0.057
	(0.082)		(0.078)
Pet	-0.109	Photography	0.296
	(0.165)		(0.294)
Politics	0.226***	Real estate	0.098
	(0.061)		(0.076)
Religion	-0.140	Science	0.220
	(0.167)		(0.135)
Society	0.187***	Sports	0.187
	(0.042)		(0.117)
Technology	0.004	Travel	0.063
	(0.078)		(0.071)
TV series	-0.088	Workplace	0.149
	(0.131)		(0.098)

Standard errors clustered at the individual level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript For Peer Review

Table 12: Summary of Measures for Facebook Dataset

Construct	Measure	Source
Salience	1: Question mark	Punctuation
	2: Ellipsis	Punctuation
	3: Wh-question sentence	NPS Chat Corpus
Importance	1: Exclamation mark	Punctuation
	2: Emphasis sentence	NPS Chat Corpus
Surprisingness	1: Reject sentence	NPS Chat Corpus
	2: KL-divergence	LDA
Valence	Sentiment of headline	qdap R package

Author Manuscript
For Peer Review

Table 13: Descriptive Statistics for Facebook Dataset

Statistics	Mean	St. Dev.	Min	Max
Binary variables			0	1
Question mark	0.035	0.183		
Ellipsis	0.294	0.455		
Wh-question	0.054	0.226		
Exclamation mark	0.004	0.061		
Emphasis sentence	0.040	0.195		
Reject sentence	0.014	0.116		
Continuous variables				
KL-divergence	1.956	0.303	0.071	3.900
Sentiment of title	-0.006	0.137	-0.951	0.962
Facebook popularity	129.363	661.182	0	49211
				N = 81,637

Author Manuscript For Peer Review

Table 14: OLS Regression Results for Facebook Data

Dependent Variable:	*ln* (Popularity+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	-0.039	-0.039	-0.181***	-0.040	-0.039	-0.181***
	(0.036)	(0.036)	(0.064)	(0.036)	(0.036)	(0.064)
Ellipsis	0.593***	0.569***	0.215***	0.590***	0.566***	0.214***
	(0.014)	(0.014)	(0.032)	(0.014)	(0.014)	(0.032)
Wh-question	0.529***	0.515***	0.148***	0.528***	0.515***	0.148***
	(0.029)	(0.029)	(0.028)	(0.029)	(0.029)	(0.028)
Exclamation mark	0.532***	0.511***	0.181**	0.535***	0.514***	0.181**
	(0.107)	(0.107)	(0.085)	(0.107)	(0.107)	(0.085)
Emphasis	0.120***	0.117***	0.077***	0.118***	0.114***	0.076***
	(0.033)	(0.033)	(0.028)	(0.033)	(0.033)	(0.028)
Reject	0.189***	0.194***	0.129***	0.191***	0.196***	0.129***
	(0.056)	(0.056)	(0.045)	(0.056)	(0.056)	(0.045)
KL-divergence	-0.160***	-0.170***	-0.034	0.705***	0.706***	0.181
	(0.027)	(0.027)	(0.041)	(0.144)	(0.144)	(0.205)
Sentiment	-0.406***	-0.398***	-0.207***	-0.409***	-0.401***	-0.208***
	(0.048)	(0.047)	(0.035)	(0.048)	(0.047)	(0.035)
Quadratic KL-divergence				-0.213***	-0.215***	-0.051
				(0.035)	(0.035)	(0.045)
LDA topic FE	No	No	Yes	No	No	Yes
Source FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
News topic FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	81,637	81,637	81,350	81,637	81,637	81,350

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. Note: In column (3) and column (6), where we add source fixed effects and LDA topic fixed effects, some observations are dropped because of the lack of sources of some articles in the Facebook dataset.

Author Manuscript

For Peer Review

Table 15: Negative Binomial Regression Results for Facebook Data

Dependent Variable:	*Popularity*					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	-0.354***	-0.353***	-0.350***	-0.354***	-0.353***	-0.351***
	(0.041)	(0.041)	(0.081)	(0.041)	(0.041)	(0.081)
Ellipsis	0.383***	0.363***	0.214***	0.383***	0.363***	0.213***
	(0.016)	(0.016)	(0.039)	(0.016)	(0.016)	(0.039)
Wh-question	0.308***	0.312***	0.166***	0.306***	0.311***	0.167***
	(0.033)	(0.033)	(0.045)	(0.033)	(0.033)	(0.045)
Exclamation_mark	0.164	0.146	0.223*	0.165	0.146	0.223*
	(0.120)	(0.120)	(0.121)	(0.120)	(0.120)	(0.121)
Emphasis	0.222***	0.207***	0.161***	0.222***	0.207***	0.160***
	(0.037)	(0.037)	(0.056)	(0.037)	(0.037)	(0.056)
Reject	0.281***	0.314***	0.350*	0.282***	0.314***	0.348*
	(0.063)	(0.063)	(0.189)	(0.063)	(0.063)	(0.188)
KL-divergence	-0.036	-0.054*	0.063	-0.152	-0.125	0.584**
	(0.029)	(0.029)	(0.053)	(0.173)	(0.174)	(0.289)
Sentiment	-0.388***	-0.349***	-0.330***	-0.387***	-0.349***	-0.332***
	(0.054)	(0.054)	(0.074)	(0.054)	(0.054)	(0.075)
Quadratic KL-divergence				0.027	0.017	-0.125*
				(0.040)	(0.040)	(0.066)
LDA topic FE	No	No	Yes	No	No	Yes
Source FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
News topic FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	81,637	81,637	78,023	81,637	81,637	78,023
Over-dispersion	0.23467	0.23543	0.48534	0.23467	0.23543	0.48539

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. Note: In column (3) and column (6), where we add source fixed effects and LDA topic fixed effects, some observations are dropped because of the lack of sources of some articles in the Facebook dataset.

Author Manuscript For Peer Review

Author Manuscript



Figure 1: WeChat Webpage

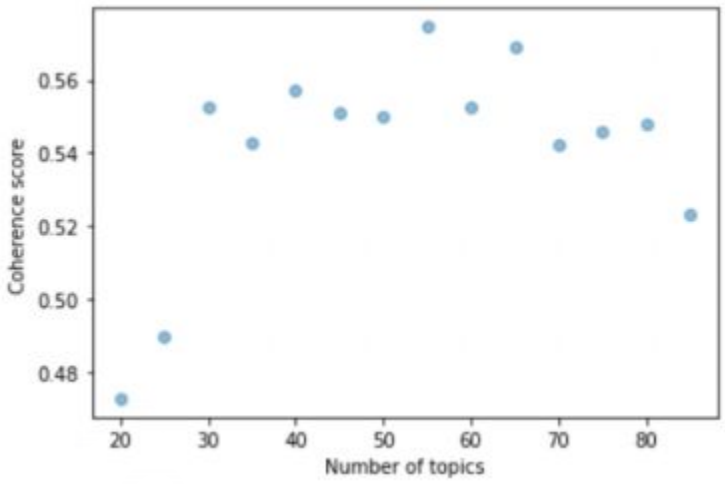


Figure 2: Scatterplot of coherence and number of topics (WeChat)

For Peer Review
Author Manuscript

Mandarin	非	别	不	没	无	勿	阻止	制止
English	not	don't	no	no	no	no	stop	stop
Mandarin	差	未	难	跌停	止	开除	莫	愁
English	poor	not	difficult	fall	stop	expel	no	sorrow
Mandarin	否	难	缺	乏	苦	艰	危	险
English	no	difficult	short of	short of	no	difficult	dangerous	risky

Figure 3: Negation words

Author Manuscript
For Peer Review

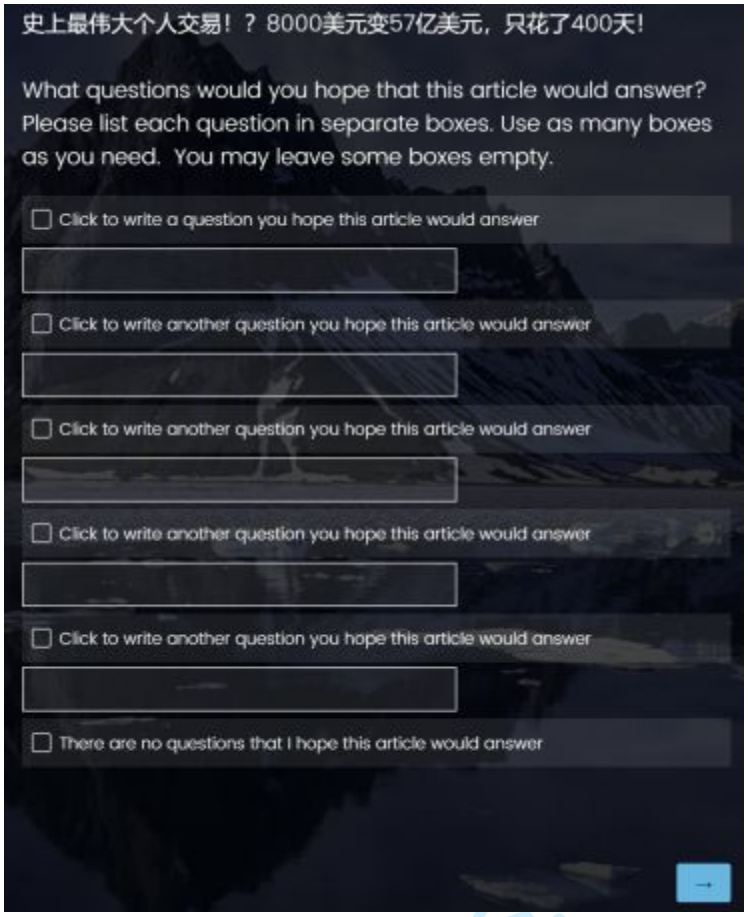


Figure 4: Screenshot of the information-gap salience validation experiment (measured by the presence of a question mark)

Author Manuscript

Originally with question mark	Originally with question mark	Originally without question mark	Originally without question mark	Semantic similarity
史上最伟大个人交易？！ 800美元变57亿美元，只花了480天！	Greatest personal deal ever? ! $8,800 became $5.7 billion in just 480 days!	收益超349%！靠这个股票，巴菲特5年狂赚1200亿美元！	The income is over 349%! With this stock, Buffet earned $120 billion in 5 years!	0.452171296
取消SAT成绩要求？！加州大学录取现在看什么？	Dropping SAT score requirements? ! What does the University of California admission look like now?	加州，不许给贫生打低分！为了推存公平！	California: Don't give poor students a low score! For education fairness!	0.522243923
全美首个成执行死刑的中国留学生，犯了什么罪？	What crime did the first Chinese student in the United States execute or execute?	让歌案震网！美国留学生爆杀案，似加此动奇！	The Jiang Ge case is sensational! The murder case of American students is also so bizarre!	0.515795899
在美国这个职业，居然没有work life balance吗？	Is there no work life balance in this profession in the United States?	在美国工作没"大饭"，就该如何节心的活着	How can I live happily if I don't have a "big bag" to work in the United States	0.483206153
美国将要发280万张绿卡！排期还能交钱插队？！	The United States will issue 2 million green cards! Can I still pay to cut the queue? !	视频了！特朗16万移民直接拿绿卡！	I'm sorry! Amazon for 160,000 immigrants to get green cards directly!	0.516702403
交了5个月房租，发现房东竟是我老婆！！	After paying 5 months rent, I found out that the landlord is my wife! ?	消区canada炒房经历，1个月内买到自住房！	Buy Area condo bargain hunting experience, buy a house within 1 month!	0.481148703
什么职业既优雅，又有"钱"景？	What sideline is both elegant and "money"?	推荐一个很适合女性的副业	Recommend a side business that is very suitable for women	0.468831824
实习生是写大厂代码是什么体验！	What is the experience of the intern rewriting the code of the big factory?	分享analyscall跳槽到大厂的经历和感想	Share the experience and impressions of analytics jumping to Dachang	0.458224714
为什么跳槽，肯定要比养上去的拿得多！	Why jump jobs, must get more than the promotion!	今年2次跳槽，工资从$4.5万涨到$13万！	Twice this year, the salary has increased from $45,000 to $130,000!	0.469226803
被面试官用十招处死面试？没永攻略来啦！	Was interviewed by the interviewer with ten tricks? Counterattack strategy is here!	加拿大找工求职逆袭开这些坑！	Please avoid these pits when looking for a job in Canada!	0.458164698

Figure 5: Headlines in the information-gap salience validation experiment (measured by the presence of a question mark). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review

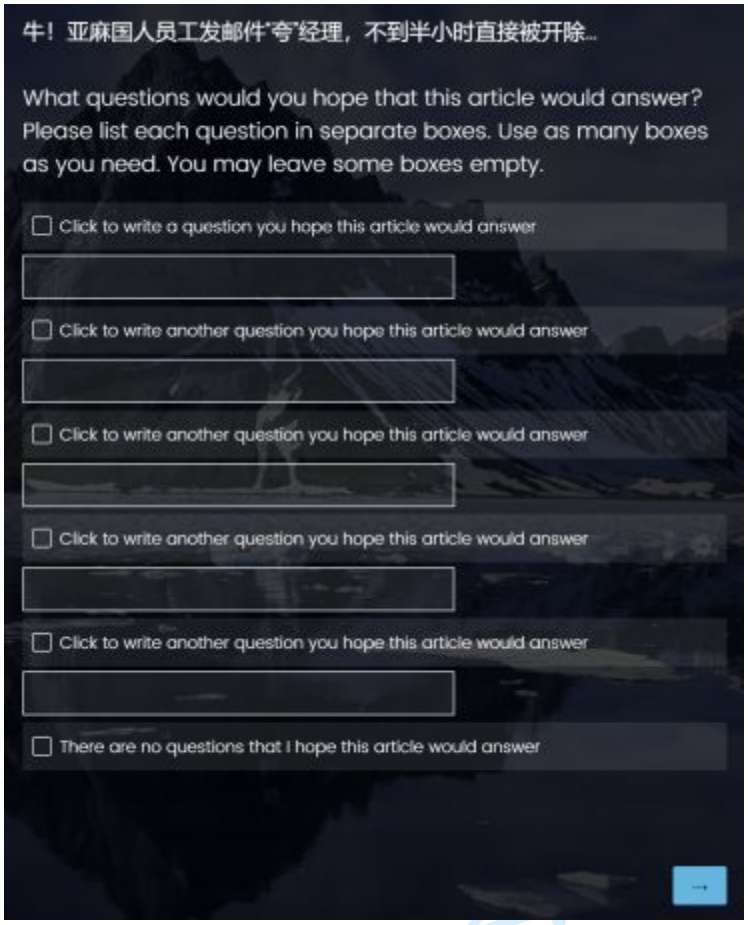


Figure 6: Screenshot of the information-gap salience validation experiment (measured by the presence of an ellipsis).

Author Manuscript

Originally with ellipsis	Originally with ellipsis	Originally without ellipsis	Originally without ellipsis	Semantic similarity
牛！亚裔员工发邮件"全"经理，不到半小时直接被开除…	Staggering! How workers sent emails to "praise" the manager and were fired directly in less than half an hour…	忘了！老板下班后发信息给我，马上！	Dead! It's illegal for my boss to text me after getting off work!	0.407233924
2022美国最卡月薪榜，Zoom跌至垫底，第一是…	In the 2022 Top 100 U.S. Employers List, Zoom fell to the bottom, the first is…	2021，科技公司最全工资揭秘！	In 2021, the most complete salary range of technology companies!	0.463159582
再见！超过58%的湾区人正打算永远离开…	Goodbye! Over 58% of Bay Area residents are planning to leave forever…	从湾区到 Google 苏黎世，离开湾区这两年的阶感悟！	From the Bay Area to Google Zurich, the new insights of leaving the Bay Area in the past two years!	0.462497652
美国警评估！22岁大学生靠卖自拍，20天赚了100万美元…	Jokes become myths! 22-year-old college student made $1 million in 20 days by selling selfies…	一天稳赚七八百，女同一个超适合留学生的副业，时间地点自由！	Earn seven or eight hundred RMB a day, recommend a side business that is super suitable for international students, and the time and place are flexible!	0.41891697
突发！1人遇难4人受伤，中国留学生换乘严重车祸…	Sudden! 1 killed, 4 injured, Chinese students suffered a serious car accident…	突发！纽约高楼发生火灾！一群青少年跳窗逃脱，多人伤亡！	Sudden! New York high-rise building fire! A group of teenagers jumped out of the window to escape, many casualties!	0.458682156
最新！全球最具吸引力公司排名，华为上榜！腾讯34，阿里见…	Up to date! The world's most attractive company ranking, Huawei is on the list! TB No. 34, No. 1 is…	全球程序员收入排行榜！最高工资120万美元！	Global Programmer Income Ranking! $1.2 million maximum salary!	0.402175754
内部员工曝料：谷歌，脸书面试大造水！科技公司已沦跑单员工了……	Internal staff broke the news: Google, Facebook interviews are full of water! Tech companies are already on draw leaves is sweat…	Google已经不是最难进的大厂了！	Google is not the most difficult big company to enter!	0.407426566
3个孩子和9岁弟弟尸体同住一年，遭弃绝，他们的父母竟被释放了…	3 children live with 9-year-old brother's body for a year, parents who abandoned them were released…	揭露！美国女网红离奇死亡，警察近搞本脸大聊同发现了8具尸体！	Outrageous! The mysterious death of a female Internet celebrity in the United States, the police found 9 bodies during the pursuit of her fiancé!	0.416379511
谷歌招领9名offer！我的申请经验是这样的…	Strongly won 9 offers! My application experience is like this…	斯坦学科技公司面试官经验，9个注意点！	Come to share the experience of interviewers in technology companies, 9 points to pay attention to!	0.44659543
微软也许愿687亿美元买下它！又赚狂了…	Microsoft paid $68.7 billion to buy it! so magical…	工作三年后终于离开了微软！	Finally left Microsoft after working for three years!	0.52946582

Figure 7: Headlines in the information-gap salience validation experiment (measured by the presence of an ellipsis). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review



Figure 8: Screenshot of the information-gap importance validation experiment (measured by the presence of an exclamation mark).

Author Manuscript
For Peer Review

Originally with exclamation mark	Originally with exclamation mark	Originally without exclamation mark	Originally without exclamation mark	Semantic similarity
就调研！哥大第工厂，可是留学生们靠了	Outrageous! Columbia University strikes, but foreign students are miserable...	全校大第工？这所名校的学生连网课也上不成了...	A school-wide strike? Students in this famous university can't even take online courses...	0.671389981
英国工资该水低了吧！	Wages in the UK are too low!	法国工资比德国还低？	Wages are lower in France than in Germany?	0.626296401
做去去厂！到码太难！不如换个方向？	Want to go to a big factory? Transcoding is too difficult! Why don't we go the other way?	大厂去年，码农还能转什么？	Eight years in a big factory, what else can a code farmer do?	0.587154448
毕业3年前美买房拿绿卡？！就能滚蛋留学吗！	Graduate 3 years to stay in the United States to buy a house to take aboutly? ! Can I study abroad again?	留学生毕业，究竟要不要回国！美好你内容想想这件事	Should international students return home after graduation? I urge you to think about these things calmly	0.561803944
买房差了几十万首付，又大房贷直接的危险经历！	Buy a house lost hundreds of thousands of down payment, and lost the dangerous experience!	差5万没买到dream house，真的好难受...	I am so sad that I failed to buy a dream house for 50,000 dollars...	0.558045405
10分钟，加州一居民被骗走1160万美元巨款！	California resident cheated out of $11.6 million in 10 minutes!	10秒轻走势方主？如何成"0元购"天堂？	Stolen in 10 seconds? California's $0 paradise?	0.552510829
防不胜防！猝会时突然昏迷，天太意识，配来发视频丢了...	Impossible to guard against! Party suddenly coma, lose consciousness, wake up to find stabbed...	半夜思念老人，疫情之下不知什么时候能团聚...	Missing the elderly in the middle of the night, I do not know when to reunite under the epidemic...	0.512328651
推荐一个很适合女性的小众高薪副业，600-800元/天！	Recommend a niche, high-paying sideline for women: 600-800 RMB per day!	推荐一个很适合女性的副业	Recommend a side hustle for women	0.667689204
西区condo外观性别，1个月内买到自住房！	Bay Area condo fishing experience, buy a home in 1 month!	西区旧难自建房实录（开始设计）	Hard To Build a House in Bay Area (Start designing)	0.510864203
难道现在是房价下跌的第1个阶段！？	Is this the first stage of falling housing prices!?	都这个时候了，你还要评进去买房子？	You want to rush in and buy a house at this hour?	0.510055771

Figure 9: Headlines in the information-gap importance validation experiment (measured by the presence of an exclamation mark). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

For Peer Review

Author Manuscript

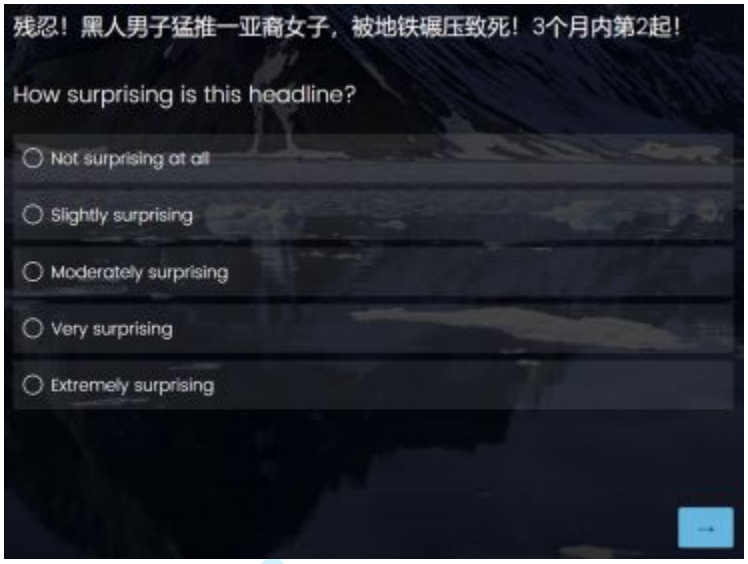


Figure 10: Screenshot of the information-gap surprisingness validation experiment (measured by classification as a negation)

Author Manuscript
Peer Review



with negation	with negation	without negation	without negation	Semantic Similarity
亚裔女生为救同胞，却惨遭黑人群殴！现场无一人施救！	Asian girls were beaten by black people to save their compatriots! No one was there to help!	残忍！黑人男子猛推一亚裔女子，被地铁碾压致死！3个月内第2起！	Cruel! A black man shoved an Asian woman and was crushed to death by the subway! 2nd in 3 months!	0.723154008
全校大罢工？这所名校的学生连网课也上不成了...	A school-wide strike? Students at this prestigious school can't even take online classes...	就离谱！哥大罢工了，可是留学生们挣了...	Outrageous! Columbia University is on strike, but the international students are miserable...	0.671105981
没绿卡？！强烈建议别去Facebook！	No permanent residency? ! Strongly recommend not going to Facebook!	活久见！花钱直接拿绿卡？！合法洗钱来了？	Live longer see more! Spend money directly to get permanent residency? ! Is it legal to make money that way?	0.578898665
差5万没买到dream house，真的好难受...	It's really sad that I didn't buy a dream house for $50,000...	100w美金，在中部能买什么样的房...	100 million US dollars, what kind of house can I buy in the middle...	0.54601264
谷歌面试官求你们了！面试的时候别这么整！	Google interviewers beg you! Don't do this during the interview!	想跳槽？你可能被谷歌安全部盯上了！	Want to jump ship? You may be targeted by Google Security!	0.545233369
骗子项目！留学花费几十万，回国不能通过学历认证！	Liar Project! Studying abroad costs hundreds of thousands, and this cannot pass the academic certification when returning to China!	留学生毕业回国后多久能赚够一百万！	How long does it take for international students to earn one million after graduation?	0.523976445
"我留了两年学，就再也没法回国了"	"I stayed for two years to study abroad, and I will never be able to return to China"	"这一届留学生，活得可太容易了。"	"Life is too easy for this class of international students."	0.523976445
炒股炒股不如纯打工？！湾区人实现财富自由的方式竟是……	Investing in stocks or currency cannot do better than just working for others? ! The way people in the Bay Area achieve financial freedom is actually...	那些早早迈入富人圈的湾区人，都做对了同一件事...	Those Bay Area people who became rich early did the same thing right...	0.527217567
接到大使馆电话，挂了7后，父母积蓄全没了...	After receiving the call from the embassy and pressing 7, my parents' savings were all gone...	留学4年，最怕给客服打电话，哭...	After studying abroad for 4 years, I am most afraid of calling customer service. Crying...	0.490797371
想去大厂？转码太难！不如换个方向？	Want to go to a big company? Transiting to a new career in software engineering is too hard! Why not change direction?	中年转行成功！拿到大厂职位！如何平衡本职工作，面试和家庭？	Successful middle-aged career change! Got a job in a big company! How do you balance work, interviews, and family?	0.486133635

Figure 11: Headlines in the information-gap surprisingness validation experiment (measured by classification as a negation). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

For Peer Review

Author Manuscript

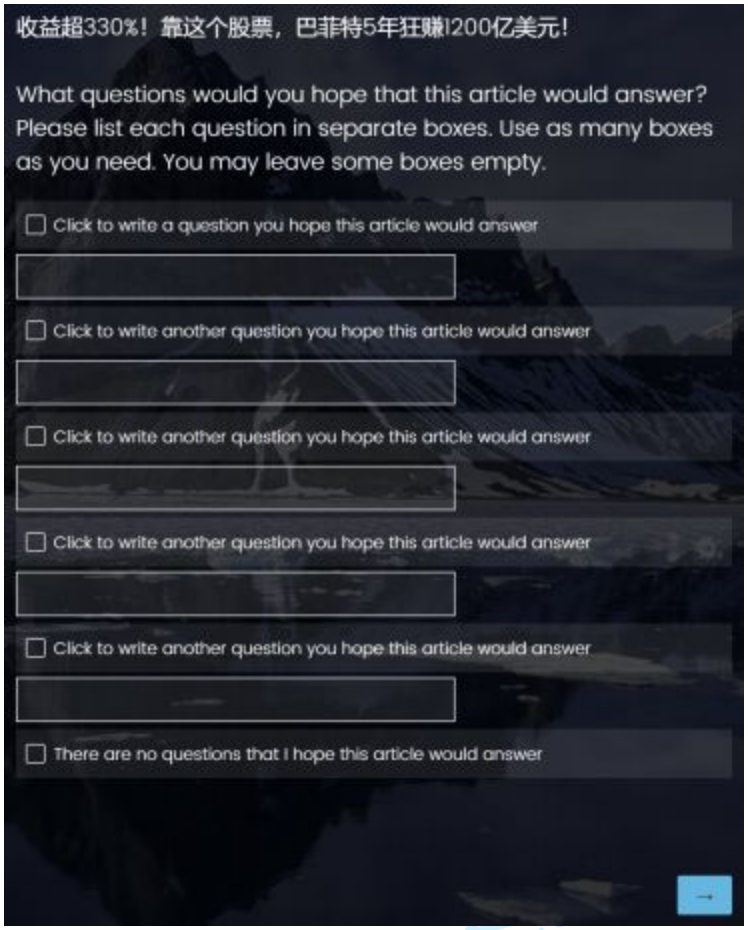


Figure 12: Screenshot of the information-gap surprisingness validation experiment (measured by KL-divergence)

Author Manuscript



High KL-divergence	High KL-divergence	KL-divergence	Low KL-divergence	Low KL-divergence	KL-divergence	Semantic Similarity
气愤！多名中国留学生遭重物击打头	Angry! Several Chinese students were hit in the head with heavy objects!	3.318968093	就离谱！育人第工了，可是接学马后悔了...	This is relentless! Columbia's on strike, but the foreign students are in trouble...	0.689418288	0.4732
发现7年间隔过的机！	7 years in the United States to step on the pit!	1.835551708	科科7年，面试7轮，最终上岸！	Seven Years of homesoding, seven rejected interviews, and finally went on the dance!	0.566308069	0.4758
没错卡？！ 强烈建议别去Facebook！	Without a green card? ! I strongly recommend not going to Facebook!	3.978028918	好消息！Facebook 的HR终于恢复办公Back to green card!	Good news! Facebook HR is finally	1.083841886	0.3891
突发！华人男子遭到枪当街射杀！华美最大华人区还安全吗？	Sudden! Chinese man shot dead in the street! Is America's largest Chinese enclave safe?	1.99468155	华人大哥被3个黑人抢劫，抢劫反杀，3死1死2伤！	The Chinese brother was mugged by 3 black men. He drew his gun and killed 1 dead and 1 wounded!	1.014323091	0.5681
惊悚！台北式爆炸再现，还是车司机神操作救了一整座城市！	Terrifying! Suicide bombing, taxi driver saved an entire city!	1.944181668	突发！纽约高楼冒生火灾！一群青少年跳窗逃脱，多人伤亡！	Sudden! New York high-rises on fire! A bunch of teenagers jumped out of a window. Multiple casualties!	1.015168744	0.4467
毕业季子弹死家中！2名前博纳实习生被保险	Chinese wife dies at home! Just bought a huge insurance policy 2 days ago...	2.8323884	痛心！悲区突发枪战，2名毕业男童中弹身亡...	Ouch! Two-year-old Chinese boy shot dead in a shootout in the Bay Area...	0.850838501	0.4662
湾区和西雅图房子一点都不便宜，为什么那么多人买？	Bay Area and Seattle houses are not a good deal. Why do so many people buy them?	1.592110566	最近西雅图比湾区便宜的...	Who says Seattle's cheaper than the Bay Area...	0.822382861	0.5974
现在社招科次进谷歌工作是不是很难使用技术提升了...	Not sure whether it is very difficult to obtain technical upgrade via social recruitment into Google as programmers now...	1.855487559	经验分享的秘密宝...	Talk about the edgucations of Google...	0.684597194	0.3898
难道现在是房价下跌的第1个阶段？？	Is this the first stage of falling housing prices??	2.882785493	都这个时候了，你还要评选去买房子?	At this point, you're gonna go in there and buy a house?	0.734891767	0.4326
大厂前高层领导，每天到底干啥？	What do the senior leaders of big companies do every day?	2.149929156	向上管理，搞定老板...	Managing up, managing the boss...	1.192484231	0.4723

Figure 13: Headlines in the information-gap surprisingness validation experiment (measured by KL-divergence). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review

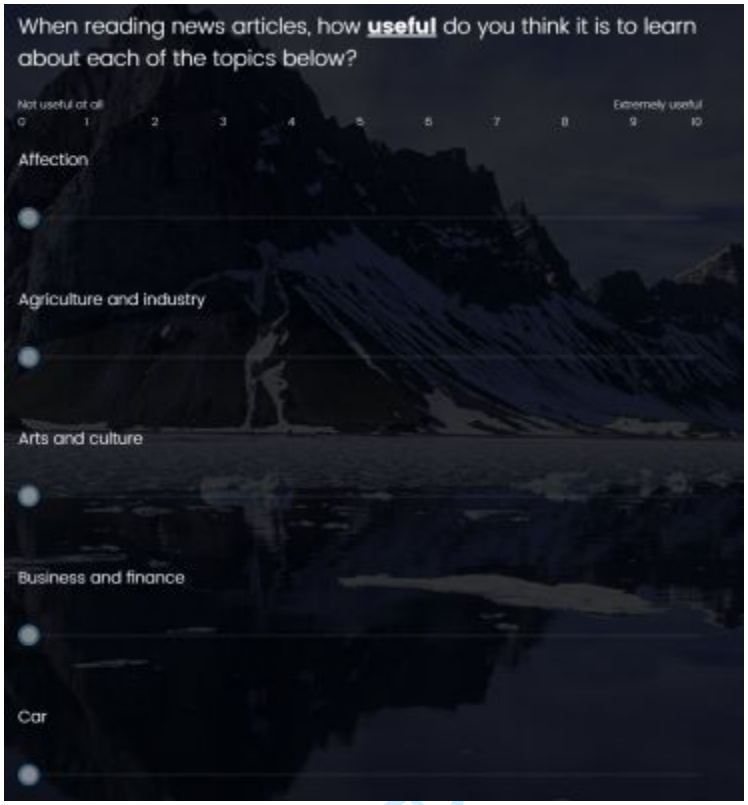


Figure 14: Screenshot of rating the usefulness of 36 different news topics on a Likert scale from 0 to 10

Author Manuscript Under Review

Author Manuscript

Appendix (For Online Publication Only)



Figure 1: WeChat Webpage

Table 5: OLS Regression Results (Excluding Duplicate Headlines)

Dependent Variable:	*ln* (Reading+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.029***	0.029***	0.033***	0.029***	0.029***	0.033***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Ellipsis	0.094***	0.094***	0.103***	0.094***	0.095***	0.104***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Exclamation mark	0.073***	0.073***	0.075***	0.073***	0.073***	0.075***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Position 1	0.677***	0.678***	0.675***	0.678***	0.678***	0.675***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Position 2	0.066***	0.066***	0.066***	0.066***	0.066***	0.066***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 3	-0.014	-0.014	-0.014	-0.014	-0.014	-0.014
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 4	-0.036**	-0.036**	-0.035**	-0.036***	-0.036***	-0.035**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 5	-0.057***	-0.057***	-0.057***	-0.057***	-0.057***	-0.057***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 6	-0.049***	-0.049***	-0.051***	-0.049***	-0.049***	-0.051***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Negation	0.019***	0.019***	0.018**	0.019***	0.019***	0.018**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
KL-divergence	0.016***	0.016***	0.039***	0.064***	0.064***	0.109***
	(0.005)	(0.005)	(0.009)	(0.019)	(0.019)	(0.026)
Sentiment	-0.077***	-0.077***	-0.068***	-0.077***	-0.077***	-0.067***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Quadratic KL-divergence				-0.013***	-0.013**	-0.018***
				(0.005)	(0.005)	(0.006)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	108,513	108,513	108,513	108,513	108,513	108,513

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript For Peer Review

Table 6: Negative Binomial Regression Results (Excluding Duplicate Headlines)

Dependent Variable:	*Reading*					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.029**	0.029**	0.034**	0.029**	0.029**	0.034**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Ellipsis	0.096***	0.096***	0.104***	0.097***	0.097***	0.105***
	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)	(0.021)
Exclamation mark	0.083***	0.083***	0.085***	0.083***	0.083***	0.085***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Position 1	0.898***	0.898***	0.893***	0.898***	0.898***	0.893***
	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)	(0.027)
Position 2	0.166***	0.166***	0.164***	0.166***	0.166***	0.164***
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Position 3	0.024	0.024	0.023	0.024	0.024	0.024
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Position 4	-0.039	-0.039	-0.039	-0.039	-0.039	-0.039
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Position 5	-0.072***	-0.072***	-0.073***	-0.071***	-0.072***	-0.072***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Position 6	-0.058**	-0.058**	-0.062**	-0.058**	-0.058**	-0.062**
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Negation	0.026***	0.025***	0.022**	0.025***	0.025***	0.022**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
KL-divergence	0.019***	0.019***	0.040***	0.073***	0.073***	0.111***
	(0.006)	(0.006)	(0.012)	(0.024)	(0.024)	(0.033)
Sentiment	-0.088***	-0.088***	-0.075***	-0.088***	-0.088***	-0.073***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Quadratic KL-divergence				-0.014**	-0.014**	-0.018**
				(0.006)	(0.006)	(0.008)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	108,363	108,363	108,363	108,363	108,363	108,363
Over-dispersion	3.6978	3.6988	3.7282	3.6982	3.6991	3.7286

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. Note: 42 fixed-effects (150 observations) removed because of only 0 outcomes.

Author Manuscript

For Peer Review

Table 7: OLS Regression Results (Imputing 171,000 as the Maximum Readings)

Dependent Variable:	*ln* (Reading+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.029***	0.029***	0.034***	0.030***	0.029***	0.034***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Ellipsis	0.087***	0.087***	0.097***	0.088***	0.088***	0.098***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Exclamation mark	0.067***	0.067***	0.070***	0.067***	0.067***	0.070***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Position 1	0.691***	0.691***	0.689***	0.691***	0.691***	0.689***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Position 2	0.076***	0.076***	0.076***	0.075***	0.076***	0.076***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 3	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 4	-0.033**	-0.033**	-0.033**	-0.034**	-0.033**	-0.033**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 5	-0.061***	-0.061***	-0.061***	-0.062***	-0.061***	-0.062***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 6	-0.049***	-0.049***	-0.051***	-0.049***	-0.049***	-0.051***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Negation	0.021***	0.021***	0.020***	0.021***	0.021***	0.021***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
KL-divergence	0.015***	0.015***	0.036***	0.066***	0.066***	0.112***
	(0.005)	(0.005)	(0.009)	(0.020)	(0.020)	(0.027)
Sentiment	-0.073***	-0.073***	-0.065***	-0.073***	-0.073***	-0.064***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Quadratic KL-divergence				-0.014***	-0.013***	-0.019***
				(0.005)	(0.005)	(0.006)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript For Peer Review

Table 8: Negative Binomial Regression Results (Imputing 171,000 as the Maximum Readings)

Dependent Variable:	*R*eading					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.031**	0.031**	0.036***	0.031**	0.031**	0.037***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Ellipsis	0.089***	0.089***	0.097***	0.090***	0.090***	0.098***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Exclamation mark	0.076***	0.076***	0.078***	0.076***	0.076***	0.079***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Position 1	0.920***	0.920***	0.915***	0.920***	0.920***	0.915***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Position 2	0.177***	0.177***	0.175***	0.177***	0.177***	0.176***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Position 3	0.032	0.032	0.031	0.032	0.032	0.031
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Position 4	-0.037	-0.037	-0.037	-0.037	-0.037	-0.037
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Position 5	-0.079***	-0.079***	-0.080***	-0.079***	-0.079***	-0.080***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Position 6	-0.060**	-0.060**	-0.064**	-0.060**	-0.059**	-0.064**
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Negation	0.029***	0.029***	0.026***	0.029***	0.029***	0.026***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
KL-divergence	0.018***	0.018***	0.038***	0.078***	0.077***	0.116***
	(0.006)	(0.006)	(0.012)	(0.024)	(0.024)	(0.034)
Sentiment	-0.085***	-0.085***	-0.071***	-0.085***	-0.085***	-0.070***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Quadratic				-0.016**	-0.016**	-0.020**
KL-divergence				(0.006)	(0.006)	(0.008)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901
Over-dispersion	3.5635	3.5645	3.5917	3.5639	3.5649	3.5922

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript

For Peer Review

Table 9: Regression Results of the Five Validation Experiments

Dependent Variables:	Salience		Importance	Surprisingness	
	(1)	(2)	(3)	(4)	(5)
Question mark	0.057*				
	(0.029)				
Ellipsis		0.058**			
		(0.028)			
Exclamation mark			0.233***		
			(0.033)		
Negation				0.071**	
				(0.035)	
KL-divergence					0.110***
					(0.023)
Individual FE	Yes	Yes	Yes	Yes	Yes
Paired headline FE	Yes	Yes	Yes	Yes	Yes
Observations	2,000	2,000	2,000	2,000	2,000

Standard errors clustered at the individual level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript
For Peer Review

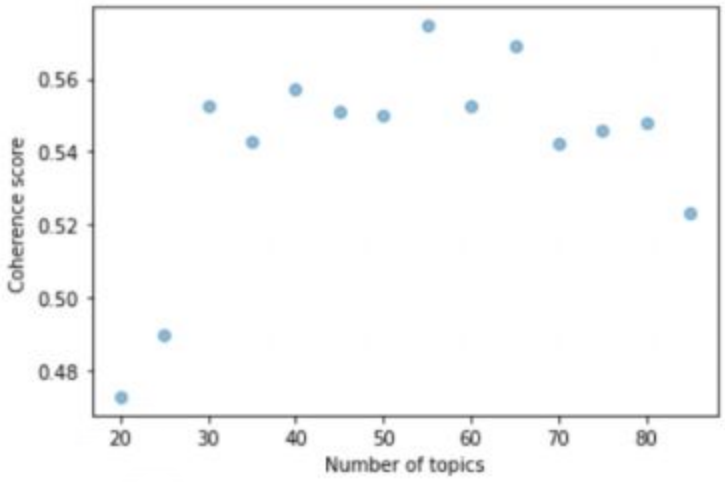


Figure 2: Scatterplot of coherence and number of topics (WeChat)

Author Manuscript
For Peer Review

Mandarin	非	别	不	没	无	勿	阻止	制止
English	not	don't	no	no	no	no	stop	stop
Mandarin	差	未	难	跌停	止	开除	莫	愁
English	poor	not	difficult	fall	stop	expel	no	sorrow
Mandarin	否	难	缺	乏	苦	艰	危	险
English	no	difficult	short of	short of	no	difficult	dangerous	risky

Figure 3: Negation words

Author Manuscript

For Peer Review

Author Manuscript

Table 10: Topic Tags' Regression Results of Specification

Topic tag		Topic Tag	
Affection	0.075	Agriculture and industry	0.020
	(0.049)		(0.122)
Art and culture	0.172*	Business and finance	0.040
	(0.088)		(0.050)
Car	0.024	Cartoon	-0.292
	(0.102)		(0.228)
Celebrity	0.405***	Documentary	0.369*
	(0.140)		(0.220)
Education and training	0.029	Fashion and skincare	0.109*
	(0.048)		(0.063)
Fate and constellation	-0.197	Fiction	-0.120
	(0.302)		(0.133)
Food and catering	-0.027	Game	0.053
	(0.069)		(0.101)
Health	0.107*	History	0.132*
	(0.060)		(0.079)
Joke	-0.052	Literature	0.060
	(0.071)		(0.052)
Military	0.241***	Marketing	0.018
	(0.092)		(0.483)
Motivation	-0.018	Movie	0.110
	(0.106)		(0.106)
Music	0.014	Parenting	-0.056
	(0.082)		(0.078)
Pet	-0.108	Photography	0.297
	(0.166)		(0.294)
Politics	0.227***	Real estate	0.098
	(0.062)		(0.076)
Religion	-0.139	Science	0.220
	(0.167)		(0.135)
Society	0.188***	Sports	0.188
	(0.042)		(0.117)
Technology	0.004	Travel	0.063
	(0.079)		(0.071)
TV series	-0.087	Workplace	0.149
	(0.132)		(0.098)

Standard errors clustered at the individual level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

For Peer Review

Table 11: OLS Regression Results with Alternative Sentiment Analysis

Dependent Variable:	*ln* (Reading+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	0.032***	0.032***	0.035***	0.032***	0.032***	0.036***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Ellipsis	0.090***	0.090***	0.098***	0.091***	0.091***	0.100***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Exclamation	0.072***	0.071***	0.073***	0.071***	0.071***	0.073***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Position 1	0.690***	0.690***	0.688***	0.690***	0.690***	0.688***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Position 2	0.075***	0.075***	0.075***	0.075***	0.075***	0.075***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 3	-0.010	-0.010	-0.009	-0.010	-0.010	-0.009
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 4	-0.034**	-0.034**	-0.033**	-0.034**	-0.034**	-0.033**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 5	-0.062***	-0.062***	-0.062***	-0.062***	-0.062***	-0.062***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Position 6	-0.050***	-0.050***	-0.051***	-0.049***	-0.049***	-0.051***
	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)
Negation	0.027***	0.027***	0.025***	0.027***	0.027***	0.025***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
KL-divergence	0.013**	0.013**	0.035***	0.066***	0.066***	0.113***
	(0.005)	(0.005)	(0.009)	(0.020)	(0.020)	(0.027)
Valence	-0.008***	-0.008***	-0.007***	-0.008***	-0.008***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Arousal	0.002	0.002	0.003	0.002	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Quadratic KL-divergence				-0.014***	-0.014***	-0.020***
				(0.005)	(0.005)	(0.006)
Topic FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
Source FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,901	112,901	112,901	112,901	112,901	112,901

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively.

Author Manuscript For Peer Review

Author Manuscript

Table 12: Summary of Measures for Facebook Dataset

Construct	Measure	Source
Salience	1: Question mark	Punctuation
	2: Ellipsis	Punctuation
	3: Wh-question sentence	NPS Chat Corpus
Importance	1: Exclamation mark	Punctuation
	2: Emphasis sentence	NPS Chat Corpus
Surprisingness	1: Reject sentence	NPS Chat Corpus
	2: KL-divergence	LDA
Valence	Sentiment of headline	qdap R package

For Peer Review

Table 13: Descriptive Statistics for Facebook Dataset

Statistics	Mean	St. Dev.	Min	Max
Binary variables			0	1
Question mark	0.035	0.183		
Ellipsis	0.294	0.455		
Wh-question	0.054	0.226		
Exclamation mark	0.004	0.061		
Emphasis sentence	0.040	0.195		
Reject sentence	0.014	0.116		
Continuous variables				
KL-divergence	1.956	0.303	0.071	3.900
Sentiment of title	-0.006	0.137	-0.951	0.962
Facebook popularity	129.363	661.182	0	49211
				N = 81,637

Author Manuscript For Peer Review

Table 14: OLS Regression Results for Facebook Data

Dependent Variable:	*l*n (Popularity+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	-0.039	-0.039	-0.181***	-0.040	-0.039	-0.181***
	(0.036)	(0.036)	(0.064)	(0.036)	(0.036)	(0.064)
Ellipsis	0.593***	0.569***	0.215***	0.590***	0.566***	0.214***
	(0.014)	(0.014)	(0.032)	(0.014)	(0.014)	(0.032)
Wh-question	0.529***	0.515***	0.148***	0.528***	0.515***	0.148***
	(0.029)	(0.029)	(0.028)	(0.029)	(0.029)	(0.028)
Exclamation mark	0.532***	0.511***	0.181**	0.535***	0.514***	0.181**
	(0.107)	(0.107)	(0.085)	(0.107)	(0.107)	(0.085)
Emphasis	0.120***	0.117***	0.077***	0.118***	0.114***	0.076***
	(0.033)	(0.033)	(0.028)	(0.033)	(0.033)	(0.028)
Reject	0.189***	0.194***	0.129***	0.191***	0.196***	0.129***
	(0.056)	(0.056)	(0.045)	(0.056)	(0.056)	(0.045)
KL-divergence	-0.160***	-0.170***	-0.034	0.705***	0.706***	0.181
	(0.027)	(0.027)	(0.041)	(0.144)	(0.144)	(0.205)
Sentiment	-0.406***	-0.398***	-0.207***	-0.409***	-0.401***	-0.208***
	(0.048)	(0.047)	(0.035)	(0.048)	(0.047)	(0.035)
Quadratic KL-divergence				-0.213***	-0.215***	-0.051
				(0.035)	(0.035)	(0.045)
LDA topic FE	No	No	Yes	No	No	Yes
Source FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
News topic FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	81,637	81,637	81,350	81,637	81,637	81,350

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. Note: In column (3) and column (6), where we add source fixed effects and LDA topic fixed effects, some observations are dropped because of the lack of sources of some articles in the Facebook dataset.

Author Manuscript

For Peer Review

Table 15: Negative Binomial Regression Results for Facebook Data

Dependent Variable:	*Popularity*					
	(1)	(2)	(3)	(4)	(5)	(6)
Question mark	-0.354***	-0.353***	-0.350***	-0.354***	-0.353***	-0.351***
	(0.041)	(0.041)	(0.081)	(0.041)	(0.041)	(0.081)
Ellipsis	0.383***	0.363***	0.214***	0.383***	0.363***	0.213***
	(0.016)	(0.016)	(0.039)	(0.016)	(0.016)	(0.039)
Wh-question	0.308***	0.312***	0.166***	0.306***	0.311***	0.167***
	(0.033)	(0.033)	(0.045)	(0.033)	(0.033)	(0.045)
Exclamation_mark	0.164	0.146	0.223*	0.165	0.146	0.223*
	(0.120)	(0.120)	(0.121)	(0.120)	(0.120)	(0.121)
Emphasis	0.222***	0.207***	0.161***	0.222***	0.207***	0.160***
	(0.037)	(0.037)	(0.056)	(0.037)	(0.037)	(0.056)
Reject	0.281***	0.314***	0.350*	0.282***	0.314***	0.348*
	(0.063)	(0.063)	(0.189)	(0.063)	(0.063)	(0.188)
KL-divergence	-0.036	-0.054*	0.063	-0.152	-0.125	0.584**
	(0.029)	(0.029)	(0.053)	(0.173)	(0.174)	(0.289)
Sentiment	-0.388***	-0.349***	-0.330***	-0.387***	-0.349***	-0.332***
	(0.054)	(0.054)	(0.074)	(0.054)	(0.054)	(0.075)
Quadratic KL-divergence				0.027	0.017	-0.125*
				(0.040)	(0.040)	(0.066)
LDA topic FE	No	No	Yes	No	No	Yes
Source FE	No	No	Yes	No	No	Yes
Published time FE	No	Yes	Yes	No	Yes	Yes
News topic FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	81,637	81,637	78,023	81,637	81,637	78,023
Over-dispersion	0.23467	0.23543	0.48534	0.23467	0.23543	0.48539

Standard errors clustered at the source level are in parentheses; *, **, and *** denote significance at the 0.1, 0.05, and 0.01 levels, respectively. Note: In column (3) and column (6), where we add source fixed effects and LDA topic fixed effects, some observations are dropped because of the lack of sources of some articles in the Facebook dataset.

Author Manuscript For Peer Review

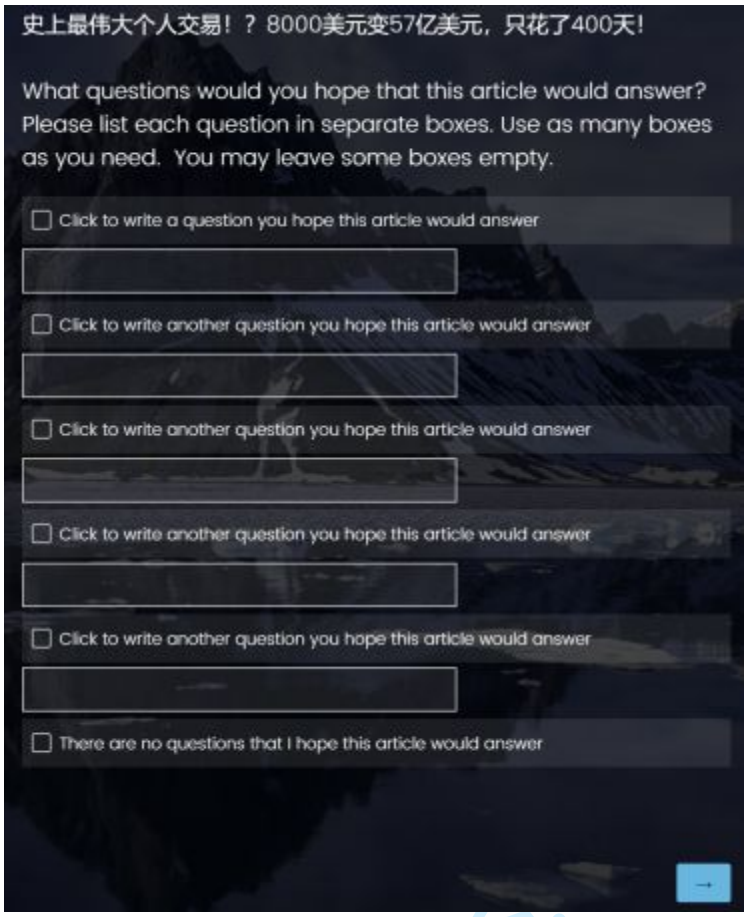


Figure 4: Screenshot of the information-gap salience validation experiment (measured by the presence of a question mark)

Author Manuscript
For Review

Originally with question mark	Originally with question mark	Originally without question mark	Originally without question mark	Semantic similarity
史上最伟大个人交易！！8000美元变57亿美元，只花了480天！	Greatest personal deal ever! ?$8,000 became $5.7 billion in just 480 days!	收益超349%！靠这个股票，巴菲特5年狂赚1200亿美元！	The income is over 349%! With this stock, Buffet earned $120 billion in 5 years!	0.452171296
取消SAT成绩要求！！加州大学录取现在看什么？	Dropping SAT score requirements?! What does the University of California admission look like now?	加州，不许给贫生打低分！为了推存公平！	California: Don't give poor students a low score! For education fairness!	0.522243923
全美首个流执行死刑的中国留学生，犯了什么罪？	What crime did the first Chinese student in the United States execute or execute?	让歌案震网！美国留学生爆杀案，似加此简单！	The Jiang Ge case is sensational! The murder case of American students is also so bizarre!	0.515795899
在美国这个职业，居然没有work life balance吗？	Is there no work life balance in this profession in the United States?	在美国工作没"大饭"，就该如何节心的活着	How can I live happily if I don't have a "big bag" to work in the United States	0.483206153
美国将要发280万张绿卡！排期还能交钱插队？！	The United States will issue 2 million green cards! Can I still pay to cut the queue? !	视频了！特朗16万移民直接拿绿卡！	I'm sorry! Amazing for 160,000 immigrants to get green cards directly!	0.516702403
交了5个月房租，发现房东竟是我老婆！！	After paying 5 months rent, I found out that the landlord is my wife! ?	惊区canada炒房经历，1个月内买到自住房！	Buy Area condo bargain hunting experience, buy a house within 1 month!	0.483148703
什么职业既优雅，又有"钱"景？	What sideline is both elegant and "money"?	推荐一个很适合女性的副业	Recommend a side business that is very suitable for women	0.468853824
实习生是写大厂代码是什么体验！	What is the experience of the intern rewriting the code of the big factory?	分享analyscall跳槽到大厂的经历和感想	Share the experience and impressions of analytics jumping to Dachang	0.458224714
为什么跳槽，肯定要比养上去的拿得多！	Why jump jobs, must get more than the promotion!	今年2次跳槽，工资从$4.5万涨到$13万！	Twice this year, the salary has increased from $45,000 to $130,000!	0.469226803
被面试官用十招处死面试？没永攻略来啦！	Was interviewed by the interviewer with ten tricks? Counterattack strategy is here!	加拿大找工求职逆袭开这些坑！	Please avoid these pits when looking for a job in Canada!	0.458164698

Figure 5: Headlines in the information-gap salience validation experiment (measured by the presence of a question mark). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review

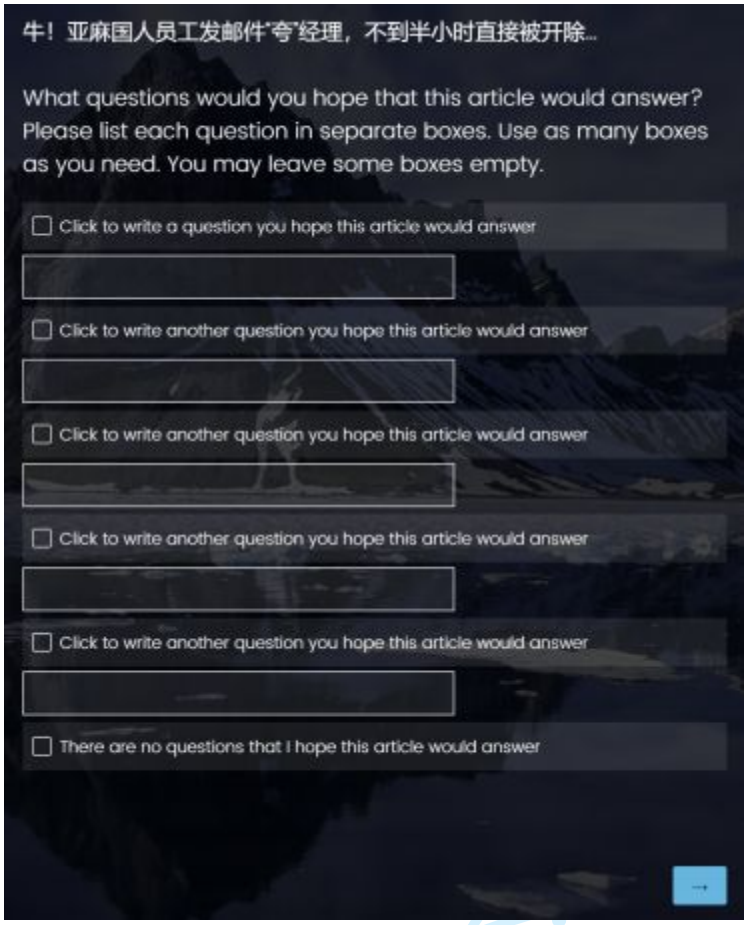


Figure 6: Screenshot of the information-gap salience validation experiment (measured by the presence of an ellipsis).

Author Manuscript
For Review

Originally with ellipsis	Originally with ellipsis	Originally without ellipsis	Originally without ellipsis	Semantic similarity
牛！亚裔员工觉醒开"全"狂喷，不到半小时高级被开除…	Staggering! How workers sent emails to "praise" the manager and were fired directly in less than half an hour…	忘了！老板下班后发信息给我，马上！	Dead! It's illegal for my boss to text me after getting off work!	0.407233924
2022美国最卡月薪榜，Zoom跌至垫底，第一是…	In the 2022 Top 100 U.S. Employers List, Zoom fell to the bottom, the first is…	2021，科技公司最全工资揭秘！	In 2021, the most complete salary range of technology companies!	0.463159582
再见！超过58%的湾区人正打算永远离开…	Goodbye! Over 58% of Bay Area residents are planning to leave forever…	从湾区到 Google 苏黎世，离开湾区这两年的阶感悟！	From the Bay Area to Google Zurich, the new insights of leaving the Bay Area in the past two years!	0.462497652
震惊觉神话！22岁大学生靠卖自拍，20天赚了100万美元…	Jokes become myths! 22-year-old college student made $1 million in 20 days by selling selfies…	一天稳赚七八百，业种一个超适合留学生的副业，时间地点自由！	Earn seven or eight hundred RMB a day, recommend a side business that is super suitable for international students, and the time and place are flexible!	0.41891697
突发！1人遇难4人受伤，中国留学生换遭严重车祸…	Sudden! 1 killed, 4 injured, Chinese students suffered a serious car accident…	突发！纽约高楼发生火灾！一群青少年跳窗逃脱，多人伤亡！	Sudden! New York high-rise building fire! A group of teenagers jumped out of the window to escape, many casualties!	0.458682156
最新！全球最具吸引力公司排名，华为上榜！腾讯34，阿里见…	Up to date! The world's most attractive company ranking, Huawei is on the list! FB No. 34, No. 1 is…	全球程序员收入排行榜！最高工资120万美元！	Global Programmer Income Ranking! $1.2 million maximum salary!	0.402175754
内部员工曝料：谷歌，脸书面试大造水！科技公司已沦跑单狂人了……	Internal staff broke the news: Google, Facebook interviews are full of water! Tech companies are already on draw leaves to sweat…	Google已经不是最难进的大厂了！	Google is not the most difficult big company to enter!	0.407426566
3个孩子和9岁弟弟尸体同住一年，遗弃他们的父母竟被释放了…	3 children live with 9-year-old brother's body for a year, parents who abandoned them were released…	揭露！美国女网红离奇死亡，警察竟搞本酷大聊同发现了8具尸体！	Outrageous! The mysterious death of a female Internet celebrity in the United States, the police found 9 bodies during the pursuit of her fiancé!	0.416379511
谷歌招很9有offer！我的申请经验是这样的…	Strongly won 9 offers! My application experience is like this…	斯坦学科技公司面试官经验，9个只意会！	Come to share the experience of interviewers in technology companies, 9 points to pay attention to!	0.44659543
微软连开价687亿美元买下它！天哪扫了…	Microsoft paid $68.7 billion to buy it! so magical…	工作三年后终于离开了微软！	Finally left Microsoft after working for three years!	0.52946582

Figure 7: Headlines in the information-gap salience validation experiment (measured by the presence of an ellipsis). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review



Figure 8: Screenshot of the information-gap importance validation experiment (measured by the presence of an exclamation mark).

Author Manuscript
For Peer Review

Originally with exclamation mark	Originally with exclamation mark	Originally without exclamation mark	Originally without exclamation mark	Semantic similarity
就调研！哥大罢工了，可是留学生们除了	Outrageous! Columbia University strikes, but foreign students are miserable...	全校大罢工？这所名校的学生连网课也上不成了...	A school-wide strike? Students in this famous university can't even take online courses...	0.671389581
英国工资该去低了吧！	Wages in the UK are too low!	法国工资比德国还低？	Wages are lower in France than in Germany?	0.626296401
做去去厂！到码太难！不如换个方向？	Want to go to a big factory? Transcoding is too difficult! Why don't we go the other way?	大厂去年，码农还能转什么？	Eight years in a big factory, what else can a code farmer do?	0.587154448
毕业3年前美买房拿身份？！就能带新得学吗！	Graduate 3 years to stay in the United States to buy a house to take identity? ! Can I study abroad again?	留学生毕业，究竟要不要回国？美好你内容想知道的事	Should international students return home after graduation? I urge you to think about these things calmly	0.561803844
买房差了几十万首付，又失房复得的危险经历！	Buy a house lost hundreds of thousands of down payment, and lost the dangerous experience!	差6万没买到dream house，真的好难受...	I am so sad that I failed to buy a dream house for 58,900 dollars...	0.558045405
10分钟，加州一居民被骗走1160万美元巨款！	California resident cheated out of $11.6 million in 10 minutes!	10秒抢走房方主？加州城市无房产天堂？	Robs in 10 seconds? California's $0 paradise?	0.552510829
防不胜防！强奸时突然昏迷，天王老识，醒来发现被光了...	Impossible to guard against! Party suddenly coma, lose consciousness, wake up to find stabbed...	半夜思念老人，疫情之下不知什么时候能团聚...	Missing the elderly in the middle of the night, I do not know when to reunite under the epidemic...	0.512328651
推荐一个很适合女性的小众高薪副业，600-800元/天！	Recommend a niche, high-paying sideline for women: 600~800 RMB per day!	推荐一个很适合女性的副业	Recommend a side hustle for women	0.667689204
西区seaside外观性别，1个月内买到自住房！	Buy Asia condo fishing experience, buy a house in 1 month!	西区亚难自建房实录（开始设计）	Hard To Build a House in Bay Area (Start designing)	0.510864203
难道现在是房价下跌的第1个阶段！！	Is this the first stage of falling housing prices?!	都这个时候了，你还要评进去买房子？	You want to rush in and buy a house at this hour?	0.510055771

Figure 9: Headlines in the information-gap importance validation experiment (measured by the presence of an exclamation mark). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review

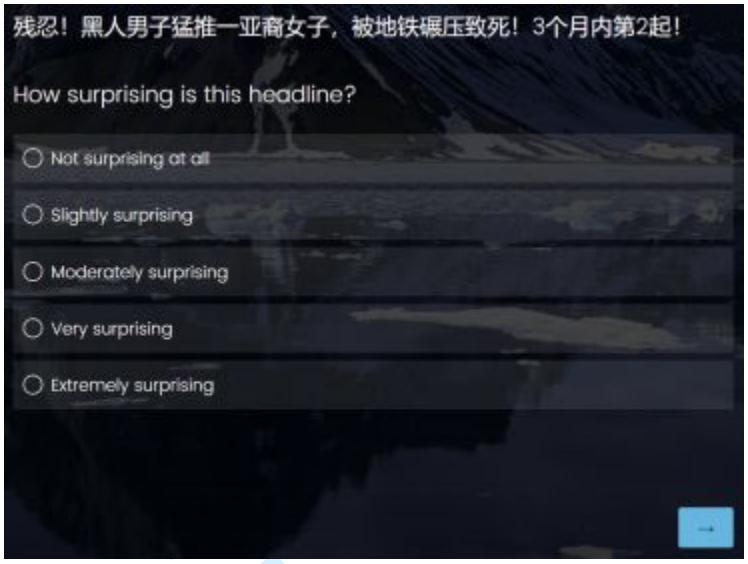


Figure 10: Screenshot of the information-gap surprisingness validation experiment (measured by classification as a negation)

Author Manuscript
Peer Review



with negation	with negation	without negation	without negation	Semantic Similarity
亚裔女生为救同胞，却惨遭黑人群殴！现场无一人施救！	Asian girls were beaten by black people to save their compatriots! No one was there to help!	残忍！黑人男子猛推一亚裔女子，被地铁碾压致死！3个月内第2起！	Cruel! A black man shoved an Asian woman and was crushed to death by the subway! 2nd in 3 months!	0.723154008
全校大罢工？这所名校的学生连网课也上不成了...	A school-wide strike? Students at this prestigious school can't even take online classes...	就离谱！哥大罢工了，可是留学生们掉了...	Outrageous! Columbia University is on strike, but the international students are miserable...	0.671105981
没绿卡？！强烈建议别去Facebook！	No permanent residency? ! Strongly recommend not going to Facebook!	活久见！花钱直接拿绿卡？！合法洗钱来了？	Live longer see more! Spend money directly to get permanent residency? ! Is it legal to make money that way?	0.578898665
差5万没买到dream house，真的好难受...	It's really sad that I didn't buy a dream house for $50,000...	100w美金，在中部能买什么样的房...	100 million US dollars, what kind of house can I buy in the middle...	0.54601264
谷歌面试官求你们了！面试的时候别这么整！	Google interviewers beg you! Don't do this during the interview!	想跳槽！你可能被谷歌安全部盯上了！	Want to jump ship? You may be targeted by Google Security!	0.545233369
骗子项目！留学花费几十万，回国不能通过学历认证！	Liar Project! Studying abroad costs hundreds of thousands, and this cannot pass the academic certification when returning to China!	留学生毕业回国后多久能赚够一百万！	How long does it take for international students to earn one million after graduation?	0.523976445
"我留了两年学，就再也没法回国了"	"I stayed for two years to study abroad, and I will never be able to return to China"	"这一届留学生，活得可太容易了。"	"Life is too easy for this class of international students."	0.523976445
炒股炒股不如纯打工？！湾区人实现财富自由的方式竟是……	Investing in stocks or currency cannot do better than just working for others? ! The way people in the Bay Area achieve financial freedom is actually...	那些早早迈入富人圈的湾区人，都做对了同一件事...	Those Bay Area people who became rich early did the same thing right...	0.527217567
接到大使馆电话，挂了7后，父母积蓄全没了...	After receiving the call from the embassy and pressing 7, my parents' savings were all gone...	留学4年，最怕给客服打电话，更...	After studying abroad for 4 years, I am most afraid of calling customer service. Crying...	0.490797371
想去大厂？转码太难！不如换个方向？	Want to go to a big company? Transiting to a new career in software engineering is too hard! Why not change direction?	中年转行成功！拿到大厂职位！如何平衡本职工作，面试和家庭？	Successful middle-aged career change! Got a job in a big company! How do you balance work, interviews, and family?	0.486133635

Figure 11: Headlines in the information-gap surprisingness validation experiment (measured by classification as a negation). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review

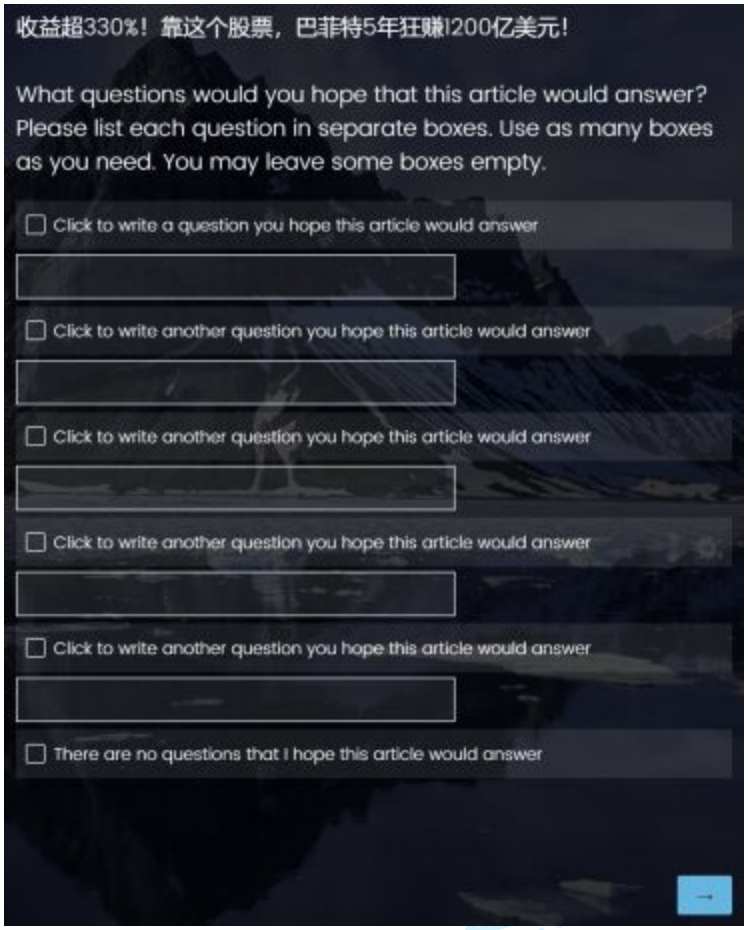


Figure 12: Screenshot of the information-gap surprisingness validation experiment (measured by KL-divergence)

Author Manuscript
For Review



High KL-divergence	High KL-divergence	KL-divergence	Low KL-divergence	Low KL-divergence	KL-divergence	Semantic Similarity
气愤！多名中国留学生遭重物击打头	Angry! Several Chinese students were hit in the head with heavy objects!	3.318968093	就离谱！育人第工了，可是接学马后悔了...	This is relentless! Columbia's on strike, but the foreign students are in trouble...	0.689418288	0.4732
发现7年间隐过的坑！	7 years in the United States to step on the pit!	1.835551708	科科7年，面试7轮，最终上岸！	Seven Years of transcending, seven rejected interviews, and finally went on the shore!	0.566308069	0.4758
没错卡了！强烈建议别去Facebook！	Without a green card? ! I strongly recommend not going to Facebook!	1.978028918	好消息！Facebook 的HR终于恢复办公Back to green card!	Good news! Facebook HR is finally	1.083841886	0.5893
突发！华人男子遭到枪当场射杀！华美最大华人区还安全吗？	Sudden! Chinese man shot dead in the street! Is America's largest Chinese enclave safe?	1.99468155	华人大哥被3个黑人抢劫，抢劫反杀，开枪1死2伤！	The Chinese brother was mugged by 3 black men. He drew his gun and killed 1 dead and 1 wounded!	1.014323091	0.5681
惊悚！自杀式爆炸再现，法雅车司机神操作救了一整座城市！	Terrifying! Suicide bombing, taxi driver saved an entire city!	1.944181668	突发！纽约高楼冒生火灾！一群青少年跳窗逃脱，多人伤亡！	Sudden! New York high-rises on fire! A bunch of teenagers jumped out of a window. Multiple casualties!	1.015368744	0.4467
毕业季子弹夺命中！2名前程明书三份保险	Chinese wife dies at home! Just bought a huge insurance policy 2 days ago...	2.8323884	晴天！然区突发枪战，2名毕业男童中弹身亡...	Ouch! Two-year-old Chinese boy shot dead in a shootout in the Bay Area...	0.850838501	0.4662
湾区和西雅图房子一点都不便宜，为什么那么多人买？	Bay Area and Seattle houses are not a good deal. Why do so many people buy them?	1.592110566	最近西雅图比湾区便宜的...	Who says Seattle's cheaper than the Bay Area...	0.822382861	0.5974
现在社招科次进谷歌工作是不是很难很得技术提升了...	Not sure whether it is very difficult to obtain technical upgrade via social recruitment into Google as programmers now...	1.835487559	经验分享|谷歌年化...	Talk about the valgarization of Google...	0.684597194	0.3898
难道现在是房价下跌的第1个阶段？？	Is this the first stage of falling housing prices??	2.882785493	都这个时候了，你还觉得这点买房子?	At this point, you're gonna go in there and buy a house?	0.734891767	0.4326
大厂前高层领导，每天到底干啥？	What do the senior leaders of big companies do every day?	2.149929156	向上管理，搞定老板...	Managing up, managing the boss...	1.192484231	0.4723

Figure 13: Headlines in the information-gap surprisingness validation experiment (measured by KL-divergence). Headlines are listed in the order the pairs are shown to participants. Within each pair, participants see both headlines in a randomized order.

Author Manuscript

For Peer Review

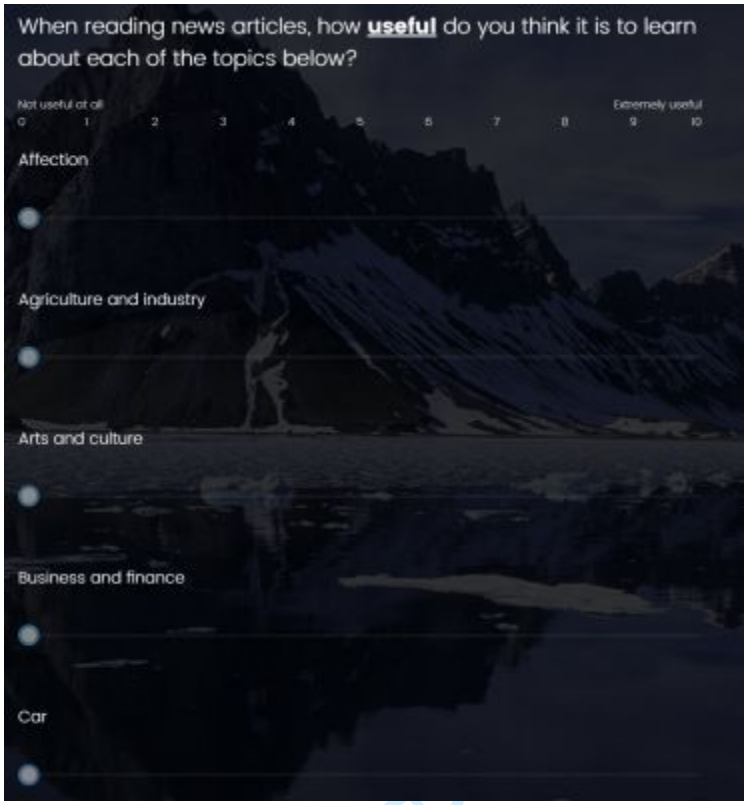
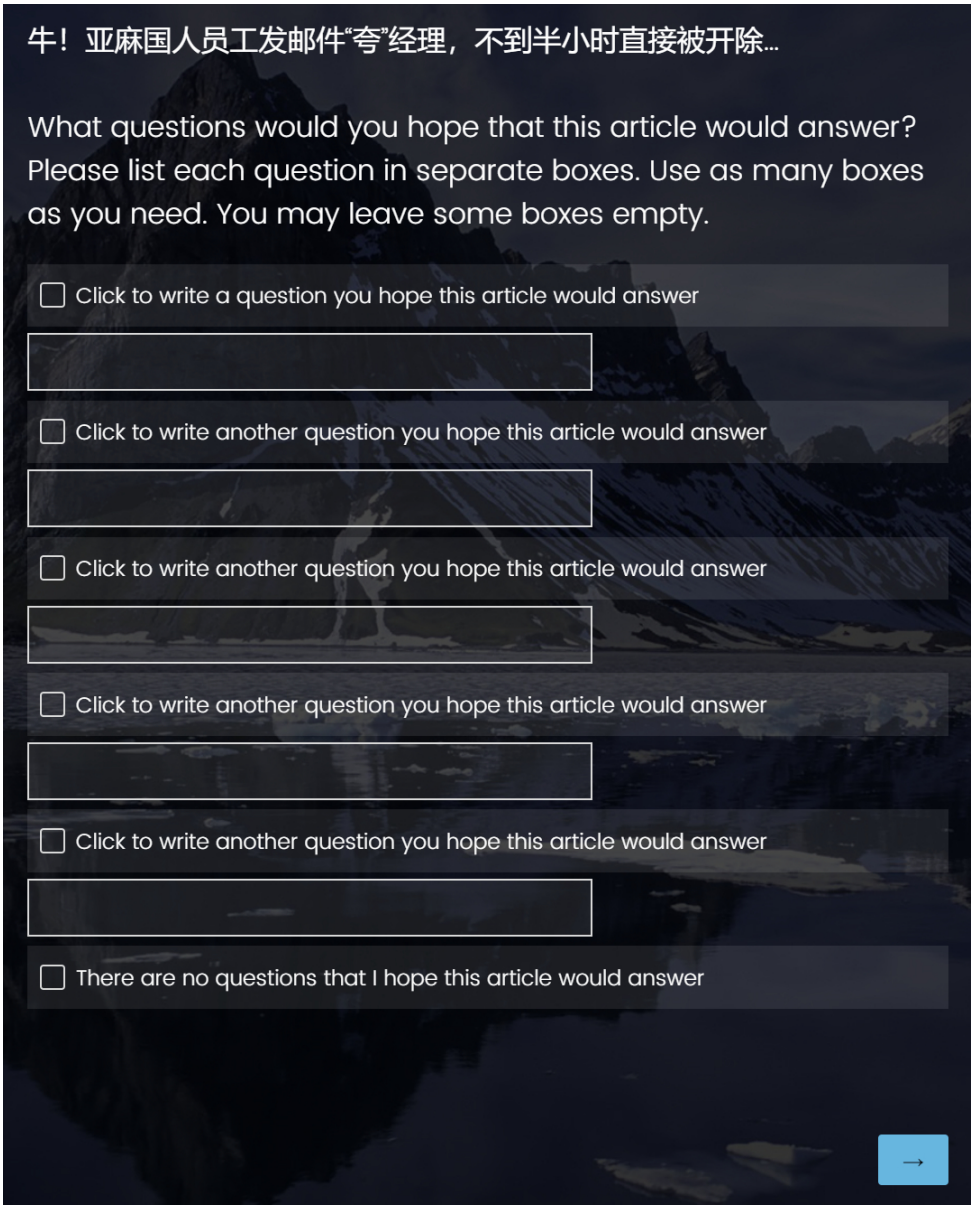


Figure 14: Screenshot of rating the usefulness of 36 different news topics on a Likert scale from 0 to 10

Author Manuscript
For Review



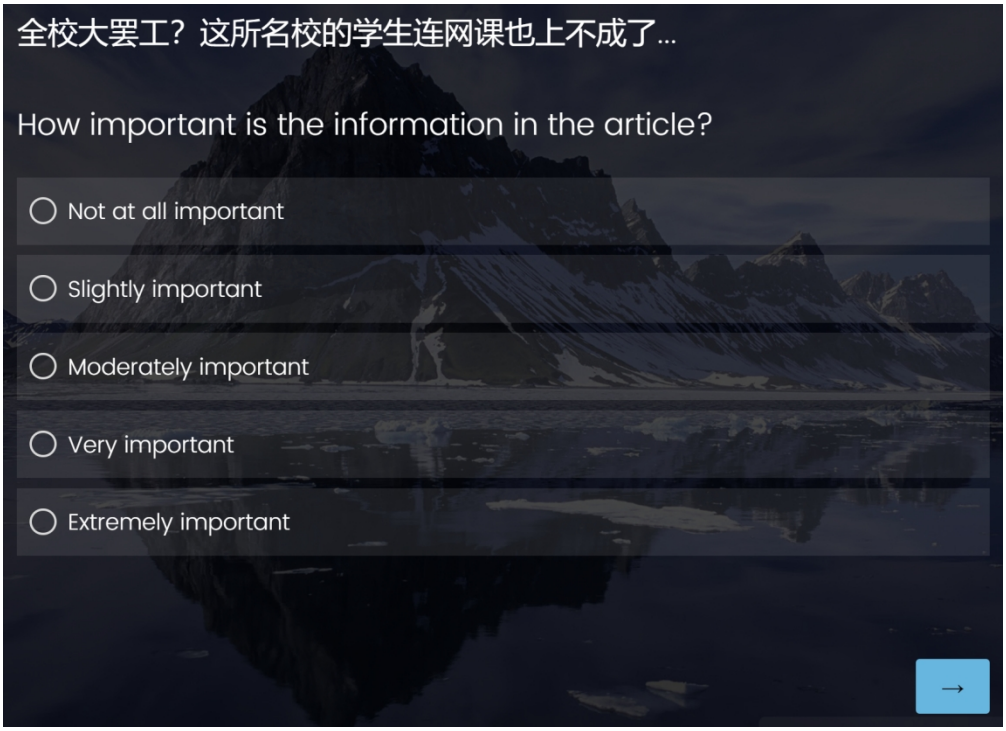
358x444mm (76 x 76 DPI)

Author Manuscript

Originally with ellipsis	Originally with ellipsis	Originally without ellipsis	Originally without ellipsis	Semantic similarity
牛！亚麻国人员工发邮件"夺"经理，不到半小时直接被开除…	Surprising! Flax workers sent emails to "praise" the manager and were fired directly in less than half an hour...	定了！老板下班后发信息给我，违法！	Deal! It's illegal for my boss to text me after getting off work!	0.407233924
2022美国雇主百强榜，Zoom跌至垫底，第一是...	In the 2022 Top 100 U.S. Employers List, Zoom fell to the bottom, the first is...	2021，科技公司最全工资范围！	In 2021, the most complete salary range of technology companies!	0.461193502
再见！超过50%的湾区人正打算永远离开...	Goodbye! Over 50% of Bay Area residents are planning to leave forever...	从湾区到 Google 苏黎世，离开湾区这两年的新感悟！	From the Bay Area to Google Zurich, the new insights of leaving the Bay Area in the past two years!	0.462497652
笑话变神话！22岁大学生靠卖自拍，20天赚了100万美元...	Jokes become myths! 22-year-old college student made $1 million in 20 days by selling selfies...	一天稳赚七八百，安利一个超适合留学生的副业，时间地点自由！	Earn seven or eight hundred RMB a day, recommend a side business that is super suitable for international students, and the time and place are flexible!	0.41891697
突发！1人遇难4人受伤，中国留学生惨遭严重车祸…	Sudden! 1 killed, 4 injured, Chinese students suffered a serious car accident...	突发！纽约高楼发生火灾！一群青少年跳窗逃脱，多人伤亡！	Sudden! New York high-rise building fire! A group of teenagers jumped out of the window to escape, many casualties!	0.430882156
最新！全球最具吸引力公司排名，华为上榜！FB第34，第1名是...	Up to date! The world's most attractive company ranking, Huawei is on the list! FB No. 34, No. 1 is...	全球程序员收入排行榜！最高工资120万美元！	Global Programmer Income Ranking! $1.2 million maximum salary!	0.402175754
内部员工爆料：谷歌、脸书面试大放水！科技公司已在跪着招人了……	Internal staff broke the news: Google, Facebook interviews are full of water! Tech companies are already on their knees to recruit...	Google已经不是最难进的大厂了！	Google is not the most difficult big company to enter!	0.407426566
3个孩子和9岁弟弟尸体同住一年，遗弃他们的父母竟被释放了...	3 children live with 9-year-old brother's body for a year, parents who abandoned them were released...	离谱！美国女网红离奇死亡，警察追捕未婚夫期间发现了9具尸体！	Outrageous! The mysterious death of a female Internet celebrity in the United States, the police found 9 bodies during the pursuit of her fiancé!	0.416379511
强势斩获9枚offer！我的申请经验是这样的...	Strongly won 9 offers! My application experience is like this...	来分享科技公司面试官经验，9个注意点！	Come to share the experience of interviewers in technology companies, 9 points to pay attention to!	0.44659543
微软竟斥资687亿美元买下它！太魔幻了...	Microsoft paid $68.7 billion to buy it! so magical...	工作三年后终于离开了微软！	Finally left Microsoft after working for three years!	0.52946502

824x318mm (76 x 76 DPI)

Author Manuscript



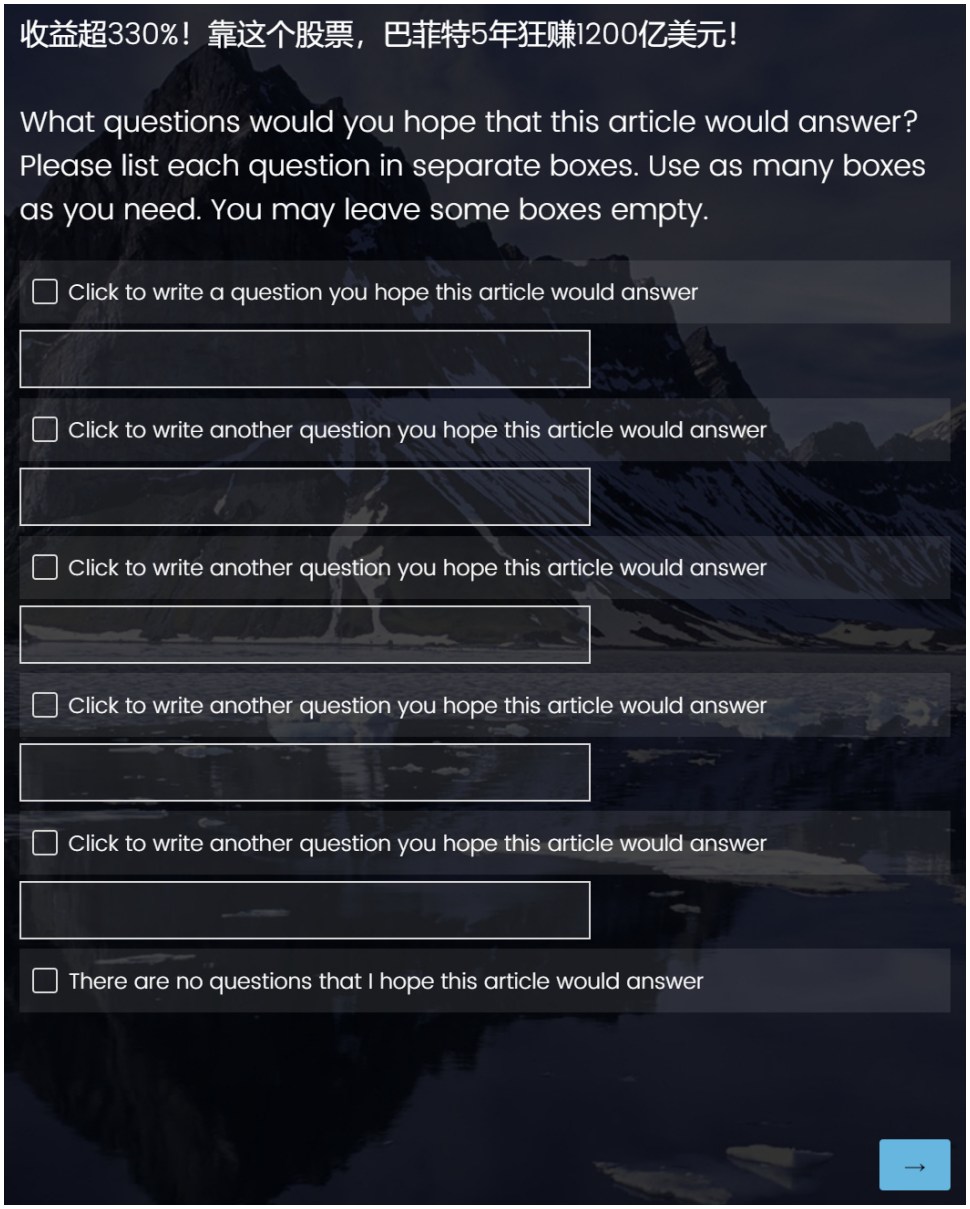
579x419mm (76 x 76 DPI)

Author Manuscript

Originally with exclamation mark	Originally with exclamation mark	Originally without exclamation mark	Originally without exclamation mark	Semantic similarity
就离谱！哥大罢工了，可是留学生们惨了	Outrageous! Columbia University strikes, but foreign students are miserable...	全校大罢工？这所名校的学生连网课也上不成了...	A school-wide strike? Students in this famous university can't even take online courses...	0.671105981
英国工资也太低了吧！	Wages in the UK are too low!	法国工资比德国还低？	Wages are lower in France than in Germany?	0.626296401
想去大厂？转码太难！不如换个方向？	Want to go to a big factory? Transcoding is too difficult! Why don't we go the other way?	大厂8年，码农还能转什么？	Eight years in a big factory, what else can a code farmer do?	0.587154448
毕业3年留美买房拿身份？！我能重新留学吗！	Graduate 3 years to stay in the United States to buy a house to take identity? ! Can I study abroad again?	留学生毕业，究竟要不要回国？我劝你冷静想想这些事	Should international students return home after graduation? I urge you to think about these things calmly	0.564003944
买房丢了几十万首付，又失而复得的危险经历！	Buy a house lost hundreds of thousands of down payment, and lost the dangerous experience!	差5万没买到dream house，真的好难受...	I am so sad that I failed to buy a dream house for 50,000 dollars...	0.536043465
10分钟，加州一居民被骗走1160万美元巨款！	California resident cheated out of $11.6 million in 10 minutes!	10秒抢走劳力士？加州成"0元购"天堂？	Rolex in 10 seconds? California a $0 paradise?	0.515210629
防不胜防！聚会时突然昏迷、失去意识，醒来发现被扎了…	Impossible to guard against! Party suddenly coma, lose consciousness, wake up to find stabbed...	半夜思念老人，疫情之下不知什么时候能团聚...	Missing the elderly in the middle of the night, I do not know when to reunite under the epidemic...	0.512128651
推荐一个很适合女性的小众高薪副业：600-800元/天！	Recommend a niche, high-paying sideline for women: 600-800 RMB per day!	推荐一个很适合女性的副业	Recommend a side hustle for women	0.667689204
湾区condo抄底经历，1个月内买到自住房！	Bay Area condo fishing experience, buy a home in 1 month!	湾区艰难自建房实录（开始设计）	Hard To Build a House in Bay Area (Start designing)	0.514964283
难道现在是房价下跌的第1个阶段！？	Is this the first stage of falling housing prices!?	都这个时候了，你还要冲进去买房子？	You want to rush in and buy a house at this hour?	0.510557771

809x281mm (76 x 76 DPI)

Author Manuscript



352x439mm (76 x 76 DPI)

Author Manuscript

High KL-divergence	High KL-divergence	KL-divergence	Low KL-divergence	Low KL-divergence	KL-divergence	Semantic Similarity
气愤！多名中国留学生遭重物击打头	Angry! Several Chinese students were hit in the head with heavy objects!	3.316964093	就离谱！哥大罢工了，可是留学生们惨了...	This is ridiculous! Columbia's on strike, but the foreign students are in trouble...	0.669410288	0.4732
来美7年间踩过的坑！	7 years in the United States to step on the pit!	1.935551708	转码7年，面试7拒，最终上岸！	Seven Years of transcoding, seven rejected interviews, and finally went on the shore!	0.566100069	0.4756
没绿卡？！强烈建议别去Facebook！	Without a green card? ! I strongly recommend not going to Facebook!	1.976028918	好消息！Facebook 的H1b终于恢复办绿	Good news! Facebook H1b is finally back to green card!	1.083041886	0.5891
突发！华人男子遭同胞当街射杀！全美最大华人区还安全吗？	Sudden! Chinese man shot dead in the street! Is America's largest Chinese enclave safe?	1.99468195	华人大哥被3个黑人抢劫，拔枪反杀，歹徒1死1伤！	The Chinese brother was mugged by 3 black men. He drew his gun and killed 1 dead and 1 wounded!	1.014321091	0.5681
恐怖！自杀式爆炸再现，出租车司机神操作救了一整座城市！	Terrifying! Suicide bombing, taxi driver saved an entire city!	1.940101668	突发！纽约高楼发生火灾！一群青少年跳窗逃脱，多人伤亡！	Sudden! New York high-rises on fire! A bunch of teenagers jumped out of a window. Multiple casualties!	1.015168744	0.4467
华裔妻子惨死家中！2天前刚购买巨额保险...	Chinese wife dies at home! Just bought a huge insurance policy 2 days ago...	2.8323984	痛心！湾区突发枪战，2岁华裔男童中弹身亡...	Ouch! Two-year-old Chinese boy shot dead in a shootout in the Bay Area...	0.850930501	0.4662
湾区和西雅图房子一点都不划算，为什么那么多人买？	Bay Area and Seattle houses are not a good deal. Why do so many people buy them?	1.592110566	谁说西雅图比湾区便宜的...	Who says Seattle's cheaper than the Bay Area...	0.82202861	0.5974
现在社招码农进谷歌工作是不是很难获得技术提升了...	Not sure whether it is very difficult to obtain technical upgrade via social recruitment into Google as programmers now...	1.835407559	说说谷歌的庸俗化...	Talk about the vulgarization of Google...	0.684593194	0.3896
难道现在是房价下跌的第1个阶段！？	Is this the first stage of falling housing prices!?	2.882783493	都这个时候了，你还要冲进去买房子？	At this point, you're gonna go in there and buy a house?	0.734891767	0.4326
大厂的高层领导，每天到底干啥？	What do the senior leaders of big companies do every day?	2.149929156	向上管理，搞定老板...	Managing up, managing the boss...	1.192484231	0.4723

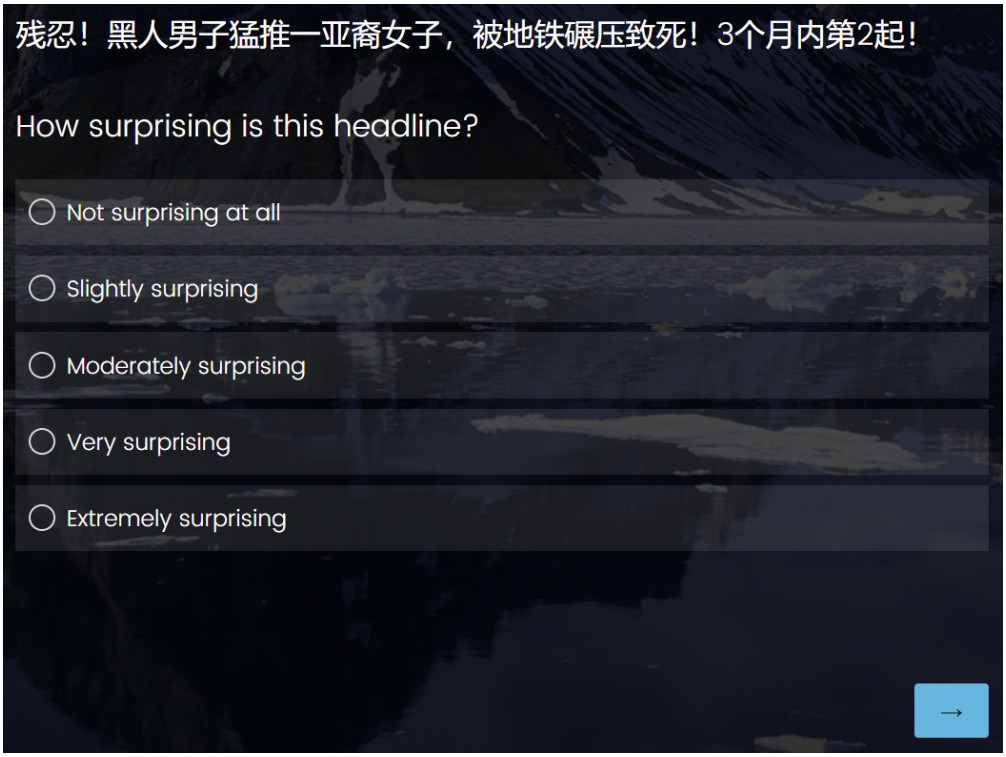
824x317mm (76 x 76 DPI)

Author Manuscript

Mandarin	非	别	不	没	无	勿	阻止	制止
English	not	don't	no	no	no	no	stop	stop
Mandarin	差	未	难	跌停	止	开除	莫	愁
English	poor	not	difficult	fall	stop	expel	no	sorrow
Mandarin	否	难	缺	乏	苦	艰	危	险
English	no	difficult	short of	short of	no	difficult	dangerous	risky

366x88mm (76 x 76 DPI)

Author Manuscript



350x262mm (76 x 76 DPI)

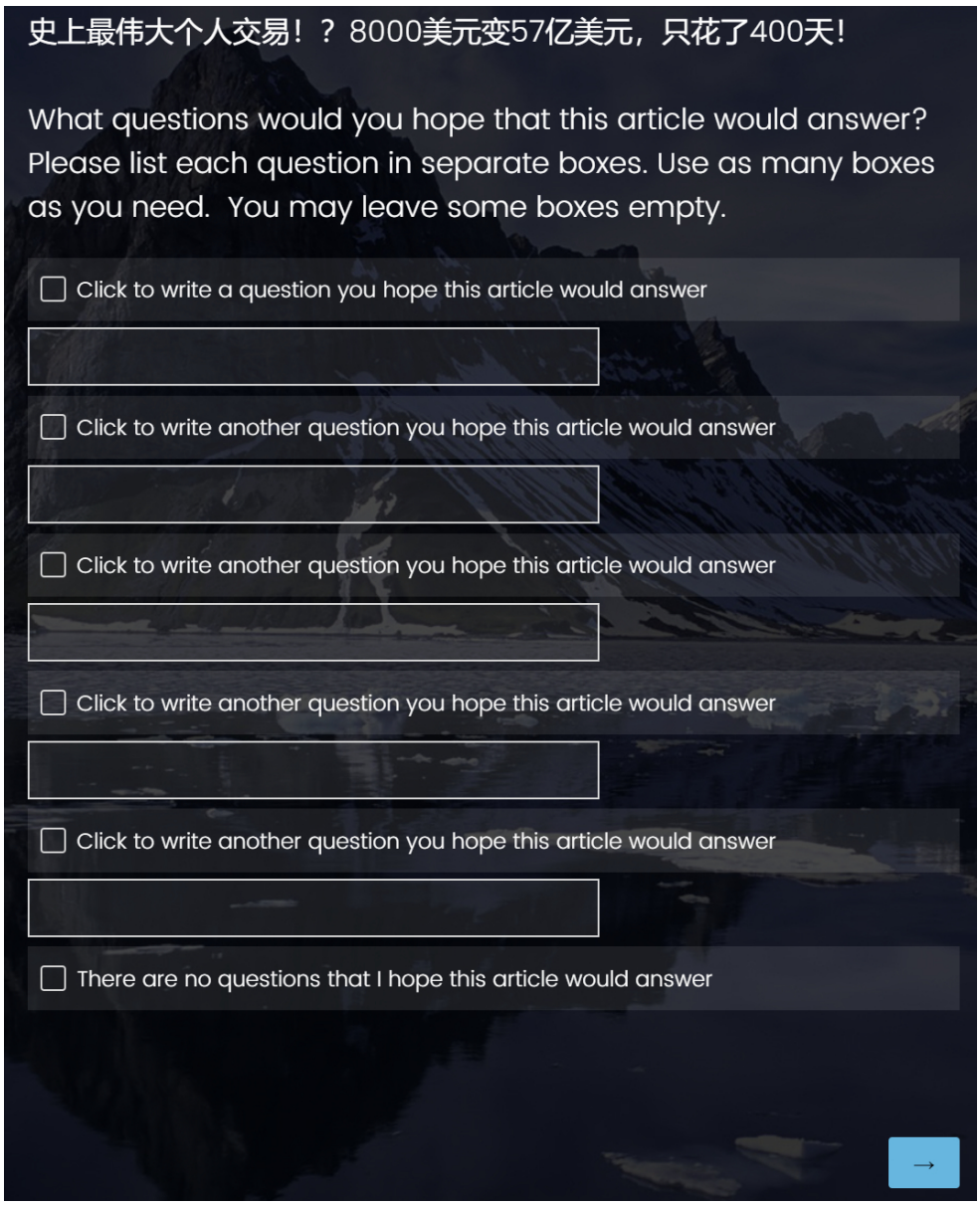
Author Manuscript

with negation	with negation	without negation	without negation	Semantic Similarity
亚裔女生为救同胞，却惨遭黑人群殴！现场无一人施救！	Asian girls were beaten by black people to save their compatriots! No one was there to help!	残忍！黑人男子猛推一亚裔女子，被地铁碾压致死！3个月内第2起！	Cruel! A black man shoved an Asian woman and was crushed to death by the subway! 2nd in 3 months!	0.723154008
全校大罢工？这所名校的学生连网课也上不成了...	A school-wide strike? Students at this prestigious school can't even take online classes...	就离谱！哥大罢工了，可是留学生们惨了...	Outrageous! Columbia University is on strike, but the international students are miserable...	0.671105981
没绿卡？！强烈建议别去Facebook！	No permanent residency? ! Strongly recommend not going to Facebook!	活久见！花钱直接拿绿卡？！合法敛财来了？	Live longer see more! Spend money directly to get permanent residency? ! Is it legal to make money that way?	0.578808665
差5万没买到dream house，真的好难受...	It's really sad that I didn't buy a dream house for $50,000...	100w美金，在中部能买什么样的房...	100 million US dollars, what kind of house can I buy in the middle...	0.54601264
谷歌面试官求你们了！面试的时候别这么整！	Google interviewers beg you! Don't do this during the interview!	想跳槽？你可能被谷歌安全部盯上了！	Want to jump ship? You may be targeted by Google Security!	0.545233369
骗子项目！留学花费几十万，回国不能通过学历认证！	Liar Project! Studying abroad costs hundreds of thousands, and this cannot pass the academic certification when returning to China!	留学生毕业回国后多久能赚够一百万？	How long does it take for international students to earn one million after graduation?	0.523976445
“我留了两年学，就再也没法回国了”	"I stayed for two years to study abroad, and I will never be able to return to China"	“这一届留学生，活得可太容易了。”	"Life is too easy for this class of international students."	0.523976445
炒币炒股不如纯打工？！湾区人实现财富自由的方式竟是……	Investing in stocks or currency cannot do better than just working for others? ! The way people in the Bay Area achieve financial freedom is actually...	那些早早迈入富人圈的湾区人，都做对了同一件事...	Those Bay Area people who became rich early did the same thing right...	0.527217567
接到大使馆电话，摁了7后，父母积蓄全没了...	After receiving the call from the embassy and pressing 7, my parents' savings were all gone...	留学4年，最怕给客服打电话，哭...	After studying abroad for 4 years, I am most afraid of calling customer service. Crying...	0.490797371
想去大厂？转码太难！不如换个方向？	Want to go to a big company? Transiting to a new career in software engineering is too hard! Why not change direction?	中年转行成功！拿到大厂职位！如何平衡本职工作、面试和家庭？	Successful middle-aged career change! Get a job in a big company! How do you balance work, interviews, and family?	0.486133635

688x376mm (76 x 76 DPI)

Author Manuscript

Author Manuscript

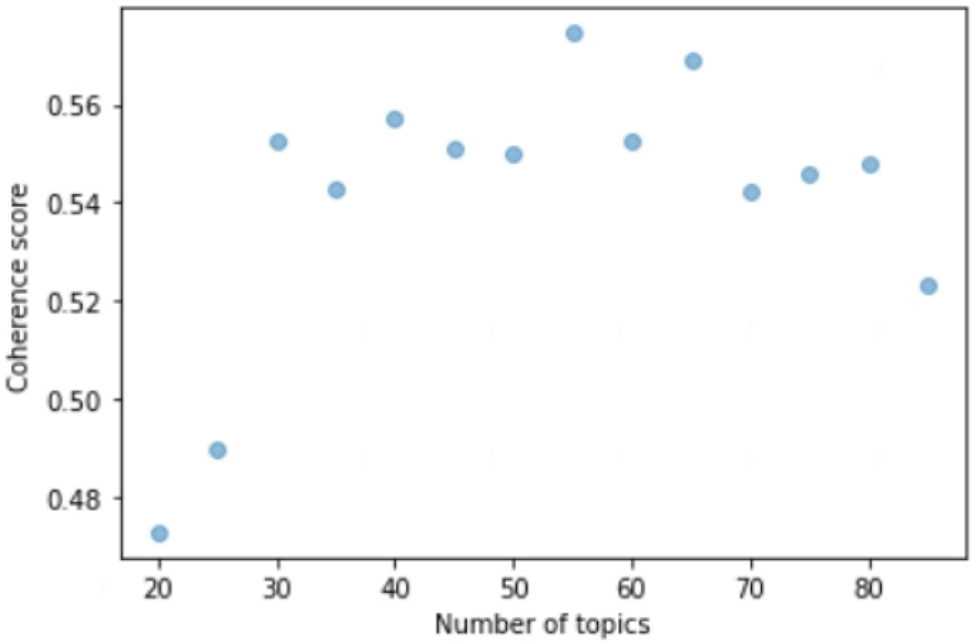


356x437mm (76 x 76 DPI)

Originally with question mark	Originally with question mark	Originally without question mark	Originally without question mark	Semantic similarity
史上最伟大个人交易！？8000美元变57亿美元，只花了400天！	Greatest personal deal ever! ? $8,000 became $5.7 billion in just 400 days!	收益超330%！靠这个股票，巴菲特5年狂赚1200亿美元！	The income is over 330%! With this stock, Buffett earned $120 billion in 5 years!	0.452171296
取消SAT成绩要求？！加州大学录取现在看什么？	Dropping SAT score requirements? ! What does the University of California admission look like now?	加州，不许给差生打低分！为了教育公平！	California: Don't give poor students a low score! For education fairness!	0.522243023
全美首个或执行死刑的中国留学生，犯了什么罪？	What crime did the first Chinese student in the United States execute or execute?	江歌案宣判！美国留学生情杀案，也如此离奇！	The Jiang Ge case is sentenced! The murder case of American students is also so bizarre!	0.515769899
在美国这个职业，居然没有work life balance吗？	Is there no work life balance in this profession in the United States?	在美国工作没"大包"，我该如何开心的活着	How can I live happily if I don't have a "big bag" to work in the United States	0.483206153
美国将拿发200万张绿卡！排期还能交钱插队？！	The United States will issue 2 million green cards! Can I still pay to cut the queue? !	我酸了！特赦16万移民直接拿绿卡！	I'm sour! Amnesty for 160,000 immigrants to get green cards directly!	0.516782403
交了5个月房租，发现房东竟是我老婆！？	After paying 5 months rent, I found out that the landlord is my wife! ?	湾区condo抄底经历，1个月内买到自住房！	Bay Area condo bargain hunting experience, buy a house within 1 month!	0.481140703
什么副业既优雅，又有"钱"景？	What sideline is both elegant and "money"?	推荐一个很适合女性的副业	Recommend a side business that is very suitable for women	0.468031824
实习生重写大厂代码是什么体验？	What is the experience of the intern rewriting the code of the big factory?	分享analytics跳槽到大厂的经历和感想	Share the experience and impressions of analytics jumping to Dachang	0.458224714
为什么跳槽，肯定要比升上去的拿得多？	Why jump jobs, must get more than the promotion?	今年2次跳槽，工资从$4.5万涨到$13万!	Twice this year, the salary has increased from $45,000 to $130,000!	0.469226003
被面试官用十招坑死面试？反击攻略来啦！	Was interviewed by the interviewer with ten tricks? Counterattack strategy is here!	加拿大找工求职请避开这些坑！	Please avoid these pits when looking for a job in Canada!	0.439164698

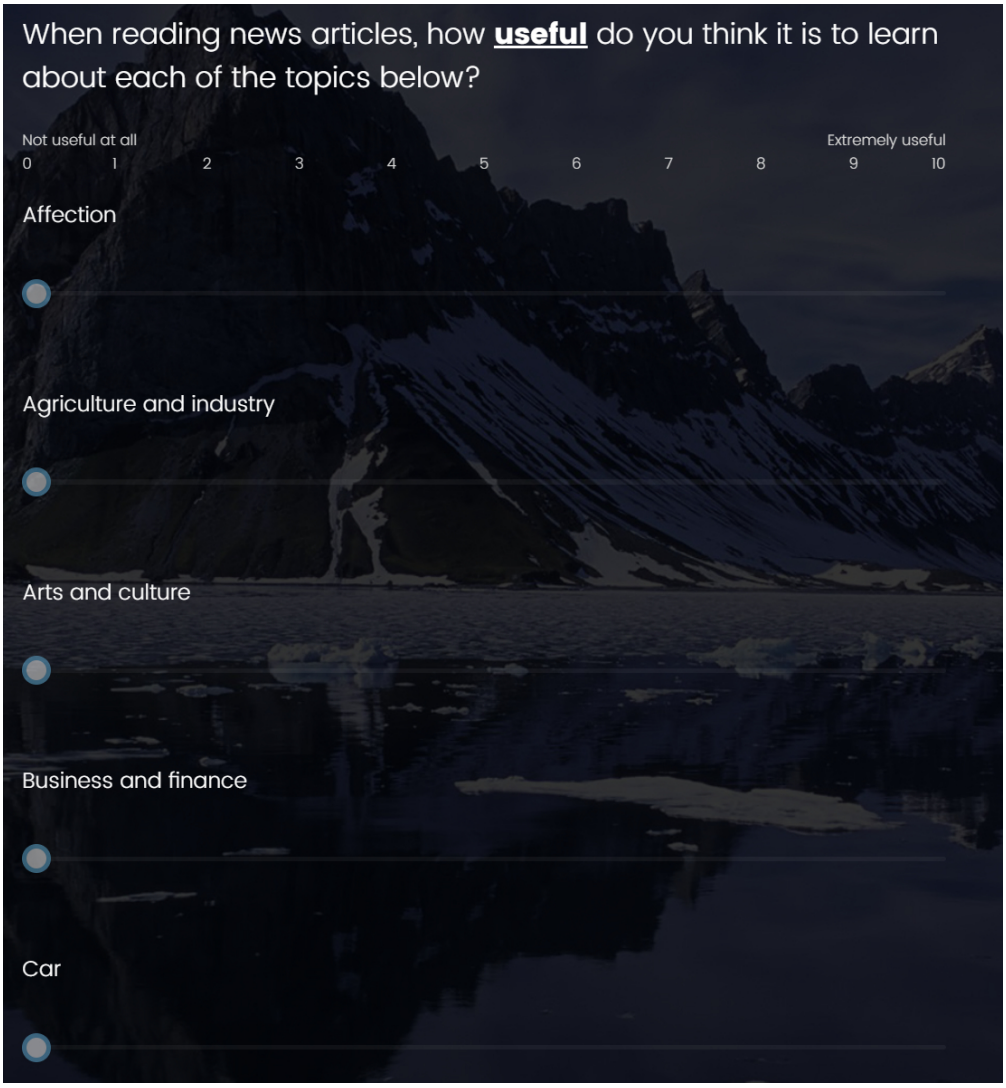
802x270mm (76 x 76 DPI)

Author Manuscript



234x154mm (76 x 76 DPI)

Author Manuscript



369x398mm (76 x 76 DPI)

Author Manuscript



309x351mm (76 x 76 DPI)

Author Manuscript