LINGUISTIC DRIVERS OF CONTENT CONSUMPTION

ABSTRACT

Marketers, media outlets, and creators alike all create content (e.g., articles, posts, and white papers) with the hope of attracting and retaining customers. But not all content generates sustained attention. Some barely get read, while other content keeps consumers engaged. Why? A multi-method investigation explores this question. We combine controlled experiments with natural language processing of 600,000 reading sessions from over 35,000 pieces of content to examine how language shapes content consumption. Results demonstrate that linguistic features associated with processing ease (e.g., concrete or familiar words) and emotion both play an important role. Rather than simply being driven by valence, though, the effects of emotional language are driven by the degree to which different discrete emotions evoke arousal and uncertainty. Consistent with this, anxious, exciting, and hopeful language encourages continued consumption while sad language discourages it. Experimental evidence underscores emotional language's causal impact and demonstrates the mediating role of uncertainty and arousal. The findings shed light on psychological drivers of content consumption, illustrate how content creators can generate more impactful content, and, as shown in a stylized simulation, have important societal implications for content recommendation algorithms.

KEYWORDS: content marketing, natural language processing, online content, digital marketing

Consuming content has become an integral part of everyday life. Consumers read the news, browse social media, and peruse various types of stories and information. People spend almost 50 hours a week online (Koetsier 2020), and much of that is reading (Cole et al. 2017).

As a result, content marketing has become big business. Rather than interrupting something people want to consume (e.g., a TV show) with an ad, companies and organizations are using content to attract and retain customers. Posts, articles, and white papers are only some of the avenues being used, and $400B a year is spent in the space (McCoy 2019).

But while some content gets viewed, less generates sustained attention (Lorenz-Spreen et al. 2019; Firth et al. 2019). Some articles are consumed in their entirety, while others barely get glanced at. Why? How does the way content is written shape *how much* of it is consumed (e.g., 25%, 50%, or 75%)? If an organization writes a post about climate change, financial literacy, or any other important issue, what language should they use to encourage people to keep reading?

Attempts to answer this question have been hampered by data availability. There's no record of how far people make it through physical content (e.g., physical newspaper articles), and while online metrics like views indicate what gets clicked on, they provide little insight into how much of a piece of content is actually consumed.

To address this gap, we analyze a unique dataset of over 600,000 reading sessions from 35,000 pieces of content to explore how language shapes content consumption. We examine whether consumers are more likely to continue consuming articles whose text should be easier to process or contains more emotional language. Results suggest that both aspects shape continued consumption, but in nuanced ways. Follow-up experiments demonstrate emotional language's causal impact, and the processes behind these effects (i.e., evoking uncertainty and arousal), and a stylized simulation highlights the implications for algorithms trained to sustain attention.

Our findings make four main contributions. First, they deepen understanding around what drives content consumption. While some research has examined what attracts attention (e.g., a catchy headline, Lai and Farbrot 2014; Kim et al. 2016), or drives word of mouth (see Berger 2014 for a review) there has been less attention to how language impacts sustained attention, or what encourages people to *keep* consuming something once they've started. We demonstrate the important role of emotional language and processing ease and show how different linguistic features shape content consumption through these key aspects.

Second, the findings help improve content design. From marketers to publishers, content creators don't just want clicks, they want readers to consume content. We show how language can help. Simple shifts can encourage sustained attention. Further, while one might think sustained attention is all about the topic (e.g., celebrity gossip beats articles about financial literacy), we show that writing *style* can make up for some of these differences. Even for "less engaging" topics, writing in particular ways can increase sustained attention. In addition, linguistic features are particularly valuable because they are actionable. Rather than focusing on how different individuals react differently, we examine ways of writing that, on average, encourage content consumption across individuals.

Third, the results highlight that sustained attention is distinct from other types of engagement. While grabbing attention or generating shares are important, we demonstrate that content characteristics that encourage these types of engagement are not always the same as what encourages sustained attention. While more certain language can increase likes and shares (Pezzuti et al. 2021), for example, when it comes to sustained attention, certain emotions are actually *detrimental*. Similarly, while some have argued that content which requires more cognitive processing should increase clicks (Kanuri et. al 2018), when it comes to sustained attention, content that requires more processing has the *opposite* effect. These differences highlight that sustained attention is a distinct and different aspect of engagement and suggest that findings from one type of engagement may not necessarily carry over to others. Consequently, when developing content, managers should think carefully about which outcomes they care most about, and design with that in mind.

Fourth, the findings have important societal implications. Online content consumption has become a critical social issue. Disinformation and hate speech have been linked to negative outcomes for individuals and society at large (e.g., Choi, Gaysynsky, and Cappella, 2020). Our results highlight the critical role language may play in this process. In a context where consumers engage more with anxious and angry content, algorithms trained to maximize sustained attention (Newberry 2022; Wall Street Journal 2021), will increase the amount of anxious and angry content users are recommended. As our simulation shows, this shapes the overall tenor of online content toward more anxious and angry emotions, an outcome which is not necessarily in the best interest of consumer welfare. This is a consequence of algorithmic design which is currently overlooked, but highlighted in this research

Types of Content Engagement

Before examining drivers of sustained attention, it is important to distinguish it from other types of engagement. Engagement with textual content can encompass a broad set of behaviors (Table 1). Advertisers, for example, want to attract attention, and research has examined how ad design impacts clickthrough (Lohtia, et al. 2003). Similarly, work on social media has examined how headlines attract attention, affecting how many views content receives

(Lai and Farbrot 2014; Kim et al. 2016). Related work has explored how times of day and message feature affect how many clicks links receive (Kanuri et. al, 2018).

TABLE 1: TYPES OF ENGAGEMENT

	Definition	**Examples of Research**	
Attracting Attention	Viewing or clicking on something	Lohtia, Donthu and Hershberger (2003)	How banner ad content and design impact advertising clickthrough
		Kim, Mantrach, Jaimes, and Oh (2016)	How words in headlines impacts clickthrough
		Lai and Farbrot (2014)	How questions in headlines increase clicks
		Kanuri, Chen, and Sridhar (2018)	How time of day and content type relates to link clicks
Likes, Comments, and Shares	Clicking the like button, writing a response, or sharing something with others (e.g., word of mouth)	Peters et al (2013)	How motives, content, and network structure drives likes and comments
		Pezzuti, Leonhardt, and Warren (2021)	How certain language affects likes, comments, and shares
		Berger and Milkman (2012)	How emotion shapes the sharing of news
		Chen (2017)	Social acceptance motives shape whether people share content
		Berger (2014)	Review of what drives word of mouth and sharing
Sustained Attention	Continued viewing or reading	The current research	How the language of content impacts whether people keep reading it

But while a particular stimulus may be more or less likely to *attract* attention, sustained attention refers to whether that stimulus *holds* attention. A catchy headline might lead someone to click on a link, for example, but once they click and start reading, how much of it do they actually consume? In some cases, people may stop reading after the first paragraph, while in others they consume some, most, or all of the article before moving on to something else. Thus while attracting attention refers to whether something garners attention, sustaining attention is about whether it *retains* the attracted attention, or how much of a piece of content is consumed.

Beyond attention, engagement can also encompass likes, comments, or shares. A burgeoning stream of research has begun to examine what drives these outcomes. Some work has examined why some content receives more likes or comments, for example, highlighting the role of consumer motives and network structure (Peters et al. 2013) or demonstrating that brand messages that include words expressing certainty are liked and commented on more (Pezzuti, Leonhardt, and Warren 2021). Other work has focused on word of mouth and what drives social sharing (see Berger 2014 for a review). Social acceptance motives, for example, shape what people share (Chen 2017) as does interpersonal closeness (Dubois, Bonezzi, and De Angelis 2016). Similarly, products that are more publicly visible, or triggered more by the environment, are talked about more (Berger and Schwartz 2011).

But while this work has provided important insights, note that sustained attention is distinct from liking, commenting, and sharing as these responses usually occur *after* content has been consumed. Someone may like, comment on, or share an article, for example, but these actions are usually *consequences* of content consumption, not causes. One does not have to read the full article to share it, but sharing is a distinct action that usually occurs after someone has attended to the article.

Overall, then, there are different types of engagement. Certain things may attract attention (i.e., cause consumers to click on or view something), or encourage liking, commenting, or sharing (i.e., passing content on to someone else), but these are distinct constructs from sustained attention (i.e., whether consumers keep reading or viewing).

Sustained Attention

Not surprisingly, content has more impact if people actually consume it, so sustained attention has a number of important consequences. The more of a piece of content consumers read, the more knowledge they gain (Ward, Zheng, and Broniarczyk 2022; also see Hidi 1990, Kintsch 1980). Content consumption affects which issues get attention, brands get word of mouth, and products get purchased (e.g., Schweidel and Moe 2016). Consequently, sustained attention also has clear implications for organizations and brands. Holding attention can deepen brand relationships and encourage action.[1] Similarly, for media companies and social media platforms, sustaining users' engagement with a piece of content or the platform increases the number of opportunities to display ads and generate ad revenue (see Yan et al., 2022).

A key question, then, is what causes sustained attention. But while decades of research in education, psychology, and communication have examined things like how readers go from eye fixations to comprehension (Just and Carpenter 1980), or generate inferences about what a text is about (Graesser, Singer, and Trabasso 1994), there has been less work on sustained attention. Some research suggests that people will be more engaged in topics they find personally interesting (see Hidi 1990 for a review), for example, but this says less about textual features that might encourage interest across people, regardless of topic.

Further, while a few papers have begun to explore textual features associated with interest (e.g., Schraw and Lehman 2001), this work has relied on forced reading paradigms. Rather than examining how much people read, laboratory participants are told to read an entire document and, at the end, report how interesting they found it (e.g., Schraw et al. 1995). Such situations provide tight experimental control but are unrealistic. In real-life, consumers don't

---

[1] Results of a pilot demonstrate the benefits of sustained attention. Participants were randomly assigned to consume either one-third, two-thirds, or all of a piece of content about a vacation destination and asked how interested they would be in visiting, their willingness to pay to get there, and memory questions about the content. Sustained attention improved responses. Consuming more content increased interest in visiting ($F(2, 88) = 10.48$, $p = .003$), willingness to pay to get there ($F(2, 88) = 5.09$, $p = .008$), and knowledge about the destination (i.e., % of questions answered correctly, $F(2, 88) = 58.97$, $p < .001$).

have to read an entire article and can opt-out at any point. Further, it's not clear that the same things that shape reported interest after forced reading would apply when people can choose to stop reading whenever they want. While reasoned arguments might be rated as interesting if people are forced to slog through them, if they are too boring in real-life, readers may just move on to something else. Consequently, what drives sustained attention in actual content consumption remains unclear.

Drivers of Content Consumption

We suggest that processing ease and emotional language are two key drivers of content consumption.

Processing Ease

In our context, processing ease describes how much cognitive effort text requires to process. Some research suggests there are benefits to increased processing. In the context of clicking on social media links, for example, Kanuri et. al (2018) argue that links that require more cognitive processing should receive more engagement (i.e., clicks). Citing prior work, they suggest that "online content that requires higher cognitive processing…receives increased engagement because of its increased level of cognitive involvement (Stieglitz and Dang-Xuan 2013)" (p. 93). Based on this, one might imagine that textual features that increase cognitive processing should encourage continued consumption

In contrast, we suggest the opposite. While there are certainly situations when more processing may be useful, in the context of continued consumption, we suggest that textual features that make content *easier* to process should have positive effects. Just as objects are more likely to keep moving when there is less friction, the easier something is to do, the more likely people are to continuing doing it (Kool, McGuire, Rosen, and Botvinick 2010; Zipf 1949). Processing ease can also generate positive affect (Alter and Oppenheimer 2019), which could encourage continued consumption.

This should play out at both the word and sentence level. The word "car," for example, is relatively straightforward to process. It is short, familiar, and concrete, all of which should make it easy to read, parse, and comprehend (Kincaid, et al. 1975; Winkielman & Cacioppo 2001; Connell and Lynott 2012). In contrast, a word like "Austalopithecus" is more difficult to process. It's longer, less familiar, and even someone who knows what it means would likely say that meaning is more abstract. Consequently, not only does it take longer to read, and comprehend, but it requires more effort as well. This should decrease the likelihood of continued content consumption.

The same logic can be applied to larger chunks of text like sentences. Longer sentences generally require more effort to read, as do sentences that are more syntactically complex. "The river near this city empties into the bay" and "The river that stopped flooding empties into the bay" are the same length, for example, but the former is syntactically simpler because the embedded prepositional phrase requires fewer syntactic nodes (Ferreira 1991). Differences in syntactic complexity, in turn, can shape how easy sentences are to read and understand (Pitler and Nenkova 2008; Schwarm and Ostendorf 2005).

In sum, we suggest that textual features that require more cognitive processing should reduce continued consumption. Further, the fact that we expect different effects for sustained attention than prior work found for clicks underscores our suggestion that sustained attention is distinct from other types of engagement (i.e., clicks) and involves different managerial insights.

Emotional Language

Beyond how easy text is to process, we suggest that emotional language will also shape content consumption. Emotions can increase attention or flag that something is important and deserves further consideration (Easterbrook 1959, Vuilleumier 2005). In particular, we suggest that how emotional language shapes sustained attention will depend on the link between specific emotions, uncertainty, and arousal.

The role of uncertainty. Emotions vary in the degree to which they are characterized by uncertainty, or not knowing or being sure of something (Smith and Ellsworth 1985, Lerner and Keltner 2000). While some emotions (e.g., anger or pride) tend to be characterized by certainty and lead people to feel certain about their environment, others (e.g., anxiety, hope, or surprise) tend to be characterized by uncertainty and uncertainty reduction (Raghunathan and Pham, 1999, Lerner and Keltner 2001; MacInnis and de Mello, 2005). When angry, for example, people tend to be certain about what they are angry about, but when anxious, people tend to be uncertain about what will occur.

Some work suggests that certain language can increase engagement. In the context of likes, comments, and shares, for example, Pezzuti et al. (2021) find that brands whose social media posts that use more certain words (e.g., always or everything) are liked, commented on, and shared more. Based on these findings, one might imagine that emotions associated with certainty (e.g., anger) might encourage sustained attention.

In contrast, we suggest the opposite. While expressing certainty may increase likes or shares because it makes brands seem more powerful (Pezzuti et al. 2021), in the context of sustained attention, other aspects of certainty may be more impactful. In particular, uncertainty can increase attention and processing as people try to resolve what will happen (Tiedens and Linton 2001, Weary and Jacobson 1997). If someone feels anxious about whether it's going to rain, the accompanying uncertainty might lead them to consume information that resolves that uncertainty (e.g., checking the weather).

Consequently, we suggest that language related to uncertain emotions should encourage content consumption. Compared to language related to certain emotions (e.g., anger), language related to uncertain emotions (e.g., anxiety) should encourage sustained attention. Further, the fact that we expect different effects for sustained attention than prior work found for likes and shares underscores the notion that sustained attention is distinct from other types of engagement and involves different managerial insights.

The role of arousal. In addition to uncertainty, emotions also vary in their level of arousal or activation (Teeny et al. 2020, Yin et al. 2017). Arousal is a state of being physiologically alert, awake, and attentive (Heilman 1997). While some emotions (e.g., anger, excitement, and anxiety) are characterized by high arousal, others (e.g., sadness or contentment) are low arousal.

While some work (Kanuri et al. 2018) finds no association between high arousal negative emotion and link clicks, in the context of content consumption, we suggest that language related to high arousal emotions should encourage continued consumption. A great deal of research finds that emotionally arousing stimuli attract attention (see Mather 2007 for a review). Work using brain imaging and skin conductance, for example, finds that threat-related stimuli are particularly attention grabbing (Ohman and Mineka 2001), in part because of the arousal involved. Arousal may also lead to an increased state of vigilance (Pham 2004) which should encourage sustained attention. Consequently, compared to low-arousal emotions (e.g., sadness or contentment), we suggest that language related to high arousal emotions (e.g., anger, anxiety, or excitement) should encourage content consumption.

Taken together, rather than suggesting that *any* emotional language should increase content consumption, we make a more nuanced prediction. Whether emotional language increases or decreases sustained attention will depend on the degree to which it is linked to specific emotions that evoke (1) uncertainty and (2) arousal. While anxious (high arousal and uncertainty) language should increase sustained attention, for example, sad (low arousal) language should decrease it.[2] Anger is high arousal and low uncertainty, so it's effect should lie somewhere in between, and should depend on the confluence of those two aspects in a particular situation. Similar effects should hold for positive emotions (e.g., excitement vs. contentment).

Empirical Tests

A multi-method approach, employing both field data and controlled experiments, tests these predictions. First, natural language processing of over 600,000 page read events from over 35,000 pieces of content examines whether consumers are more likely to continue consuming texts which should be easier to process, or contain more emotional language, and whether different specific emotions (e.g., anxiety and sadness) have different effects (Study 1).

---

[2] Certainty is not a core dimension of sadness and the relationship is more variable (Ellsworth and Smith 1988; Tiedens and Linton 2001). Sometimes people feel sad and uncertain, but other times they feel sad and certain.

Second, follow up experiments (Studies 2 and 3) test specific emotions' causal impact and the underlying process. They examine how emotional language shapes content consumption, and whether, as hypothesized, these effects are driven by uncertainty and arousal.

Study 1: Natural Language Processing of Over 35,000 Pieces of Content

Our first study uses natural language processing to analyze the consumption of over 35,000 pieces of content. We examine whether people are more likely to continue consuming text that (1) is easier to process (e.g., because it uses shorter sentences or more familiar language) and (2) uses more emotional language and whether (3) different specific emotions (e.g., anger vs. sadness) have different effects.

Data

A major content intelligence company provided a representative random sample of page-consumption events (i.e., instances where an article was opened by a user) over a two-week period from nine sites that together cover a wide range of topics (i.e., global and local news, business, sports, technology, fashion, and lifestyle content). [3] While confidentiality prohibits us from revealing the exact publishers, outlets like CNBC, the Wall Street Journal, Minnesota Star Tribune, and Jezebel provide some sense of the types of content examined. At the time of data collection, the sites used had fixed layouts (i.e., content was laid out the same way across

[3] Given our interest in textual features, we focus on articles rather than other content types (e.g., videos). We focus on page-sessions that involve some interaction. Readers may sometimes leave right after opening an article or open another browser tab and do something else. To avoid such cases where users are unlikely to have read much, if any, of the article, we rely on the company's definition, which involves focusing only on users who had at least two interactions with the page (e.g., mouse scrolls or clicks). Users are not tracked over time or across sites. Consequently, we can compare behavior across users for a given article but not repeat user behavior across articles.

articles), did not have ads within the text, and were non-responsive (i.e., regardless of whether an article was read on phone, desktop, or other device, content was not reformatted based on viewport size and line breaks were the same). This characteristic of the data simplified the analysis by eliminating the need to accommodate differences in a reader's screen size or page breaks. The final dataset involved 649,129 page-consumption events from 35,448 articles. See Table A1 and Figure A1 for summary statistics.

Processing Ease

We measure processing ease in four ways. First, to test whether shorter words and sentences (which should generally be easier to process) are linked to continued consumption, we use the standard Flesch-Kincaid approach to measure text *readability* (Kincaid, et al. 1975). This uses word and sentence length to capture how easy content is to read (i.e., what grade level it is appropriate for). Second, to test whether syntactic simplicity encourages continued consumption, we measure *parse tree height*, a standard approach in the computer science and linguistics literatures (Pitler and Nenkova 2008).[4] Finally, to test whether *familiar* or *concrete* language (which should be easier to process, Connell and Lynott 2012; Winkielman & Cacioppo 2001), is linked to continued consumption, we use ratings from Paetzold and Specia (2016).[5] To facilitate interpretation, we multiply Flesch-Kincaid and parse tree height scores by -1 such that higher scores on readability and syntactic simplicity indicate things should be easier to process.

---

[4] Parse tree height counts the number of steps to get from the top node to the bottom most node. While both "The cat on the hot tin roof meowed at my parent's house" and "The cat on the hot tin roof at my parent's house meowed" are similar sentences and involve the same words, the second one has a taller parse tree (i.e., 8 edges vs. 6 edges tall).

[5] Paetzold and Specia (2016) used bootstrapping with word embeddings to extend the MRC Psycholinguistic Database (Coltheart, 1981) from around 9,000 words to over 85,000. Participants in the original studies rated words based on either how familiar they were (e.g., Gilhooly and Logie 1980; 1 = never seen, heard, or used and 7 = seen, heard, or used every day) or how concrete they were (e.g., Spreen and Schulz 1966; 1 = least concrete, 7 = most concrete). Words referring to objects, materials, or people received high concreteness ratings. Words referring to abstract concepts that could not be experienced by the senses (e.g., the word "facts") received low concreteness ratings. Using Brysbaert et al. (2014)'s concreteness measure finds the same results.

We predict that readability, syntactic simplicity, linguistic familiarity, and linguistic concreteness should all make content easier to process and thus encourage continued consumption.

Emotional Language

While we focus on individual specific emotions, one could wonder whether all emotional language encourages content consumption, and whether positive or negative content is more likely to encourage reading. Consequently, before exploring specific emotions, we conduct two simpler specifications: one that groups all emotional language together and one that separates positive and negative language. To do this, we measure emotional valence using Linguistic Inquiry and Word Count (LIWC, Pennebaker et al. 2015).

We also measure language linked to specific emotions. Specific negative emotions are easier to distinguish from one another than positive emotions (Keltner and Lerner 2010), and more sophisticated natural language processing tools exist to extract negative emotions from text, so we focus on specific negative emotions (we explore specific positive emotions in more detail in Study 3). There are reliable, well-validated tools to measure anger, anxiety, and sadness, so we focus on those. While LIWC's positive and negative emotion categories have been validated, the specific emotion categories have less empirical support. Consequently, rather than simply measuring the presence or absence of individual words, we rely on a more continuous approach developed by Mohammad and Bravo-Marquez (2017) which allows for more accurate variation (LIWC provides similar results, Table A5).[6] They gave raters four pieces of content and asked which had the highest and lowest intensity of different specific emotions (i.e., anger, anxiety, and sadness). Then, they performed machine learning on this training set

[6] Pilot testing found that Mohammad and Bravo-Marquez (2017)'s approach was more predictive of manual coding.

and used a variety of features (e.g., word embeddings and affect lexicons) to extrapolate responses to a broader set of content.

We use this approach to measure the amount of angry, anxious, and sad language in each sentence, averaging across sentences to get a score for each paragraph. (See Table A2 for summary statistics, Table A3 for correlations between variables, and Table A4 for example paragraphs that score highly on each dimension).[7]

Dependent Variable and Analysis Strategy

While one could model consumption of an article as a function of textual features of the entire article, behavior can't be influenced by content that hasn't been read, so it's important to focus on content that appears before a user stopped. Further, article level analysis ignores within article variation (e.g., some paragraphs use lots of anxious language while others do not).

Consequently, we take a more fine-grained approach, examining whether a user continues consuming content (i.e., moves to the next paragraph) based on the text of each paragraph. To capture paragraph-to-paragraph consumption, we measure how far down the page a user scrolls (see Appendix for more detailed description and example). This is determined using code embedded on the publishers' sites and executed on the user's browser when an article is loaded (i.e., for each consumption event). The code records the pixel position a user scrolls to on the page, defined as the top position that is visible on the user's screen, starting from 0 and increasing up to the length of a given article. We then map pixel length to position within the article using a custom CSS selector, unique for each site.[8] The conversion from pixel length, and

---

[7] We normalize the processing ease and emotional language variables for use in our analyses.

[8] We downloaded each article and extracted the pixel location of the top of each paragraph to know whether the user read past this point during a given page-consumption event. The content is not scaled and text is not re-flowed based

selection of sites with non-responsive layouts, ensures that recorded content consumption is independent of site layout and consistent across devices, screen resolutions, and window sizes.[9] Table A7 contains descriptive statistics for the page consumption events.

We conceptualize each page-consumption event, *I*, as a sequence where at the end of each paragraph the user either continues consuming content or stops. We denote the action made after paragraph *j* of session *i* as Y_{ij}, in which $Y_{ij}=1$ if the user continues to the next paragraph and $Y_{ij}=0$ if they do not. We assume that the probability of continuing past paragraph *j* in session *i* is a function of the paragraph-level content variables and control variables. Formally, we estimate individual i's probability of continuing past paragraph j as:

$$Y_{ij} \sim Bernoulli(p_{ij}) \text{ where}$$

$$logit(p_{ij}) = \beta_0 + \sum_k \beta_k \cdot X_{ijk} + \sum_c \gamma_c \cdot Z_{ijc}$$

where X_{ijk} denotes the k^{th} independent variable that characterizes the content of paragraph *j* in event *i* and Z_{ijc} denotes the c^{th} control variable.

This analysis is consistent with a discrete-time survival analysis. Effectively, we model the "time" (measured in paragraphs) until someone stops reading, recognizing that some individuals may read the entire article. For an individual who is observed to stop consuming after seeing paragraph T, this likelihood is given by decisions to continue consuming after paragraphs 1,2,…,T-1 and to stop consuming after paragraph T. If the predictor variables were assumed to

on screen resolution or browser window size. For example, the same page would be 1000 pixels on both a low-resolution mobile device and high-resolution screen.

[9] We do not mean to suggest that visitors are reading every single word. Indeed, people sometimes skim rather than reading articles in depth. That said, if visitors were not consuming any words, that should make it harder to find effects of textual features on scrolling because they were not exposed to enough words for the words to shape behavior. Thus, unless skimming or deeper scrolling is somehow driven by some alternative feature that is also correlated with the textual features we examine, this measure provides a conservative test of our hypotheses. The less people are reading the weaker any relationships between textual features and reading should be. While one could argue that familiarity with the article or subject could encourage skimming, note that out article level topic controls, and controlling for familiar language at the article level, help address this possibility.

be constant, our analysis would be equivalent to assuming that the time until an individual stops reading the paragraphs of an article follows a geometric distribution. Similar methods have been used to model binge viewing (Schweidel and Moe 2016) and clickstream behavior (Sismeiro and Bucklin 2004, Moe 2006).

Control Variables

While we are interested in the effects of emotional language and linguistic features linked to processing ease, external factors (e.g., device or publisher), aspects of the article (e.g., topical content), reader progression (e.g., how much they have read so far), and other aspects of the paragraph (e.g., topical content) may also shape consumption, so we control for them to cast doubt on alternative explanations and test the robustness of the effects.

Device. Different types of people may use different devices (e.g., mobile vs. desktop), use different devices at different times, and device itself may impact behavior (Ransbotham et al. 2018). To control for these possibilities, we use dummy variables to control for whether users read an article on mobile, desktop, or tablet (0.5% of page-reads are from an unknown device).

Publisher. Different publishers may attract different types of people, attract them when they have more or less time, or publish articles that encourage different content consumption. To control for these possibilities, we use publisher-specific fixed effects.

Temporal controls. Time of day (i.e., early morning, morning, afternoon, evening and overnight) or day of week may also impact consumption, so we include fixed effects for each.

Article Content. Beyond these external factors, aspects of the article itself may also shape consumption. Before reading an article, consumers are usually exposed to its headline or a brief summary. This content may impact who starts to read the article, when they read it, and their initial interest level. We control for this in two ways.

First, we control for topical focus. Articles about certain topics may attract different types of people, attract them when they have more time, or impact content consumption in other ways. The websites in our dataset include the article's headline or a related summary in the article's URL. As an illustration, an article in *The Wall Street Journal* entitled "IKEA Sales Boosted by China" is associated with the URL https://www.wsj.com/articles/ikea-sales-boosted-by-china-1410246927. We extract the article summary (in this case, "ikea sales boosted by china 1410246927") for each article and perform topic modeling on the resulting words, allowing each summary be represented as a proportion of different topics. We use latent Dirichlet allocation (e.g., Blei et al. 2003), a common topic modeling framework (e.g., Berger and Packard 2018, Tirunillai and Tellis 2014)[10] and include the posterior topic probabilities as control variables to represent the relative prevalence of each topic at the article level.

Second, we measure the presence of the same emotions measured in the body of the article, and control for those as well.

Article popularity. Article popularity may also pick up variation in what attracts and holds attention that is not reflected in other aspects of the content. We account for this by including the logarithm of the number of unique readers of the article.

Reading Progress. We also control for aspects within the article. More time spent consuming an article may make readers more or less likely to continue reading, so we control for how much content someone has consumed so far using content length in words up to that point.

[10] We increase the number of topics considered until validation perplexity increases. While not all topics readily lend themselves to easy interpretation, note that we are interested in controlling for and accommodating variation in topics across articles, not the effects of the specific topics themselves. Nonetheless, a review of the topics reveals articles about technology (words such as facebook, data and smartphone), social activities (words including friends, coffee and weekend), and sports (words such as nfl, series, game and score).

Content consumption may also depend on where someone is in the article, so we also control for percentage consumed so far. We use both linear and quadratic terms to allow for non-linearities.

Paragraph topics. We also control for other features of the focal paragraph. Just as the article's headline may impact reading, so too might the topical content of the focal paragraph. Following the procedure for article summaries, we control for topics across all paragraphs (see Appendix Table A6 for distribution of topics across articles and sample words), which includes things like government (i.e., words like state and govern$), sports (i.e., game and team) and personal technology (i.e. app and google). We include the posterior topic probabilities topics as control variables to represent the relative prevalence of each topic in each paragraph.

Other linguistic features. We also control for other linguistics features of the focal paragraph (i.e., baskets of words linked to other social or psychological constructs from LIWC dictionaries such as cognitive processes, sociality, perception, motivation, time, and formality).

Paragraph length. Given limited attention spans, the longer a paragraph is, the less likely people may be to read the next one. Consequently, we control for paragraph length using the number of words.

User Heterogeneity. Finally, to accommodate differences across users in content consumption behavior, we adopt a modeling approach that allows for unobserved heterogeneity in users' baseline levels of sustained attention (see Model Specification below).

Model Specification

We estimate several models. Model 1 groups all emotional language together (i.e., emotionality), Model 2 separates positive and negative language, and Model 3 explores specific emotions. In addition to controlling for the factor mentioned above, each examines effects of processing ease (i.e., readability, syntactic simplicity, familiarity, and concreteness).

To accommodate user heterogeneity, we model user-specific fixed effects governed by a discrete mixture model (Vilcassim and Jain 1991). Specifically, we allow β_0 to be characterized by N finite supports with probability masses $[q_1\ q_2\ \ldots\ q_N]$. As we increase the number of supports from N=1 to N=3, we the substantive findings are unchanged. For N=3, the smallest probability mass is less than 1%. Based on this, we present our empirical findings associated with N=2.[11]

To accommodate heterogeneity across articles, we control for article-level differences through the content of the headlines/summaries (i.e., topics and emotional content) and differences in views across articles.[12]

Results

Processing ease. Features which should make content easier to process were linked to continued consumption (Table 2). Consumers were more likely to continue consuming content that was more readable, syntactically simple, and used familiar or concrete language (all coefficients are positive and ps < .001).

For example, the paragraph "It's the great retirement debate: How much can retirees spend each year without running out of money before they run out of breath?" was high on familiar language (0.69) and has a higher completion rate (90%) than the paragraph "In a letter to

[11] Results are similar using a hierarchical Bayesian logistic regression that allows for unobserved heterogeneity across both individuals and articles (e.g., Ansari et al. 2000, Moe and Schweidel 2012). Given computational resources required to estimate such a model on our full dataset, we randomly sample 5% of users (i.e. 32,024 individuals) or 350,505 paragraph consumption events from 32,941 reading events spanning 10,284 articles.
[12] We also assessed the accumulation of emotions and topics over the course of the article through the use of a stock model. Not allowing for the accumulation of these factors best fits the data. See Web Appendix for detail.

a friend, the manager of a Florida urology practice worried in 2010 that her company would attract federal scrutiny for its frequent use of an expensive bladder-cancer test." (familiarity = -0.95, completion rate = 81%).

Table 2: Content Characteristics and Continued Consumption

		Emotionality	Valence	Specific Emotions
		(1)	(2)	(3)
Emotion	Emotionality	0.0072***		
		0.0005		
	Positive Emotion		0.023***	0.025***
			0.0015	0.0015
	Negative Emotion		0.0060***	
			0.0005	
	Anger			0.011***
				0.0013
	Anxiety			0.042***
				0.0013
	Sadness			-0.036***
				0.0013
Processing Ease	Readability	0.065***	0.066***	0.070***
		0.0014	0.0014	0.0014
	Syntactic Simplicity	0.051***	0.050***	0.051***
		0.0013	0.0013	0.0013
	Familiarity	0.028***	0.028***	0.029***
		0.0014	0.0014	0.0014
	Concreteness	0.048***	0.051***	0.046***
		0.0015	0.0015	0.0015
Controls	Device dummies	Yes	Yes	Yes
	Publisher dummies	Yes	Yes	Yes
	Temporal controls	Yes	Yes	Yes
	Paragraph word count	Yes	Yes	Yes
	Reading progress	Yes	Yes	Yes
	Additional LIWC features	Yes	Yes	Yes
	Headline content	Yes	Yes	Yes
	Article popularity	Yes	Yes	Yes
	User heterogeneity	Yes	Yes	Yes
	Observations	6,994,372	6,994,372	6,994,372
	LL	-1969437	-1969377	-1969152
	AIC	3939088	3938971	3938524

Emotional language. Model 1 suggests that people are more likely to continue consuming content that uses more emotional language ($\beta = 0.0072$, $p < .001$), and Model 2 suggests that this is true for both positive ($\beta_{\text{Positive Language}} = 0.023$, $p < .001$) and negative ($\beta_{\text{Negative Language}} = 0.0060$, $p < .001$) emotions.

Exploring specific emotions (Models 3), however, suggests a more nuanced picture. While people were more likely to continue consuming content that used more anxious ($\beta = 0.042$, $p < .001$) or angry ($\beta = 0.011$, $p < .001$) language, they were *less* likely to continue consuming content that used sad language ($\beta = -0.036$, $p < .001$).

For example, this paragraph "In the winter of 1947, an American tourist arrived in New York City on a bus from Mexico, feeling feverish and stiff. He checked into a hotel and did some sightseeing before his condition worsened. A red rash now covered his body. He went to a local hospital, which monitored his vital signs and transferred him to a contagious disease facility, where he was incorrectly diagnosed with a mild drug reaction. He died a few days later of smallpox." was high on sad language (1.23) and was less likely to be completed (91%) than this paragraph "This winter, the Santa Cruz Museum of Art & History will feature an exhibit of works relating to the ocean, with paintings and sculptures by established artists alongside works by local residents. According to a call for submissions, that includes not just watercolors of Pacific sunsets, but that awesome GoPro footage you took while surfing and your two-year-old's drawing of the beach that's been on the fridge for five months." (below average sad language = -2.71, 94% completion rate). See Table A4 for more examples.

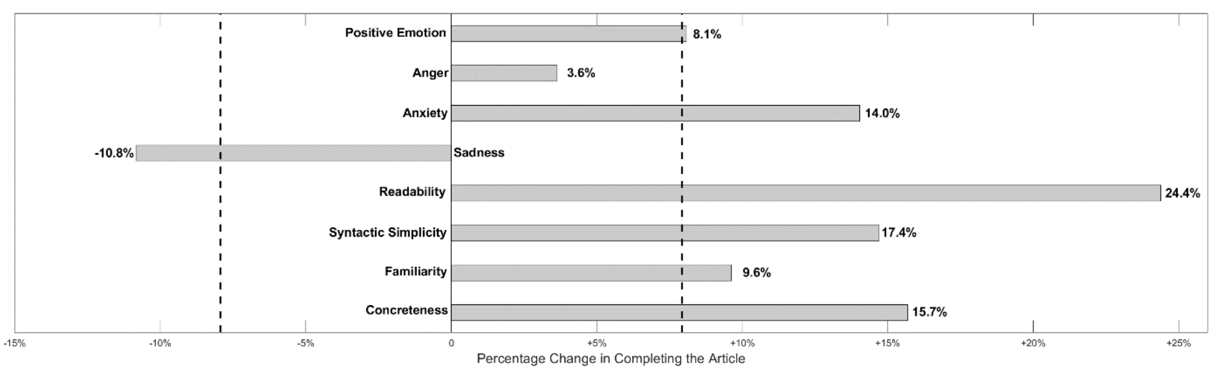
The fact that different specific negative emotions have different effects suggests that emotion's effect on content consumption is driven by more than valence alone. Further, the effects are consistent with the hypothesized role of arousal and uncertainty. Anxiety (high arousal and uncertainty) was associated with continued consumption, anger (high arousal but certain) was associated with continued consumption but to a lesser degree, and sadness (low arousal) seemed to discourage continued consumption.

Discussion

Natural language processing of over 600,000 reading events from over 35,000 pieces of content is consistent with our theorizing. First, results suggest that processing ease encourages content consumption. Consumers were more likely to continue consuming content that was more readable, syntactically simple, and used familiar or concrete language.

To illustrate the impact of linguistic features, Figure 1 shows the effect of a standard deviation increase in each linguistic feature on the likelihood an article is completed. Consumers are around 25% more likely to finish articles that use shorter word and sentences (i.e., has higher readability), and around 14% more likely to finish articles that use more anxious language.

Figure 1: How Textual Features Change Content Consumption



Note: Bars represent effect of a standard deviation increase in each textual feature on continued consumption, relative to "baseline" article (i.e., Wall Street Journal article consumed on desktop with average posterior probabilities for each topic, and average emotion and control measures). Dashed vertical lines reflect average impact of the article's topical content. To derive this, we calculate the absolute value of each topic's impact by increasing the posterior topic probability by one standard deviation, reducing the topic probabilities of the remaining 24 topics by 1/24 of this amount, and average across the 25 topics to arrive at the average impact of topics.

By comparing the effects of each textual feature to the average topic effect (dotted lines), the figure also illustrates the relative impact of topic versus writing style. Results suggest that language use has a similar, and in several case even larger, effect on content consumption than the average topic. This suggests that while topic (e.g., sports or science) certainly shapes content consumption, how that topic is discussed (i.e., the *language* used) also plays an important role.

Second, emotional language plays an important role, but the effects are more nuanced than just emotionality or valence alone. Instead they are consistent with our suggestions regarding arousal and uncertainty. While people were more likely to continue consuming content that used anxious or angry language, they were *less* likely to continue consuming content that used sad language.

We focused on anger, anxiety, and sadness because there are reliable tools to extract these emotions, but exploratory analyses on surprise and disgust underscore the hypothesized role of uncertainty and arousal (see Appendix). Surprise, which is low certainty and high arousal, seems to encourage continued content consumption, while disgust, which is associated with certainty, seems to discourage continued consumption.

Further analyses are also consistent with the suggested role of uncertainty and arousal. Uncertainty was measured using LIWC's certainty (i.e., words like "always" and "never") and tentative language (e.g., "maybe" and "perhaps") measures and arousal was measured using ratings from Warriner et al. (2013). Results are consistent with the notion that uncertainty and arousal encourage continued consumption. People were more likely to continue consuming content that used less certain ($\beta = -0.001$, $p = .07$) or more tentative language ($\beta = 0.006$, $p < .001$) or more language associated with arousal ($\beta = 0.084$, $p < .001$).

Including controls and robustness checks helps rule out alternative explanations, but to further test our theorizing, and more directly test language's causal impact, we turn to experiments. This also allows us to better test the hypothesized process underlying the effects.

Given the difficulty of orthogonally manipulating all features studied, and the fact that the emotional language effects are more complex (i.e., different emotions seem to have different effects), the experiments focus on emotional language.

Study 2: Experimentally Manipulating Emotion

Study 2 has two main goals. First, while the results of Study 1 are consistent with the notion that emotional language influences content consumption, one could wonder whether the effects are truly causal. To provide more direct evidence, we use an experiment. We take an article portion, manipulate whether it uses sad, anxious, or angry language, and measure the resulting impact on whether people choose to continue reading. Consistent with the field data, we predict that different specific emotions will have different effects: compared to when an article uses sad language, anxious or angry language will encourage continued reading.

Second, Study 2 more directly tests the underlying process behind these effects. Using mediational analyses, we examine whether specific emotions impact content consumption because they evoke uncertainty and arousal in readers.

Method

Participants (N = 278, recruited through Mturk) completed an experiment as part of a larger group of studies. They were told experimenters were interested in perceptions of content and that they would read the beginning of a news article and respond to some questions.

First, we took the part of an article and manipulated language associated with specific emotions. All participants received a similar article about the stock market, but we manipulated whether the article used sad, anxious, or angry language. The [sad, anxious, angry] version said:

> "Recent stock market performance has made investors really [sad, anxious, angry]. Most markets are down over 25%, the average American has lost tens of thousands of dollars, and many have [helplessly, nervously, furiously] watched as their retirement savings have dwindled. "I'm [heartbroken, worried, frustrated], said one New Jersey man, "my family is really [devastated, confused, bitter]."

The article portion was the same length and structure across conditions, and the only difference was the specific emotion words used. These words in particular were chosen based on their membership in specific emotion dictionaries (i.e., helpless, heartbroken, and devastated are all part of the sadness dictionary). A manipulation check (Appendix) shows that the manipulation worked as intended, reliably activating the intended emotions and not others. A second pretest further demonstrated that there was no difference between conditions on a variety of other measures (i.e., personal relevance, concreteness, extremity, or evoking hope).

After reading the article portion, participants completed the dependent variable: choosing whether they wanted to continue reading the rest of the article or switch to something else.

Finally, we measured the hypothesized underlying processes: uncertainty and arousal. Uncertainty was measured using three items adapted from Faraji-Rad and Pham (2017): "how does this article make you feel?" 7-point scale anchored by unsure/sure, hesitant/determined, and don't feel confident/feel confident ($\alpha = .93$, reverse-scored and averaged to an uncertainty index). Arousal was measured using three items adapted from Berger (2011): "how does this article make you feel" 7-point scale anchored by very low energy/very high energy, very passive/very active, very mellow/very fired up (αs = .92, averaged to arousal index).

Results

Sustained attention. As predicted, compared to using sad language (M = 52%), anxious (80%) or angry language (71%) led people to choose to continue reading the article (Anxiety χ^2 (1) = 18.99, $p < .001$, Anger χ^2 (1) = 7.66 $p = .006$).

Underlying processes. As predicted, a one-way ANOVA found that emotional language influenced how much uncertainty the article evoked ($F(2, 336) = 4.73$, $p = .009$). As expected, using anxious rather than angry language boosted feelings of uncertainty (M = 4.50 vs. 3.80; $t(336) = 3.01$, $p = .003$). As noted, sad language can introduce either certainty or uncertainty, so we did not have a specific prediction. That said, in this particular context, sad language reduced uncertainty compared to anxious language (M = 4.12; t(336) = 2.05, $p = .041$), but directionally increased uncertainty compared to angry language (t(336) = 1.10, $p = .27$)

Emotional language also affected how much arousal the article evoked ($F(2, 336) = 7.62$, $p = .001$). As expected, compared to using sad language (M = 3.88), using anxious (M = 4.34; $t(336) = 2.23$, $p = .026$) or angry language (M = 4.71; $t(336) = 3.87$, $p < .001$) boosted arousal.

Mediation. More importantly, a series of bias-corrected simultaneous mediation models (Hayes 2017) found that, as predicted, the combination of uncertainty and arousal drove the effects of specific emotions on continued reading. Compared to using sad language, anxious language encouraged continued reading because it boosted both uncertainty (*ab* = .15, 95% CI .01 to .38) and arousal (*ab* = .23, 95% CI .04 to .54). Similarly, using angry rather than sad language encouraged continued reading because it boosted arousal (*ab* = .44, 95% CI .20 to .82, uncertainty's indirect effect did not reach significance = -.06, 95% CI -.26 to .03). Finally, also consistent with our theorizing, the difference between anxious and angry language was driven by uncertainty (*ab* = .18, 95% CI .009 to .51, arousal's indirect effect did not reach significance *ab* = -.24, 95% CI -.58 to .04).

Discussion

Results of Study 2 bolster the findings of the field data in a controlled setting and provide evidence for the hypothesized underlying process. First, manipulating emotional language influenced people's choice to continue reading, and the effect depended on the specific emotion considered. Consistent with Study 1, compared to sad language, angry or anxious language led people to want to read more. The fact that these effects held even when taking the same article, and simply manipulating a few words, underscores the causal impact of specific emotion.

Second, as predicted, these effects were driven by how much uncertainty and arousal different specific emotions evoked. Using angry or anxious (rather than sad) language evoked arousal and using anxious (rather than angry) language evoked uncertainty. Both uncertainty and arousal encouraged continued reading, and, in combination, drove specific emotion's effects.[13]

Study 3: Positive Emotions

Studies 1 and 2 focused on negative emotions because there are reliable tools for measuring such features in language, but to explore whether positive emotions show similar effects, and whether they are driven by the same underlying processes, Study 3 manipulates

---

[13] Arousal alone is insufficient to explain the results. While both anger and anxiety are high arousal, anxious language encouraged reading more than angry language. Anxiety's greater uncertainty encouraged reading.

positive emotions. While contentment is a relatively low-arousal positive emotion, excitement, and to some degree hope, are relatively high arousal (Berger and Milkman, 2012; Cavanaugh, Bettman, and Luce 2015; Kim, Park, and Schwarz 2010; MacInnis and de Mello 2005). Similarly, while contentment is a relatively certain positive emotion, hope is characterized by more uncertainty (i.e., people are hoping but not sure something with happen) and excitement may be as well (Cavanaugh, et al. 2015; Kim, et al., 2010; MacInnis and de Mello 2005). Consequently, we predict that compared to contentment, both excitement and hope should increase interest in continuing to read more because they increase excitement and arousal.[14]

Further, to provide additional control, and rule out alternative explanations, Study 3 manipulates emotions outside the content itself (i.e., through a seemingly unrelated task, see Cavanaugh et al. 2015, Berger 2011, for similar approaches). By keeping the focal reading content identical across conditions, and manipulating emotion incidentally, we ensure that any observed difference between conditions is driven by emotion rather than some other factor. Participants wrote about a time they felt either content, excited, or hopeful, and then, as part of an ostensibly unrelated experiment, read part of a neutral article and reported their interest in continuing to read more. If specific emotion impacts reading, as we suggest, then the emotion induced in the first task should spill over into the second. Even though everyone read the same article, the incidental emotion manipulation should impact reading, with the effects driven by how these discrete emotions impact uncertainty and arousal.

[14] Note that the exact effects of excitement and hope will depend on the degree to which each evokes arousal and uncertainty, so we do not have a prediction about any difference between them.

Method

Two hundred and forty-eight Prolific participants completed two ostensibly unrelated studies and were randomly assigned to one of three between subject conditions (content vs. excited vs. hopeful). The study was pre-registered at https://aspredicted.org/FNP_91B.

First, we manipulated specific emotions. Adapting manipulations used in prior work (Griskevicius, Shiota, and Neufeld, 2010), participants described something that made them feel the focal emotion. In the excited condition, for example, participants were asked to recall a time when they felt excited, take a minute to remember it vividly, and then write a paragraph about it in as much detail as they could. The prompt was similar in the content and hopeful conditions condition except that participants were encouraged to think about a time they felt those emotions instead. Manipulation checks indicate the manipulations worked as intended.[15]

Second, we measured the hypothesized processes using the uncertainty and arousal measures from Study 2

Third, we measured the dependent variable. After completing the first "study" participants moved on to the "second." This involved reading content from a news article (about wireless charging, adapted from the *New York Times*) and answering some follow up questions. After reading the first paragraph, participants were asked "how likely they would be to read more of the article" (1 = not at all, 7 = very much) if they came across it while browsing online.

[15] As precited, participants in the excited condition reported feeling more excitement than participants in the other conditions ($t(246) = 5.09$, $p < .001$), participants in the content condition reported feeling more content than participants the other conditions ($t(246) = 1.83$, $p = .068$), and participants in the hopeful condition reported feeling more hopeful than participants in the other conditions ($t(246) = 4.54$, $p < .001$).

Results

Specific Emotions. As predicted, a one-way ANOVA found that emotion shaped continued reading ($F(2, 246) = 4.84$, $p = .009$). Consistent with our theorizing, compared to contentment ($M = 3.57$), both excitement ($M = 4.44$, $t(246) = 3.03$, $p = .003$) and hope ($M = 4.15$, $t(246) = 2.088$, $p = .038$) made people more interested in reading more of the article.

Underlying processes. A one-way ANOVA found similar effects on uncertainty ($F(2, 246) = 7.05$, $p = .001$) and arousal ($F(2, 246) = 6.97$, $p = .001$). Compared to contentment ($M = 2.31$), both excitement ($M = 3.04$, $t(246) = 3.64$, $p < .001$) and hope ($M = 2.80$, $t(246) = 2.54$, $p = .012$) made people feel more uncertain. Similarly, compared to contentment ($M = 4.34$), both excitement ($M = 5.04$, $t(246) = 3.41$, $p = .001$) and hope ($M = 4.93$, $t(246) = 2.94$, $p = .004$) made people feel greater arousal.

Mediation. Finally, bias-corrected simultaneous mediation models (Hayes 2017) found that uncertainty and arousal drove the effects on discrete emotions on reading. First, compared to contentment, excitement encouraged continued reading because it boosted both uncertainty ($ab = .20$, 95% CI .04 to .45) and arousal ($ab = .27$, 95% CI .10 to .54). Similarly, compared to contentment, hope encouraged continued reading because it boosted both uncertainty ($ab = .17$, 95% CI .03 to .41) and arousal ($ab = .20$, 95% CI .05 to .45).

Discussion

Study 3 underscores the findings of the first two studies. First, the results demonstrate that the previously observed effects extended to discrete positive emotions. Compared to contentment, excitement or hope increased people interested in continuing to read further. Second, consistent with our theorizing, these effects were driven by arousal and uncertainty.

Third, manipulating emotion incidentally casts doubt on the possibility that something else, beyond discrete emotions, is driving the effect. Fourth, by studying positive emotions, we de-couple the effects of a specific emotion (e.g., anxiety) with the effects of arousing emotions more generally. The fact that we find similar effects for both positive and negative discrete emotions suggests that the effects are not about any one individual emotion per se, but the larger dimensions different discrete emotions are associated with.

Implications for Algorithmic Design

Taken together, the results of the field data and experiments have implications for the algorithms digital platforms use to recommend content. Many publishers and social media platforms employ algorithms to select what content to present, as suggested stories or part of a user's feed. Traditionally, these algorithms have relied on manual categorization of stories into broad categories or identifying and promote topics matching user preferences. More recently, however, algorithms identify and promote content that increases user engagement, including user interactions with the content and sustained attention. In addition to considering likes and shares, for example, Facebook ranks content according to "time spent" when determining what content to include in a user's feed (Newberry 2022).

Our findings demonstrate that content features beyond just topic influence engagement, specifically sustained attention. In particular, they suggest that algorithms trained to increase engagement may ultimately privilege types of content (e.g., anxiety-producing) that have negative consequences for individual users, and the overall tone of content in the community.

To demonstrate how our findings interact with algorithm design, we simulated 10,000 users whose engagement behaviors are determined by a baseline tendency, topic preference, and given it had the largest effect in the field data, response to anxiety-producing content (this could be replaced with any content feature or multiple content features). In this paper, engagement represents a reader's sustained attention to a piece of text-based content and measured by reading depth. But it can be broadly construed and measured by a variety of behavioral metrics more generally. TikTok uses "dwell time," for example, or the amount of time a user's screen remains on a specific piece of content, as a metric to maximize (Wall Street Journal 2021). In our simulation, we simulate engagement behavior and do not specify the specific measure of engagement as that can be defined to accommodate a variety of platforms and data signals.

In this stylized simulation, we specify user *i*'s probability to engage with a specific piece of content as follows:

$$p_{ij} = logit(b_{0i} + b_{1i}TOPIC_j + b_{2i}ANXIETY_j)$$

where b_{0i} represents user *i*'s baseline tendency to engage with any topic, b_{1i} represents *i*'s tendency to engage with specific topics, and b_{2i} represents *i*'s response to anxiety-producing content.

We start with a simulated environment where each of these three parameters are normally distributed across users with mean 0 and standard deviation of 1. For simplicity, we assume TOPIC is 1 or 0, representing the presence or absence of a topic, and likewise for ANXIETY. These simple and stylized assumptions help establish a baseline behavior where consumer preferences for topic and emotional content are normally distributed across the population. Our simulation considers the presence or absence of a single topic or content feature (in this case, anxiety), but can be generalized to accommodate multiple topics or different emotions. User *i*'s engagement with a piece of content is 0 or 1 according to a Bernoulli draw of p_{ij}.

The content recommendation algorithm suggests new content to each user based on their previous consumption. We consider two types of algorithms, one that is trained to serve content that matches the user's topical interests and one that is trained to maximize content engagement, driven by both content topic and emotionality. In both cases, the simulation begins with randomly selected content (i.e., both TOPIC and ANXIETY are randomly chosen) being served. Depending on whether or not a user attends to the served content, subsequent recommendations evolve, depending on the algorithm being used and the user's engagement decision from the previous round. For the topic-focused algorithm, content that matches the topic served previously would be served again if the user attends to it; otherwise, the other topic would be served. For the algorithm that also considers emotionality, both the topic and emotionality would be matched if the user attended to it in the last round.

Not surprisingly, results (see Table 4) indicate that incorporating linguistic features into the algorithm improved engagement. After 10 rounds of recommendations, the topic-only algorithm achieved engagement among 52% of users (compared to 50% in the initial round when the content was randomly chosen). When the algorithm also considered whether the content evokes anxiety, engagement increased to 54%. This suggests that if users are driven by both topic preferences and emotions, as our findings suggest, then algorithms designed to consider more than just topical interests will increase engagement.

That said, given our empirical findings show that some emotional content (e.g., anxiety inducing) deepens engagement more than others, we also simulate an environment where b_{2i} is distributed N(1,1) to mirror that tendency. In other words, we create a simulated world where consumers are drawn to anxiety inducing content like we observe in our empirical analyses. In this scenario, engagement increases to 64% after 10 rounds, but the recommended content has a mean anxiety score of 0.63. This suggests that the content being served becomes notably more anxiety-inducing after rounds of algorithmic recommendations (whereas there was no difference in the previous scenarios).

Table 4: Simulation results

		Engagement Rate		Average Topic		Average Emotionality	
Scenario	Algorithm	Round 1	Round 10	Round 1	Round 10	Round 1	Round 10
Topic and emotions preferences randomly distributed N(0,1)	Topic focused	50%	52%	0.50	0.52	0.50	0.50
	Engagement focused	50%	54%	0.50	0.52	0.50	0.52
Topic preference distributed N(0,1) and emotionality preference distributed N(1,1)	Engagement focused	50%	64%	0.50	0.51	0.50	0.63

While highly stylized, the simulation highlights our finding's implications for content recommendation algorithms. If user engagement is influenced by specific emotions (e.g., anxiety), as our findings show, then algorithms trained to increase engagement based on content features will systematically serve content that caters to those preferences, shaping the overall tone of the content in ways that may be negative for users, and society.

Consequently, firms that design and employ algorithms that serve content designed to increase engagement should do so with caution, as such algorithms have the potential to increase negativity in their community. Instead, these firms should consider increasing engagement as one goal among many. Other goals may include aligning with user preferences or increasing the diversity of recommendations. Further study into the impact of these various algorithmic goals can lead to improved customer experience.

General Discussion

Content consumption is an integral part of consumer behavior. People spend hours a day reading posts, articles, and other content. Consequently, optimizing content has become big business. Beyond traditional media outlets, all sorts of companies and organizations are turning to content marketing to engage with customers: designing posts, articles, and white papers to attract and retain audiences. But why does some content generate more sustained attention?

Our multi-method investigation examines how language shapes content consumption. Consistent with the notion that processing ease encourages continued consumption, consumers were more likely to continue consuming content that was more readable, syntactically simple, and used familiar or concrete language.

Emotional language also plays a role. But rather than any emotion increasing consumption, or the effects being driven by valence alone, the results demonstrate the effect of arousal and uncertainty. Whether emotional language encouraged or discouraged continued consumption depended on the degree to which it evoked these aspects. People were more likely to consume content that uses anxious, exciting, or hopeful language, for example, because it evoked uncertainty and arousal in readers.

Contributions

These findings make several contributions. First, on a theoretical level, they contribute to understanding content consumption. Decades of research have examined different aspects of reading and attention. But while this work has provided important insights into comprehension, memory, and other topics, there has been less attention what drives sustained attention. This paper demonstrates that linguistic features associated with processing ease (e.g., familiar or concrete words) seem to encourage continued consumption. Emotional language, particularly emotions that evoke uncertainty and arousal, has similar effects.

Second, on a practical level, the findings have clear takeaways for increasing engagement. Subtle shifts in how things are written should encourage continued reading. Replacing abstract words with more concrete ones (e.g., "product" with "phone," or describing a car in terms of its color) for example, and less familiar words with more familiar synonyms, should encourage engagement. The right emotional language should have similar effects. Rather than just relying on facts, for example, a non-profit focused on climate change might benefit from leveraging emotions that evoke uncertainty and arousal. Using exciting or hopeful language, for example, should encourage sustained attention.

That said, the benefits of content consumption should be balanced against other outcomes. Increasing anxiety in advertising might encourage sustained attention, for example, but hurt brand equity. As our simulation suggests, there are also important societal implications. Platforms can use measures of sustained attention to recommend future content, but this may lead readers to be served an endless stream of anxiety-producing information. Addressing this issue, though, is not straightforward. Hoping that readers will be drawn to useful, informative content simply because it has those characteristics is unlikely to solve the problem. By understanding why some content is more likely to be consumed, hopefully content creators can level the playing field and help useful information get more sustained attention.

Third, the results demonstrate that content consumption depends on more than just the topic alone. Organizations often lament that it is easier to get people to read about "frivolous" topics (e.g., celebrity gossip) than "weightier" ones (e.g., policy discussions and environmental appeals). But while topic certainly plays a role in driving sustained attention, our results demonstrate that they are not the only factor. Even controlling for what an article is about (i.e., its topics), how that topic was discussed (i.e., the language used) played an important role. This is good news for organizations trying to encourage people to read about less "engaging" topics. While the topic itself may not be the most engaging, writing about it in the right ways can deepen sustained attention. Writing style can compensate for topic.

Future Research

As with any preliminary effort, more remains to be done. Research might examine other textual features. How does the similarity between an article and its outlet shape reading? For a sports outlet, for example, are people more likely to deeply read a typical article or one that is more atypical (e.g., about player's personal lives)? Similarity could also be examined within content. Any story, article or narrative (e.g., book or movie) can be broken down into chunks. How might the similarity between those chunks impact sustained attention? Content that flows between similar chunks should be easier to process and follow, which might deepen sustained attention. Alternatively, chunks being too similar might feel repetitive or like the plot is not advancing fast enough.

Future work could also examine where readers came from. While our data does not allow us to examine this, compared to readers that come to an outlet's homepage, for example, those that come in through social media may not read as much. Similarly, does reading one article impact reading behavior on a subsequent article? If people leave one article because it is not engaging enough, might that increase their impatience and decrease their likelihood of sustained attention in a subsequent article? One could also examine individual differences in the emotional state of readers, or factors that affect the mood of the general populace. Similarly, one could examine whether user preferences are stable throughout the course of reading an article or whether they change (e.g., emotion has a larger impact later rather than earlier in a piece) or whether there might be complex interactions within content over time (e.g., whether positive content followed by negative content might have differing effects then the opposite order).

It would also be interesting to consider how content consumption affects return visits and subscriptions. One would imagine that the more attention readers give an article, the more likely they will be to return to that content provider, thereby generating more advertising revenue and possibly choosing to subscribe. If most articles only hold people's attention for a couple paragraphs, they're less likely to keep coming back. This highlights the downsides of overly attention-grabbing headlines. Clickbait may be great for attracting views, but to maintain long term value (e.g., Du et al. 2021), deeper sustained attention may be needed. Further research into content's ability to sustain attention may also shed light on the effectiveness of embedded advertising (e.g., Schweidel and Moe 2016, Fossen and Schweidel 2019).

Future work might also examine the boundaries of the findings observed here. What drives sustained attention on a very technical website, for example, might be different from what drives sustained attention for the news or content on social media. Similarly, while we found that uncertainty increased sustained attention, in situations where it casts doubt on whether the continuing to read will lead the reader to find the information they are looking for, it may have the opposite effect.

Finally, while there are many cases where most consumers might respond the same way, future work might explore whether, when, and why different consumers might respond differently to the same content. t A news article might report on a Republican election victory with a positive tone, for example, but while Republicans might react positively, Democrats might react negatively. Similarly, while most investors would be sad or anxious about a tanking stock market, short sellers might react more positively. Future work could examine which types of content are more likely to have heterogeneous effects and why.

Conclusion

In conclusion, sustained attention has important implications for consumers, organizations, and society. It shapes what issues get attention and what brands consumers learn about and purchase. Consequently, sustained attention has an important impact on brands and societal discourse more generally. By better understanding why certain content attracts more sustained attention, hopefully we can improve outcomes for all these audiences.

References

Alter, Adam L. and Daniel M. Oppenheimer (2019), "Uniting the Tribes of Fluency to Form a Metacognitive Nation," *Personality and Social Psychology Review*, 13, 219–235.

Ansari, Asim, Skander Essegaier, and Rajeev Kohli (2000), "Internet Recommendation Systems," *Journal of Marketing Research,* 37 (3), 363–375.

Berger, Jonah (2011), "Arousal Increases Social Transmission of Information," *Psychological Science*, 22 (7), 891–93.

Berger, Jonah and Eric M. Schwartz (2011), "What drives immediate and ongoing word of mouth?" *Journal of marketing research*, 48 (5), 869–880.

Berger, Jonah, and Katherine Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research,* 49 (2), 192–205.

Berger, Jonah (2014), "Word of mouth and interpersonal communication: A review and directions for future research," *Journal of consumer psychology*, 24 (4), 586–607.

Berger, Jonah and Grant Packard (2018), "Are Atypical Things More Popular?" *Psychological Science*, 29 (7), 1178–1184.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Brysbaert, Marc, Amy Beth Warriner, and Victor Kuperman (2014), "Concreteness Ratings for 40 Thousand Generally Known English Word Lemmas," *Behavior Research Methods*, 46, 904–911.

Cavanaugh, Lisa A., James R. Bettman, and Mary Frances Luce (2015), "Feeling Love and Doing More for Distant Others: Specific Positive Emotions Differentially Affect Prosocial Consumption," *Journal of Marketing Research*, 52 (5), 657–673.

Chen, Zoey (2017), “Social Acceptance and Word of Mouth: How the Motive to Belong Leads to Divergent Word of Mouth with Strangers and Friends,” *Journal of Consumer Research*, 44 (3), 613–632.

Choi, Wen-Ying Sylvia, Anna Gaysynsky, and Joseph N. Cappella (2020), “Where We Go From Here: Health Misinformation on Social Media,“ *American Journal of Public Health* 110, 273–275.

Cole, Jeffrey I., Michael Suman, Pheobe Schramm, and Liuning Zhou (2017), “Surveying the Digital Future 2017,” *Center for the Digital Future at USC Annenberg*.

Coltheart, Max (1981), “The MRC Psycholinguistic Database,” *The Quarterly Journal of Experimental Psychology*, 33 (4), 497–505.

Connell, Louise and Dermot Lynott (2012), “Strength of perceptual experience predicts word processing performance better than concreteness or imageability,” *Cognition*, 125 (3), 452–465.

Du, Rex Yuxing, Oded Netzer, David A. Schweidel, and Debanjan Mitra (2021), “Capturing Marketing Information to Fuel Growth,” *Journal of Marketing*, 85 (1), 163–183.

Dubois, David, Andrea Bonezzi and Matteo De Angelis (2016), “Sharing with Friends versus Strangers: How Interpersonal Closeness Influences Word-of-Mouth Valence,” *Journal of Marketing Research*, 53 (5), 712–727.

Easterbrook, James A. (1959), “The Effect of Emotion on Cue Utilization and the Organization of Behavior,” *Psychological Review*, 66 (3), 183.

Ellsworth, Phoebe C. and Craig A. Smith (1988), “From appraisal to emotion: Differences among unpleasant feelings,” *Motivation and emotion*, 12 (3), 271–302.

Faraji-Rad, Ali and Michael Tuan Pham (2017), “Uncertainty Increases the Reliance on Affect in Decisions,” *Journal of Consumer Research*, 44 (1), 1–21.

Ferreira, Fernanda (1991), “Effects of Length and Syntactic Complexity on initiation Times for Prepared Utterances,” *Journal of Memory and Language,* 30, 210–233.

Firth, Joseph, John Torous, Brendon Stubbs, Josh A. Firth, Genevieve Z. Steiner, Lee Smith, Mario Alvarez-Jimenez, John Gleeson, Davy Vancampfort, Christopher J. Armitage, and Jerome Sarris (2019), “The Online Brain: How the Internet May be Changing Our Cognition,” *World Psychiatry*, 18 (2), 119–129.

Fossen, Beth L. and David Schweidel (2019), “Social TV, Advertising, and Sales: Are Social Shows Good for Advertisers?” *Marketing Science*, 38 (2), 274–295.

Graesser, Arthur C., Murray Singer, and Tom Trabasso (1994), “Constructing Inferences During Narrative Text Comprehension,” *Psychological Review*, 101 (3), 371.

Gilhooly, Ken J. and Robert H. Logie (1980), “Age-of-Acquisition, Imagery, Concreteness, Familiarity, and Ambiguity Measures for 1,944 Words,” *Behavior Research Methods and Instrumentation*, 12, 395–427.

Griskevicius, Vladas, Michelle N. Shiota, and Samantha L. Neufeld (2010), “Influence of different positive emotions on persuasion processing: A functional evolutionary approach,” *Emotion,* 10 (2), 190–206.

Hayes, Andrew F. (2017), “Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach,” *Guilford Publications.*

Heilman, Kenneth M. (1997), “The Neurobiology of Emotional Experience,” *The Neuropsychiatry of Limbic and Subcortical Disorders*, 133–142.

Just, Marcel A. and Patricia A. Carpenter (1980), “A Theory of Reading: From Eye Fixations to

Comprehension," *Psychological Review*, 87 (4), 329.

Hidi, Suzanne (1990), "Interest and Its Contribution as a Mental Resource for Learning," *Review of Educational Research*, 60 (4), 549–571.

Kanuri, Vamsi, Yixing Chen, and Shrihari Sridhar (2018), "Scheduling Content on Social Media: Theory, Evidence, and Application," *Journal of Marketing*, 86, 89-108.

Keltner, Dacher and Jennifer S. Lerner (2010), "Emotion," In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology,* 317–352.

Kim, Hyun Suk, Heather Forquer, Joseph Rusko, Robert C. Hornik, and Joseph N. Cappella (2016), "Selective Exposure to Health Information: The Role of Headline Features in the Choice of Health Newsletter Articles," *Media Psychology*, 19 (4), 614–637.

Kim, Hakkyun, Kiwan Park, and Norbert Schwarz (2010), "Will this trip really be exciting? The role of incidental emotions in product evaluation," *Journal of Consumer Research*, 36 (6), 983–991.

Kincaid, J. Peter, Robert P. Fishburne Jr, Richard L. Rogers, and Brad S. Chissom (1975), "Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel," *Institute for Simulation and Training*, 56.

Kintsch, Walter (1980), "Learning from Text, Levels of Comprehension, or: Why Anyone Would read a Story Anyway," *Poetics*, 9 (1-3), 87–98.

Koetsier, John (2020), "Why 2020 Is A Critical Global Tipping Point for Social Media," *Forbes*.

Kool, Wouter, Joseph T. McGuire, Zev B. Rosen, and Matthew M. Botvinick (2010), "Decision Making and the Avoidance of Cognitive Demand," *Journal of Experimental Psychology General*, 139 (4), 665–682.

Lai, Linda and Audun Farbrot (2014), "What Makes You Click? The Effect of Question Headlines on Readership in Computer-mediated Communication," *Social Influence*, 9 (4), 289–299.

Lerner, Jennifer S. and Dacher Keltner (2000), "Beyond Valence: Toward a Model of Emotion-Specific Influences on Judgement and Choice," *Cognition & Emotion*, 14 (4), 473–493.

Lerner, Jennifer S. and Dacher Keltner (2001), "Fear, Anger, and Risk," *Journal of Personality and Social Psychology*, 81 (1), 146.

Lohtia, Ritu, Naveen Donthu, and Edmund K. Hershberger (2003), "The Impact of Content and Design Elements on Banner Advertising Click-through Rates," *Journal of Advertising Research*, 43 (4), 410–418.

Lorenz-Spreen, Phillip, Bjarke Mørch Mønsted, Philipp Hövel, and Sune Lehmann (2019), "Accelerating Dynamics of Collective Attention," *Nature Communications*, 10 (1), 1–9.

MacInnis, Deborah J., and Gustavo E. De Mello (2005), "The Concept of Hope and its Relevance to Product Evaluation and Choice," *Journal of Marketing*, 69 (1),1 –14.

Mather, Mara (2007), "Emotional Arousal and Memory Binding: An Object-Based Framework," *Perspectives on Psychological Science*, 2 (1), 33–52.

McCoy, Julia (2019), "Why Content Marketing is Set to Be an Industry Worth $412.88 Billion by 2021," *MarTech Advisor*.

Moe, Wendy W. (2006), "A Field Experiment to Assess the Interruption Effect of Pop-Up Promotions," *Journal of Interactive Marketing*, 20 (1), 34–44.

Moe, Wendy W., and David A. Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science*, 31 (3), 372–386.

Mohammad, Saif M. and Felipe Bravo-Marquez (2017), "Emotion Intensities in Tweets," *Proceedings of the Sixth Joint Conference on Lexical and Computational Semantics*, 65–77, Vancouver (August 3–4).

Newberry, C. (2022). "How the Facebook algorithm works in 2022." *Hootsuite* (February 28), Retrieved from https://blog.hootsuite.com/facebook-algorithm/

Öhman, Arne and Susan Mineka (2001), "Fears, Phobias, and Preparedness: Toward an Evolved Module of Fear and Fear Learning," *Psychological Review*, 108 (3), 483.

Paetzold, Gustavo and Lucia Specia (2016), "Inferring Psycholinguistic Properties of Words," *Proceedings of the 2016 Conference of NACL: Human Language Technologies*, 435–440, San Diego (June 12-17).

Pennebaker, James W., Ryan L. Boyd, Kayla Jordan, and Kate Blackburn (2015b), "The Development and Psychometric Properties of LIWC2015," Austin, TX: University of Texas at Austin.

Pennebaker, James W., Roger J. Booth, Ryan L. Boyd, and Martha E. Francis (2015a), "Linguistic Inquiry and Word Count: LIWC2015," Austin, TX: Pennebaker Conglomerates.

Peters, Kay, Yubo Chen, Andreas M. Kaplan, Bjorn Ognibeni, and Koen Pauwels (2013), "Social Media Metrics—A Framework and Guidelines for Managing Social Media," *Journal of Interactive Marketing*, 27 (4), 281–298.

Pezzuti, Todd, James M. Leonhardt, and Caleb Warren (2021), "Certainty in Language Increases Consumer Engagement on Social Media," *Journal of Interactive Marketing*, 53, 32–46.

Pham, Michel Tuan (2004), "The Logic of Feeling," *Journal of Consumer Psychology*, 14 (4), 360–369.

Pitler, Emily and Ani Nenkova (2008), “Revisiting Readability: A Unified Framework for Predicting Text Quality,” *Proceedings of The Conference on Empirical Methods in Natural Language Processing*, 186–195, Honolulu (October 25-27).

Raghunathan, Rajagopal and Michel Tuan Pham (1999), “All Negative Moods are Not Equal; Motivational Influences of Anxiety and Sadness on Decision Making,” *Organizational Behavior and Human Decision Processes*, *79* (1), 56–77.

Ransbotham, Sam, Nicholas H. Lurie, and Hongju Liu (2018), “Creation and Consumption of Mobile Word of Mouth: How are Mobile Reviews Different?” *Marketing Science*, 38 (5), 773-792.

Schraw, Gregory and Stephen Lehman (2001), “Situational Interest: A Review of the Literature and Directions for Future Research,” *Educational Psychology Review*, 13 (1), 23–52.

Schraw, Gregory, Roger Bruning, and Carla Svoboda (1995), “Sources of Situational Interest,” *Journal of Reading Behavior*, 27 (1), 1–17.

Schwarm, Sarah E. and Mari Ostendorf (2005), “Reading Level Assessment Using Support Vector Machines and Statistical Language Models,” *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 523–530, Ann Arbor, (June 25-30).

Schweidel, David A. and Wendy W. Moe (2016), “Binge Watching and Advertising,” *Journal of Marketing*, 80 (5), 1–19.

Sismeiro, Catarina and Randolph E. Bucklin (2004), “Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach,” *Journal of Marketing Research*, 41, 306–323 (August).

Smith, Craig A. and Phoebe C. Ellsworth (1985), “Patterns of Cognitive Appraisal in Emotion,” *Journal of Personality and Social Psychology*, 48 (4), 813.

Spreen, Otfried and Rudolph W. Schulz (1966), “Parameters of Abstraction, Meaningfulness, and Pronunciability for 329 Nouns,” *Journal of Verbal Learning and Verbal Behavior*, 5 (5), 459–468.

Stieglitz, Stefan and Linh Dang-Xuan (2013), “Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior,” *Journal of Management Information Systems*, 29 (4), 217–248.

Teeny, Jacob, Xiaoyan Deng, and H. Rao Unnava (2020), “The “Buzz” Behind the Buzz Matters: Energetic and Tense Arousal as Separate Motivations for Word of Mouth,” *Journal of Consumer Psychology*, 30 (3), 429–446.

Tiedens, Larissa Z. and Susan Linton (2001), “Judgment Under Emotional Certainty and Uncertainty: The Effects of Specific Emotions on Information Processing,” *Journal of Personality and Social Psychology*, 81 (6), 973.

Tirunillai, Seshadri and Gerard J. Tellis (2014), “Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation,” *Journal of Marketing Research*, 51 (4), 463–479.

Vilcassim, Naufel J. and Dipak C. Jain (1991), “Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity,” *Journal of Marketing Research*, 28 (1), 29–41.

Vuilleumier, Patrik (2005), “How Brains Beware: Neural Mechanisms of Emotional Attention,” *Trends in Cognitive Sciences*, 9 (12), 585–594.

The Wall Street Journal (2021), “TikTok to Adjust Its Algorithm to Avoid Negative Reinforcement,” (December 16), https://www.wsj.com/articles/tiktok-to-adjust-its-algorithm-to-avoid-negative-reinforcement-11639661801

Ward, Adrian F., Frank Zheng, and Susan Broniarczyk (2022), "I Share, Therefore I Know? Sharing—Even Without Reading—Inflates Subjective Knowledge" *Working Paper.*

Warriner, Amy Beth, Victor Kuperman, and Marc Brysbaert (2013), "Norms of valence, arousal, and dominance for 13,915 English lemmas," *Behavior research methods*, 45 (4), 1191–1207.

Weary, Gifford and Jill A. Jacobson (1997), "Causal Uncertainty Beliefs and Diagnostic Information Seeking," *Journal of Personality and Social Psychology*, 73 (4), 839.

Winkielman, Piotr and John T. Cacioppo (2001), "Mind at Ease Puts a Smile on the Face: Psychophysiological Evidence that Processing Facilitation Elicits Positive Affect," *Journal of Personality and Social Psychology*, 81 (6), 989.

Yan, Shunyao, Klaus M. Miller, and Bernd Skiera (2022), "How Does the Adoption of Ad Blockers Affect News Consumption?" *Journal of Marketing Research*.

Yin, Dezhi, Samuel D. Bond, and Han Zhang (2017), "Keep Your Cool or Let it Out: Nonlinear Effects of Expressed Arousal on Perceptions of Consumer Reviews," *Journal of Marketing Research*, 54 (3), 447–463.

Zipf, George K. (1949), *Human behavior and the principle of least effort.* Addison-Wesley Press.