Article

Journal of Marketing
2023, Vol. 87(3) 428-450

DOI: 10.1177/00222429221127927
journals.sagepub.com/home/jmx

Does Topic Consistency Matter? A Study of Critic and User Reviews in the Movie Industry

Eunsoo Kim, MengQi (Annie) Ding, Xin (Shane) Wang, and Shijie Lu

Abstract

Online review platforms often present reviews from both critics and general users. In this research, the authors propose a measure called "topic consistency" to capture the degree of overlap between critic and user review content. High topic consistency suggests greater information recall due to repeated presentation of the same topics, which may increase the memorability of movie attributes and therefore positively affect movie demand. The authors measure the topic consistency between critic and user reviews using topic models and further study the financial consequences of this measure using data for movies released in the United States. Topic consistency is positively associated with subsequent box office revenue, suggesting a positive relationship between topic consistency and movie demand. Furthermore, the effect of topic consistency on demand is the greatest for movies with mediocre review ratings and when the review ratings from critics are close to those from users. Using lab experiments, the authors provide evidence of the causal link between topic consistency and consumers' willingness to watch a movie, and support for the potential mediation through the information recall of reviews. Movie producers and advertisers should consider highlighting or inducing a central theme for critics and users to discuss, as the more the review content of critics and users overlaps, the higher a movie's revenue.

Keywords

critic reviews, machine learning, movies, online word of mouth, topic consistency, topic model, user reviews

Online supplement https://doi.org/10.1177/00222429221127927

Consumers often read online reviews before visiting movie theaters. In 2018, 63% of U.S. adults indicated moderate to heavy reliance on online reviews before seeing a movie (Statista 2018). These reviews may come from different sources, as both critics and general users can leave reviews on popular movie review platforms such as Rotten Tomatoes and the Internet Movie Database (IMDb). Unsurprisingly, however, critics and general users do not always share similar opinions about a movie's quality or focus. Although media outlets speculate that this divide in opinions may impact consumers' moviegoing decisions (Chamary 2018; Youngs 2017), these conjectures have never been supported by rigorous empirical evidence. Accordingly, in this research, we examine review texts to quantify the division between the opinions of general users and those of critics and explore its financial consequences in the movie industry.

User and critic reviews differ in several aspects. Whereas user reviews tend to be more reflective of a typical consumer's tastes, critic reviews can provide additional information about the product grounded in expert judgment (Holbrook 1999). Moreover, critics, who are professional reviewers, often offer their opinions before the release of the product (Faber and O'Guinn 1984; Litwak 1986), whereas users typically provide their feedback after the product release (Liu 2006). Furthermore, critics and users hold different tastes and knowledge about the product (Hennig-Thurau and Houston 2019) and face different incentives (Burtch et al. 2018), which may lead to discrepancies in review content and review ratings. Thus, the different opinions conveyed by reviews from critics and users, and the interaction between them, can affect moviegoing decisions in a complex way. In the literature on online reviews, an underexplored facet of reviews is the review topic. Critics and users can write about different topics, such

Eunsoo Kim is Assistant Professor of Marketing, Nanyang Business School, Nanyang Technological University, Singapore (email: eunsoo@ntu.edu.sg). MengQi (Annie) Ding is Doctoral Candidate in Marketing, Ivey School of Business, Western University, Canada (email: ading.phd@ivey.ca). Xin (Shane) Wang (corresponding author) is Professor of Marketing, Pamplin College of Business, Virginia Tech, USA (email: xswang@vt.edu). Shijie Lu is Howard J. and Geraldine F. Korth Associate Professor of Marketing, Mendoza College of Business, University of Notre Dame, USA (email: slyu@nd.edu).

as movie plot, acting, and cinematography. Therefore, we ask important research questions regarding topic consistency: How do consumers evaluate a movie when users and critics mention common versus separate topics in their reviews? How does the effect of topic consistency between users and critics vary by the consistency of numeric ratings and the average movie ratings?

The goal of this research is to investigate the impact of the potential overlap between user and critic reviews on box office revenues. We capture this overlap with a measure that we call "topic consistency." We classify a movie as having high topic consistency if there is a large degree of overlap in the underlying topics conveyed by user and critic review texts. We then posit that such an overlap in review content (i.e., topic consistency) can affect movie demand, via the following mechanisms. High topic consistency reflects a collective voice from both types of reviewers, which makes the movie and the associated attributes discussed in reviews more salient than they would be if users and critics focused on different aspects of the movie. Because of the positive impact of salience on consumers' recall of information (Alba, Hutchinson, and Lynch 1991), repeating the same movie attributes among both parties (i.e., a greater overlap) helps create a stronger association between these movie attributes and movies, which in turn makes it easier for moviegoers to retrieve these movie attributes at the time of evaluation. In other words, topic consistency can increase consumer purchases through better recall.

To test our proposition, we collected both user and critic reviews from Rotten Tomatoes for all movies released between January 2013 and December 2017 in the United States. We also obtained advertising expenditures, daily box office revenues, and other movie characteristics from Box Office Mojo and Kantar Media. Our final sample contains 750 movies. For each movie, we quantify the topic consistency using the degree of overlap in the underlying topics conveyed by user and critic reviews. The underlying topics of review texts were recovered using a bigram latent Dirichlet allocation (LDA) model because this topic model can effectively capture the names of directors and actors.

We find that topic consistency has a significant and positive association with box office revenue—controlling for conventional WOM variables such as review valence, volume, and variance—and that this effect is the greatest for movies that received mediocre review ratings. In addition, the more similar critics' and users' numeric ratings were, the stronger the effect of topic consistency on box office revenue. We conducted further online experiments to replicate the main findings from our observational data, test the causal links, and provide evidence for the proposed mechanisms. Our results suggest that higher topic consistency improves recall of review information, which in turn increases an individual's willingness to watch the movie.

This research contributes to the word-of-mouth (WOM) literature by adding to an understudied area of research, the interplay between different sources of user-generated content. At the same time, our work offers practical implications as well. Because of the positive relationship between topic consistency and movie demand, we recommend that movie studios and partnering advertising agencies strike a careful balance between critics' tastes and the preferences of general consumers when positioning and promoting films. In particular, they should avoid promoting the film in a way that caters to only one party without considering potential feedback from the other.

The remainder of this article is organized as follows. We first review previous research related to this study. Next, we describe the empirical setting and data, followed by the definition and operationalization of topic consistency. Building on this foundation of topic consistency, we present our model and empirical findings. We then examine the moderating role of review rating and valence gap between critics and users and present experiments conducted to investigate potential mechanisms that may drive the effect of topic consistency. We conclude by outlining the managerial implications of our findings, as well as avenues for future research.

Literature Review

Online WOM

As this research aims to examine a new aspect of online reviews (i.e., topic consistency), a highly relevant stream of literature is that pertaining to online WOM. Arguably, two of the most studied facets of online WOM are volume and valence: volume is the intensity of public interest in the product, and valence is the sentiment of WOM (Babić Rosario et al. 2016). Previous research documents a positive relationship both between WOM volume and sales (Duan, Gu, and Whinston 2008; Liu 2006) and between WOM valence and sales (Chen and Xie 2008; Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Dellarocas, Zhang, and Awad 2007; Liu, Steenkamp, and Zhang 2018). In addition to aggregate volume and valence, researchers have also created metrics to capture the variation of volume and valence across people (Clemons, Gao, and Hitt 2006; Godes and Mayzlin 2004; Moe and Schweidel 2012; Moe and Trusov 2011; Sun 2012) and across time (Gelper, Peres, and Eliashberg 2018; Godes and Silva 2012).

Recently, with the aid of text mining tools, researchers have begun to explore topics as a new dimension of WOM content (Buschken and Allenby 2016; Lee and Bradlow 2011; Liu, Lee, and Srinivasan 2019; Liu, Vir Singh, and Srinivasan 2016; Tirunillai and Tellis 2014; Toubia et al. 2019). However, these studies focus on either the aggregate distribution of topics (Mankad et al. 2016; Timoshenko and Hauser 2019) or the change of topics over time (Zhong and Schweidel 2020). In this article, we focus on the variation of topics discussed across people (see Table 1). Specifically, we create a measure from review content, topic consistency, to reflect the degree of overlap or variation among the topics discussed between critics and users. We show that topic consistency can offer additional diagnostic cues on quality when other more accessible cues (e.g., valence) are ambiguous or nondiagnostic.

Table 1. Dimensions of WOM Literature.

Classification of WOM Dimension	Aggregate	Variation Across People		Variation over Time
		Within Groups	**Between Groups**	
Volume	For example, Dellarocas and Narayan (2006); Liu (2006)	For example, Godes and Mayzlin (2004)	Our study	For example, Gelper, Peres, and Eliashberg (2018); our study
Valence	For example, Dellarocas, Zhang, and Awad (2007); Hennig-Thurau, Wiertz, and Feldhaus (2015); Jansen et al. (2009)	For example, Chintagunta, Gopinath, and Venkataraman (2010); Moe and Trusov (2011); Sun (2012)	For example, Deng et al. (2021); our study	For example, Godes and Silva (2012); our study
Topic	For example, Mankad et al. (2016); Timoshenko and Hauser (2019)	For example, Schweidel and Moe (2014)	For example, Deng et al. (2021); our study (main focus)	For example, Puranam, Narayan, and Kadiyali (2017); Zhong and Schweidel (2020); our study

Notes: We have used an exhaustive rule to show our study in all the cells for WOM dimensions considered in this study. For other studies, we have used a "for example" rule so that these studies only appear in some cells as illustrations.

Our research differs substantially from previous literature that measures topics or text similarity in the following ways. First, Homburg, Ehm, and Artz (2015), Ma, Sun, and Kekre (2015), and Wang and Chaudhry (2018) investigate the effect of firm intervention and/or responses on consumer WOM (i.e., rating, consumer sentiment), whereas we study the effect of conversation overlap between different groups of consumers on product sales. Second, instead of measuring the extent of similarity within one consumer group (Herhausen et al. 2019; Schweidel and Moe 2014), we focus on the interactions and dynamics between the WOM of two distinct groups of consumers, namely critics and users. Third, we extend the similarity measure from Wang and Chaudhry (2018) and Deng et al. (2021) by incorporating topic importance in the topic similarity measure and using a daily measure to better capture the temporal variation in topic overlap between critic and user reviews. Lastly, to our knowledge, this research is the first to show that topic variation between groups can affect sales.

Source of Movie Reviews

This research focuses on the interplay between two types of consumer reviews: critic reviews and user reviews. Critics and their reviews play a particularly notable role in the entertainment industry (Handel 1950). According to industry jargon, the term "critics" describes "person[s] usually employed by newspapers, television stations, or other media who screen newly released movies and provide their subjective views and comments on the movie for the public's information" (Cones 1992, p. 120). In line with Holbrook (1999), we view critic reviews as the expert judgments of professional critics, defined as those who assess the artistic success of films from a relatively detached and long-term perspective that focuses on accepted standards for excellence.

Previous research on critic reviews has mainly focused on their role in predicting movie success (Boatwright, Kamakura, and Basuroy 2007; Zhang and Dellarocas 2006). Critics with opinions that are correlated with early box office revenues are called "influencers," while critics with opinions that are correlated with overall box office revenues are called "predictors" (Eliashberg and Shugan 1997). Eliashberg and Shugan (1997) find that critic reviews are correlated with late and overall box office revenues, acting as a mere representative of the audience with no significant influence on the initial box office revenues; thus, they conclude that critics are predictors rather than influencers. In contrast, Basuroy, Chatterjee, and Ravid (2003) find that both positive and negative critic reviews are correlated with weekly box office revenue over eight weeks, suggesting that critics can play dual roles, both influencing and predicting box office revenue. Reinstein and Snyder (2005) find weak evidence of an influence effect after accounting for the spurious correlation between critic reviews and movie quality using a difference-in-differences approach. In addition, critic reviews can affect the movie studio's stock returns in the financial market (Chen, Liu, and Zhang 2012).

Table 2 lists prior empirical research investigating both critic and user reviews under a unified conceptual framework. The columns indicate the type of WOM variables examined in each study. This line of work mainly focuses on whether critic and user reviews are substitutive or complementary in affecting consumer choices (e.g., Amblee and Bui 2007; Moon, Bergey, and Iacobucci 2010). For example, Dellarocas, Zhang, and Awad (2007) show that the information provided by users is complementary to the information provided by professional movie critics in the earlier stage of a movie's life cycle. Further, Chakravarty, Liu, and Mazumdar (2010) demonstrate that the reliance on critics or user ratings can vary by audience. The authors find that the persuasive influence of online user reviews is stronger for infrequent moviegoers than for frequent moviegoers; the latter tend to rely more on critic reviews. Previous research also examines the mediating role of user review valence (Holbrook and Addis 2007) and user review

Table 2. Prior Research Considering Both Critic and User Reviews.

	Context	Valence of Critic Reviews	Valence of User Reviews	Volume of Critic Reviews	Volume of User Reviews	Review Text by Users	Review Text by Users and Critics	Topic Consistency	Link to Financial Outcome	Main Findings
Liu (2006)	Movies	✓	✓	✓	✓				✓	Review volume from both types of reviews affects sales, while valence does not matter
Amblee and Bui (2007)	Freeware	✓	✓	✓	✓					The impacts of both types of reviews are nearly identical
Dellarocas, Zhang, and Awad (2007)	Movie	✓	✓	✓	✓				✓	Online review metrics increase the accuracy of forecasting models for movie sales
Holbrook and Addis (2007)	Movie	✓	✓							Ordinary evaluation mediates the relationship between expert judgment and popular appeal
Chakravarty, Liu, and Mazumdar (2010)	Movie	✓	✓							The persuasive effect of online WOM is stronger on infrequent moviegoers
Moon, Bergey, and Iacobucci (2010)	Movie	✓	✓	✓	✓				✓	High early movie revenues enhance subsequent movie ratings, and sequel movies receive lower ratings but reap higher revenue
Li et al. (2013)	Mobile phone, laptop	✓	✓			✓				Source- and content-based review features have direct impact on product review helpfulness
Wang, Liu, and Fang (2015)	Movie, digital camera, book	✓	✓	✓	✓				✓	Variance of critic and user reviews can lead to customer breadth and depth effect
Zhou and Duan (2016)	Software	✓	✓	✓	✓					Volume of user reviews mediates the positive effect of critic reviews on number of download
Deng (2020)	Movie	✓	✓	✓	✓	✓			✓	User reviews impact sales through aggregate ratings; critic reviews exert their impact through textual narratives
Deng et al. (2021)	Restaurant	✓	✓	✓	✓	✓	✓			Editorial review program leads to a herding effect: users write more like critics do
This study	Movie	✓	✓	✓	✓	✓	✓	✓	✓	We propose a metric to capture the overlap of topics discussed between groups and show that information recall mediates the positive relationship between topic consistency and sales

volume (Zhou and Duan 2016) in the effect of critic reviews on popular appeal and user choices. Wang, Liu, and Fang (2015) show that critic review variance positively moderates the effects of user review variance on sales. More recently, some researchers have started to examine the review content, either by manipulating the content abstractness in an experimental setting (Li et al. 2013) or by extracting prespecified characteristics related to sentiment and social terms (Deng 2020).

Yet despite the growing attention to both user and critic reviews in the WOM literature, we still have scant knowledge about the impact of the interplay between reviews from multiple sources on consumer purchase decisions. To our knowledge, only one study has considered review content by both parties (Deng et al. 2021). Its authors focus on the impact of critic reviews on subsequent user reviews in the food industry. In contrast, we focus on the impact of the overlap in content between critic and user reviews on movie demand. We complement the previous research by bringing these two review sources (i.e., critics and users) together and examining the consequence of their review content jointly through the lens of topic consistency.

Effect of Topic Consistency

As outlined previously, we posit that the potentially positive association between topic consistency and consumer purchases can be attributed to information recall. Notably, the underlying topics of each movie review can be thought of as different attributes of a particular movie. As individuals read reviews, they receive cues along with these movie attributes. When critics and users comment on the same topic, thus attaining high topic consistency, the potential moviegoers will receive more focused cues along with the more salient movie attributes associated with the topic. The more focused cues will then make these attributes more memorable. Following this reasoning, repeated presentation of the same topic from both users and critics may enhance recall, which may in turn increase the impact of associated movie attributes on moviegoing choices. Since the association between review topics and movie attributes is harder to retrieve from memory than the average rating of movies, we expect the effect of topic consistency to be dampened in the presence of more accessible cues, such as extremely positive or negative average movie ratings. In contrast, when more accessible cues are nondiagnostic or ambiguous, consumers are more motivated to use less accessible cues such as movie attributes in decision making (Gardial and Biehal 1985), leading to a more salient effect of topic consistency.

Consumers often rely on memory to form consideration sets (Alba, Hutchinson, and Lynch 1991), and repetition can facilitate recall (Crowder 2015). According to Cacioppo and Petty's (1980) illustration, formulations of cognitive responses and attitudes for repeated messages depend on the convincingness of the message, the information processing effort involved, and reexposure, which provides new opportunities for the message to be attended to, comprehended, and encoded. Of course, each review's content is rarely repeated verbatim in the movie review setting, unless the same person posts the same review multiple times. However, at the topic level—that is, at the level of "what they talk about"—there are likely to be repeated topics that both critics and reviewers mention (e.g., casts). When the number of total reviews from critics and users is held constant, a large overlap in topics can reexpose the topic, providing opportunities to attend to and comprehend the topic and encode the movie. Such repetition of a topic among both critics and users will help consumers create a clear association between the topic and the movie, which can later be easily retrieved from memory at the time of purchase.

We expect the effect of the repetition of topics between critics and users to be more pronounced than repetition of topics within critics or users because the variance in the sources of information (critics vs. users) could make information more convincing and memorable. According to the encoding variability theory (Lord and Putrevu 1998), consumers who encounter the same message in more than one context have access to multiple retrieval cues, resulting in a stronger, clearer, and more accessible information network in the brain, which further leads to more accurate recall (Burmester et al. 2015; Stammerjohan et al. 2005). Therefore, whether topic consistency is measured within or between groups can be another boundary condition in addition to movie valence.

Setting and Data

Movie Review Data

We use the online review platform Rotten Tomatoes as our primary source of review data. Rotten Tomatoes is widely known for publishing both critics' and users' movie reviews. Critics express their judgment of the movie in review text and give the movie an overall label of either "rotten" (unfavored) or "fresh" (favored). Each user can also leave a review text and a numeric rating on a scale of 0 through 5 (5 being the highest). Because of the platform's popularity among movie critics and general users, we find Rotten Tomatoes suitable for collecting critics' and users' opinions about movies. For each movie in our sample, we collected all reviews by critics and users, including prerelease reviews, up to eight weeks after the movie's release because the first eight weeks account for 97.97% of the box office revenue in the data.

Using the movie list provided by Rotten Tomatoes, we sampled 951 movies released in the United States between January 2013 and December 2017. From these movies, we gathered a sample of 712,126 reviews from Rotten Tomatoes: 128,983 reviews from critics (18.11%) and 583,143 reviews from users (81.89%). We use this movie review data set to identify the key topics from review texts, which are further used to measure the topic consistency.

We also collected the review rating for both user reviews (0–5 scale) and critic reviews (rotten–fresh rating) to capture the valence. To make the ratings comparable, we converted the binary ratings of critic reviews (i.e., rotten = 0, fresh = 1) to a

0–5 scale (rotten = 0, fresh = 5).[1] In addition to numeric review ratings, we applied sentiment analysis to review texts to create a text-based valence measure for each review using the Bing lexicon dictionary (please see Web Appendix A for more details). The text-based review valence is between −1 and 1, with 1 being the most positive and −1 being the most negative.

Movie Box Office Data

We collected box office revenue from reporting site Box Office Mojo. Of the original sample of 951 movies, 936 had daily box office revenue information during the first eight weeks of the movie's release. We further dropped 186 movies that did not have ad spending information available from Kantar Media or had a limited release, leaving 750 movies in the subsequent empirical analysis. We focus on movies with a wide release to alleviative the potential bias caused by sequential rollout found in Chintagunta, Gopinath, and Venkataraman (2010). The 186 dropped movies were mostly from niche distributors (categorized as "others"), accounting for only 2.09% of the market share. We gathered three additional pieces of information for the 750 remaining movies from both Box Office Mojo and IMDb: number of theaters, genre, and distribution studio. To capture the effect of competition, we computed the number of movies in play from the same genre as the focal movie.

Table 3 lists the key variables used in this study along with their descriptions. The descriptive statistics of these variables are in Table 4. On average, the volume of cumulative user reviews is five times more than that of critic reviews. As expected, the mean valence (i.e., numeric rating) of cumulative critic reviews (2.75) is lower than that of cumulative user reviews (3.18). We observe a similar pattern in review texts as the mean text-based valence is .16 for cumulative critic reviews and .35 for cumulative user reviews. On average, each movie played in 1,465 theaters and generated $1.25 million per day.

Topic Consistency Between Users and Critics

In this section, we explain step by step how we measured topic consistency, defined as the degree of overlap in underlying topics between user and critic reviews.

Preprocessing Review Data

Before analyzing the review text, we removed the noninformative textual content, transformed certain words to increase the reliability, and reduced the computational overhead. First, we lowercased all review text and removed punctuation. Second, we removed numbers, excessive spaces, and stop words that are generally recognized as meaningless (see Schofield, Magnusson, and Mimno 2017). Third, we transformed the remaining words into stem words, that is, into their root form (e.g., "like" for "liked" or "liking"). Lastly, to ensure that outlier words would not influence the results, we only included terms with a term frequency–inverse document frequency value of at least .1 (Blei and Lafferty 2009) for the text analysis.

Topic Model

We employed LDA to extract the underlying topics from critic and user reviews. The unigram LDA, introduced by Blei, Ng, and Jordan (2003), is one of the most widely employed techniques in text mining and has been applied to the field of marketing (e.g., Tirunillai and Tellis 2014). Following the terminology used by Blei, Ng, and Jordan, we regard each movie review as a document. A movie review with a sequence of N words is denoted by $r = (w_1, w_2, \ldots, w_N)$, where each w denotes a word. Assuming there are M reviews in total, our corpus is a collection of total M reviews denoted by $D = \{r_1, r_2, \ldots, r_M\}$. Let T denote the total number of topics and W denote the number of unique words appearing in M reviews. The data-generating process of the unigram LDA can be described as follows:

1. The distribution over all the words for each topic z is drawn from $\beta_z \sim \text{Dirichlet}(\eta)$, where η is a W-dimensional vector.
2. For each movie review r:
   a. The topic proportion for each movie review θ_r is drawn from $\theta_r \sim \text{Dirichlet}(\alpha)$, where α is a T-dimensional vector.
   b. For each word contained in the movie review r:
      i. Draw a topic assignment variable $z \sim \text{Multinomial}(\theta_r)$,
      ii. Draw a word $w \mid z, \beta \sim \text{Multinomial}(\beta_z)$, where β_z denotes the probability of words occurring in topic z.

Bigram LDA. We employed a bigram LDA, which extends the unigram LDA by incorporating a notion of word order.[2] The unit of analysis in a bigram model (w) is two consecutive words rather than a single word as used in the conventional unigram LDA (Lau, Baldwin, and Newman 2013). We utilized bigram LDA in this study because it can better detect names of actors, directors, and movie titles, which usually involve multiple sequential words. For example, for the term "New York," "New" and "York" are considered separate terms in the unigram LDA, whereas "New York" can be represented as one term in the bigram LDA. To implement the bigram LDA, we used the bigram tokenizer to convert the word-level review matrix (so-called document term matrix) into a two-word-based matrix. The rest of the procedure is the same. Notably, the bigram LDA is more computationally demanding

[1] We show in Table W1 in Web Appendix A that our main finding is robust to the mapping of critic reviews to numeric ratings.

[2] We reran our analysis with unigram LDA as a robustness check, and the results are consistent (Web Appendix A, Table W2).

Table 3. Variable Description.

Variable Name	Description	Source
Dependent variable		
$LnBoxRev_{i,t}$	Log of box office revenue for movie i on day t	Box Office Mojo
Independent variables		
Review volume, valence, variance measures		
$LnCumCriticVol_{i,t}$	Log of number of cumulative critic reviews for movie i until day t	Rotten Tomatoes
$LnCumUserVol_{i,t}$	Log of number of cumulative user reviews for movie i until day t	Rotten Tomatoes
$LnCumCriticVal_{i,t}$	Log of average ratings of cumulative critic reviews for movie i until day t	Rotten Tomatoes (rotten = 0; fresh = 5)
$LnCumUserVal_{i,t}$	Log of average ratings of cumulative user reviews for movie i until day t	Rotten Tomatoes (0–5 scale)
$VarCumCriticVal_{i,t}$	Variance of ratings by critics for movie i until day t	Rotten Tomatoes
$VarCumUserVal_{i,t}$	Variance of ratings by users for movie i until day t	Rotten Tomatoes
$Abs(ValenceGap)_{i,t}$	Absolute gap between $LnCumCriticVal_{i,t}$ and $LnCumUserVal_{i,t}$	Rotten Tomatoes
$CumCriticTextVal_{i,t}$	Mean text-based valence of cumulative critic reviews for movie i until day t	—
$CumUserTextVal_{i,t}$	Mean text-based valence of cumulative user reviews for movie i until day t	—
$VarCumCriticTextVal_{i,t}$	Variance of the text-based valence of cumulative critic reviews for movie i until day t	—
$VarCumUserTextVal_{i,t}$	Variance of the text-based valence of cumulative user reviews for movie i until day t	—
Review topic consistency measures		
$TopicConsistency(C, U)_{i,t}$	Degree of the topic consistency between cumulative critic and user reviews for movie i until day t	—
$TopicConsistency(U, U)_{i,t}$	Degree of the topic consistency within cumulative user reviews for movie i until day t	—
$TopicConsistency(C, C)_{i,t}$	Degree of the topic consistency within cumulative critic reviews for movie i until day t	—
$TopicConsistency(All)_{i,t}$	Degree of the topic consistency for all cumulative reviews for movie i until day t	—
$ValConsistency(C, U)_{i,t}$	Degree of consistency in the valence of topics between cumulative critic and user reviews for movie i until day t	—
Controls		
$NComp_{i,t}$	Number of other movies from the same genre played on day t after movie i's release	—
$LnTheaters_{i,t}$	Log of number of theaters that play movie i on day t after the release	Box Office Mojo
$LnAdv_{i,t}$	Log of daily advertisement expenditure for movie i on day t after the release	Kantar Media
Age, Year, Month, Day of Week, Holiday	Dummies for the number of days after the release, year, month, day of the week, and holiday	—

Notes: We calculate $ValConsistency(C, U)_{i,t}$ in the following way. Using the top 100 bigrams from the LDA model, we first gather the sentences that contains the bigrams from the critic review and user review until day t respectively, and compute the text sentiment with the Bing dictionary. Next, we calculate the sentiment proportion (positive/(positive + negative)) for each topic for user review and critic review. We then compute the valence consistency on the basis of the cosine similarity of the sentiment proportion vectors.

than the unigram LDA because of the inclusion of a much larger set of vocabulary (i.e., $W \times (W - 1)$).

Optimal number of topics. In LDA, the number of topics T is a hyperparameter that must be predetermined. We followed the literature to select the optimal number of topics T using the perplexity score (Blei, Ng, and Jordan 2003). In particular, we ran multiple bigram LDA topic models with different values of T to pinpoint the optimal number of topics using the movie review data. We selected $T = 25$ on the basis of the lowest average perplexity score from the fivefold cross-validation.

Identified topics. The bigram LDA model returns topic association probabilities for each review, and the bigram term probabilities associated with each topic. The topic model does not return an actual label or name for each topic. Instead, the label of each topic is often determined by human coders on the basis of terms with the highest associated probabilities for each topic. To aid the topic interpretation, we select the topic label on the basis of topic relevance, which is the weighted average of bigram-topic probabilities and its lift (Sievert and Shirley 2014). Table W6 in Web Appendix B shows the top ten bigram words in each topic.

Table 4. Descriptive Statistics.

			Percentiles				
	Mean	**SD**	**5th**	**25th**	**50th**	**75th**	**95th**
Box Office Revenue (in $1,000)							
$BoxRev_{i,t}$	1,247.593	3,585.394	3.445	35.848	220.259	993.509	5,476.589
Review-Related Variables							
$CumCriticVol_{i,t}$	129.210	68.330	21	80.750	123	175	250
$CumUserVol_{i,t}$	658.900	1,005.469	28	148	297	671	2,753.650
$CumCriticVal_{i,t}$	2.749	1.433	.446	1.488	2.917	4.070	4.750
$CumUserVal_{i,t}$	3.180	.700	1.902	2.751	3.238	3.700	4.180
$CumCriticTextVal_{i,t}$	.155	.281	−.297	−.034	.154	.356	.600
$CumUserTextVal_{i,t}$	.345	.290	−.179	.159	.376	.553	.758
$VarCumCriticVal_{i,t}$	4.209	1.795	.980	2.769	4.626	5.849	6.290
$VarCumUserVal_{i,t}$	2.274	.892	.987	1.664	2.203	2.768	3.856
$TopicConsistency(C, U)_{i,t}$	.888	.017	.860	.878	.889	.899	.913
$TopicConsistency(U, U)_{i,t}$	.865	.027	.822	.850	.867	.883	.906
$TopicConsistency(C, C)_{i,t}$	.919	.007	.907	.915	.919	.923	.929
$TopicConsistency(All)_{i,t}$	.887	.017	.859	.874	.887	.898	.915
$ValConsistency(C, U)_{i,t}$	.671	.189	.302	.568	.707	.811	.917
Control Variables							
$NComp_{i,t}$	5	5.431	0	2	5	8	19
$Theaters_{i,t}$	1,465	1,376.586	17	203	1,044	2,705	3,761
$Adv_{i,t}$ (in $1,000)	46.036	139.147	0	0	0	3.486	287.056
Genre							
Action	.146	.353					
Adventure/sci-fi	.143	.350					
Comedy	.176	.381					
Drama	.275	.446					
Family	.047	.211					
Horror	.069	.253					
Musical	.013	.112					
Romance	.034	.181					
Thriller	.099	.299					
Distribution studio							
Disney	.016	.127					
Fox	.135	.341					
Lionsgate	.091	.288					
Paramount	.071	.258					
Sony	.076	.265					
Universal	.096	.294					
Warner Brothers	.132	.339					
Others	.383	.486					

Notes: Descriptive statistics of non-log-transformed variables are reported.

We show the proportions of user and critic reviews over the 25 topics in Figure 1 to visualize the extent to which users and critics "talk" similarly about the movie. To do so, we first calculate the topic–review association by identifying the key topic that has the highest probabilistic association with each critic review. We then compute the distribution of critic reviews across 25 topics by dividing the number of critic reviews associated with each topic by the total number of critic reviews. We generate the distribution of user reviews in a similar way.

We observe a moderate discrepancy between the topics that frequently appear in critic reviews and the topics in user reviews in Figure 1. However, overall, we find that subjective judgments on movies are more directly expressed in user reviews than in critic reviews, as can be seen in Topics 1 (Ambivalence), 3 (Enjoyment), and 7 (Usage of resources). Critic reviews, however, seem to focus on details or provide additional insight on the film, such as in Topics 8 (Movie details), 14 (Award assessment), 17 (Story-related), and 25 (Entertainment). The correlation of the two distributions is .45, suggesting that the

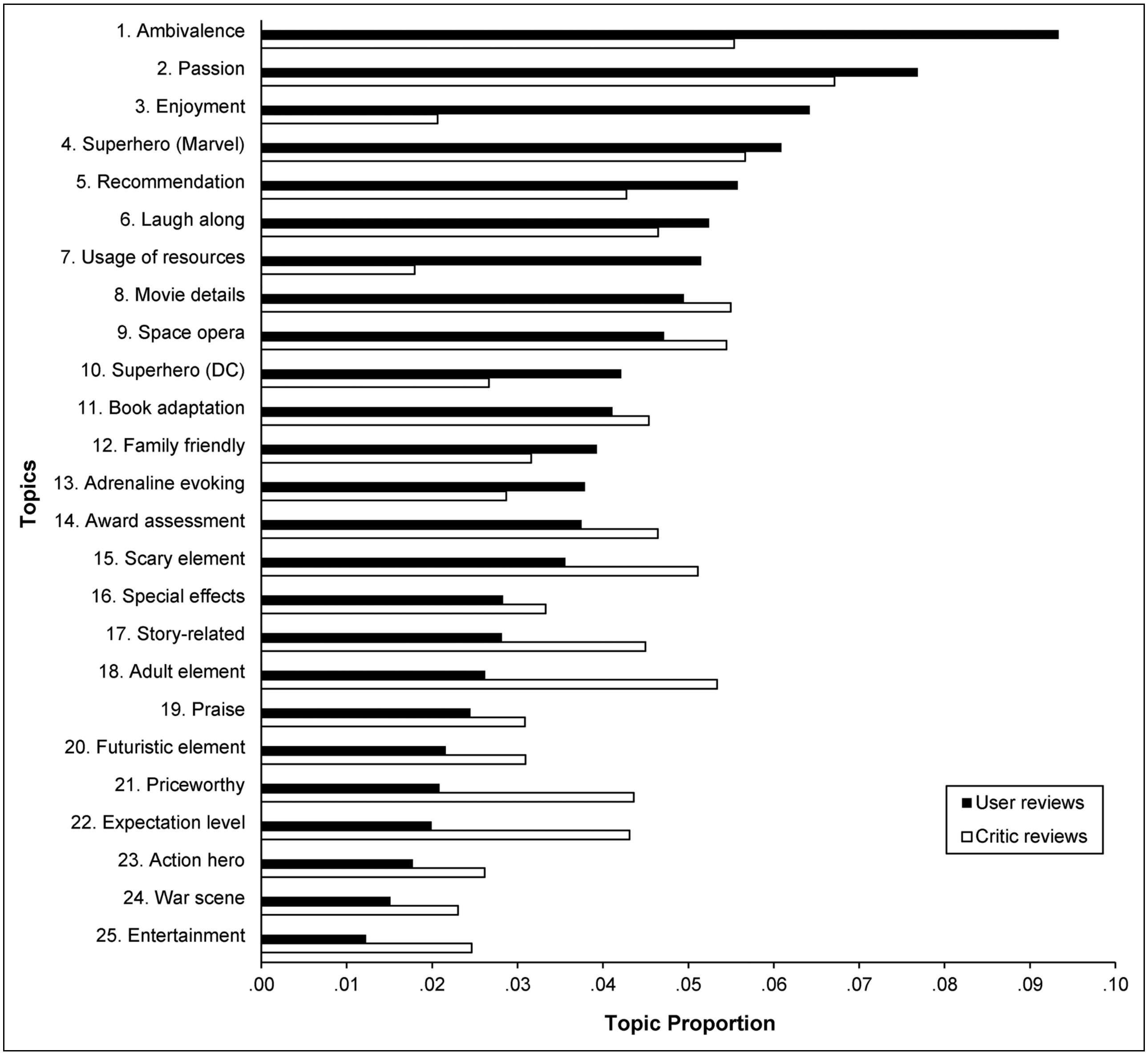

Figure 1. Topic Coverage Between Users and Critics.

topics mentioned by users are not always aligned with the key topics mentioned by critics.

Measuring Topic Consistency

With the discrepancy between user and critic review topics in mind, we measured the topic consistency between the two groups using the weighted cosine similarity on user–critic pairwise topic weight. Let $\boldsymbol{\theta}_i$ denote a $T \times M$ probability matrix of T topic mixtures for movie i with total M reviews from users and critics. Then, $\theta_i(t, r)$ stands for a topic probability of topic t for review r, and $\boldsymbol{\theta}_i(:, r)$ stands for a $T \times 1$ vector of all topic probabilities associated with review r. Assume that there are J cumulative critic reviews and K cumulative user reviews for movie i on day d after release so that $M = J + K$.

We operationalize the movie-level topic consistency in three steps. First, note that not all topics are equally important, as some topics appear more frequently than others in review texts. To account for the relative importance of each topic t, we measure the relative topic importance denoted by $w_{id}(t)$ as follows. We first define the relative topic importance from user reviews $w_{idU} = \left[\sum_{j=1}^{J_{id}} \boldsymbol{\theta}_i(:,\ j)\right] / J_{id}$ as the mean of topic weight vectors from J cumulative user reviews on day d where j stands for user review. Similarly, we define the relative topic importance from critic reviews $w_{idC} = \left[\sum_{k=1}^{K_{id}} \boldsymbol{\theta}_i(:,\ k)\right] / K_{id}$ as the mean of topic weight vectors from K cumulative critic

Topic	Critic 1	Critic 2	User 1	User 2	User 3
	La La Land is probably going to win Best Picture at next year's Oscars.	La La Land is a romance, and everyone is in love with everything—most of all, old Hollywood dreams. It is, quite simply, magic.	Charmingly un-ironic, yet self-aware. Bravura opening and closing scenes. Will win many awards and will deserve the ones for art direction and Emma Stone.	This should win best picture!	I LOVED THIS FILM. If you like the old time musicals you will love love love this film. It's been a long time since I felt good walking out of a theater!
1. Ambivalence	.052	.033	.030	.038	.032
2. Passion	.034	**.067**	.045	.038	**.143**
3. Enjoyment	.034	.033	.030	.038	.048
4. Superhero (Marvel)	.034	.033	.030	.038	.048
5. Recommendation	.052	.033	.061	.038	.032
6. Laugh along	.034	.033	.030	.038	.032
7. Usage of resources	.034	.050	.045	.038	.032
8. Movie details	.034	.050	.061	.038	.032
9. Space opera	.034	.033	.030	.038	.032
10. Superhero (DC)	.034	.033	.030	.038	.032
11. Book adaptation	.052	.033	.030	.038	.032
12. Family friendly	.034	.050	.045	.038	.095
13. Adrenaline evoking	.034	.033	.030	.038	.032
14. Award assessment	**.103**	.050	**.121**	**.077**	.032
15. Scary element	.034	.033	.030	.038	.032
16. Special effects	.034	.033	.030	.038	.032
17. Story-related	.034	.033	.030	.038	.032
18. Adult element	.034	.033	.030	.038	.032
19. Praise	.034	.050	.045	.038	.032
20. Futuristic element	.052	.033	.030	.038	.032
21. Priceworthy	.034	.033	.045	.038	.032
22. Expectation level	.034	.050	.045	.038	.032
23. Action hero	.034	.033	.030	.038	.032
24. War scene	.034	.050	.030	.038	.032
25. Entertainment	.034	.050	.030	.038	.032
Total	1	1	1	1	1

Figure 2. Examples of Critic and User Reviews for *La La Land*.

reviews on day d.[3] We then define the relative topic importance for a given topic t as $w_{id}(t) = [w_{idU}(t) + w_{idC}(t)]/2$, the mean of relative topic importance from user and critic reviews. A greater $w_{id}(t)$ means a more frequently used topic in reviews. We show in Table W3 in Web Appendix A that the choice of the relative weight between user and critic reviews in the construction of $w_{id}(t)$ does not much affect our main findings. Next, we calculate the similarity for each pair of user review j and critic review k using the cosine similarity between $\boldsymbol{\theta}_i(:, j)$ and $\boldsymbol{\theta}_i(:, k)$, weighted by the vector of relative topic importance ($\mathbf{w}_{id}$). Let $\text{WTopicSimilarity}_{ijkd}$ denote the pairwise weighted topic similarity between a critic review j and a user review k for a movie i on day t. We can express $\text{WTopicSimilarity}_{ijkd}$ as follows:

$$\text{WTopicSimilarity}_{ijkd} = \frac{\sum_{t=1}^{T} w_{id}(t)\, \theta_i(t, j)\theta_i(t, k)}{\sqrt{\sum_{t=1}^{T} w_{id}(t)\theta_i^2(t, j)}\sqrt{\sum_{t=1}^{T} w_{id}(t)\theta_i^2(t, k)}}. \quad (1)$$

Lastly, we aggregate the pairwise weighted topic similarity to the movie and daily level by taking the mean of $\text{WTopicSimilarity}_{ijkd}$ across all $J \times K$ pairs of reviews. Specifically, we operationalize the topic consistency between critics and users denoted by $\text{TopicConsistency}(C, U)_{id}$ as

$$\text{TopicConsistency}(C, U)_{id} = \frac{1}{J \times K}\sum_{k=1}^{K}\sum_{j=1}^{J} \text{WTopicSimilarity}_{ijkd}. \quad (2)$$

To better illustrate the calculation of topic consistency, we provide an example in Figure 2, which presents two reviews from critics and three reviews from users for the movie *La La Land*. Each column reports a 25-by-1 vector of topic probabilities of the corresponding review (i.e., $\boldsymbol{\theta}_i(:, r)$), where the bar visualizes the topic probability for each of the 25 topics. To measure $\text{WTopicSimilarity}_{ijkd}$, we compute the weighted cosine similarity of topic probability between each pair of critic and user reviews. With two critic reviews and three user reviews, we have six pairs in this example. The vector $\mathbf{w}$ represents the relevance of all 25 topics in these five reviews. As Critic 1, User 1, and User 2 all commented on the award potential of the movie (i.e., topic 14), this award-related topic receives the highest weight. Finally, we take the mean of $\text{WTopicSimilarity}_{ijkd}$ across the six pairs of reviews to obtain $\text{TopicConsistency}(C, U)_{id}$.

The topic consistency between cumulative critic and user reviews ranges from .67 to 1, with a mean of .89 (see Table 4). We also depict the histogram of topic consistency in Figure 3, which shows significant variation across movies and days. We report the correlations between the WOM volume, valence, and topic consistency in Table 5. The relatively

[3] These are the relative topic importance values we used to calculate TopicConsistency(U, U) and TopicConsistency(C, C) respectively. For TopicConsistency(All), the relative topic importance is defined as $w_{idAll} = \left[\sum_{m=1}^{M_{id}} \boldsymbol{\theta}_i(:, m)\right]/M_{id}$ where m stands for all the user and critic reviews.

small correlations between topic consistency and log-transformed cumulative average critic valence ($r = -.186$) and between topic consistency and log-transformed cumulative average user valence ($r = .041$) suggest that the overlap in review content does not necessarily associate with critics' and users' attitudes toward a movie. We also compute the correlation between topic consistency and the absolute valence gap between critics and users (i.e., $|\text{CumCriticVal} - \text{CumUserVal}|$). Again, we do not find a strong correlation ($r = .188$). As all correlations between topic consistency and the volume- or valence-related measures are relatively small, the proposed topic consistency seems to capture an additional facet of WOM revealed by the review texts from critics and users.

Empirical Analysis

Box Office Revenue Model

Let i denote movies and t denote the number of days since the movie's release date. The dependent variable $\text{LnBoxRev}_{i,t}$ represents the log-transformed daily box office revenue for movie i on day t. We investigate the relationship between topic consistency and box office revenue by considering the following model:

$$\text{LnBoxRev}_{i,t} = \underbrace{\beta_1 \text{LnCumCriticVal}_{i,t-1} + \beta_2 \text{LnCumUserVal}_{i,t-1} + \beta_3 \text{LnCumCriticVol}_{i,t-1}}_{\text{traditional WOM measures}}$$

$$\underbrace{+\beta_4 \text{LnCumUserVol}_{i,t-1} + \beta_5 \text{VarCumCriticVal}_{i,t-1} + \beta_6 \text{VarCumUserVal}_{i,t-1} + \beta_7 \text{Abs(ValenceGap)}_{i,t-1}}_{\text{traditional WOM measures}}$$

$$\underbrace{+\beta_8 \text{TopicConsistency(C, U)}_{i,t-1} + \beta_9 \text{ValConsistency(C, U)}_{i,t-1} + \beta_{10} \text{CumCriticTextVal}_{i,t-1}}_{\text{text-based WOM measures}}$$

$$\underbrace{+\beta_{11} \text{CumUserTextVal}_{i,t-1} + \beta_{12} \text{VarCumCriticTextVal}_{i,t-1} + \beta_{13} \text{VarCumUserTextVal}_{i,t-1}}_{\text{text-based WOM measures}}$$

$$\underbrace{+\beta_{14} \text{LnAdv}_{i,t-1} + \beta_{15} \text{LnTheaters}_{i,t} + \beta_{16} \text{NComp}_{i,t} + \beta_{17} \text{HOLIDAY}_{i,t} + \sum_{j=1}^{55} \alpha_j \text{I}(\text{AGE}_{i,t} = j)}_{\text{controls}}$$

$$\underbrace{+\sum_{j=1}^{6} \gamma_j \text{I}(\text{DAYOFWEEK}_{i,t} = j) + \theta_i + \omega_t + \epsilon_{i,t},}_{\text{controls}} \quad (3)$$

where $\text{I}(\cdot)$ is an indicator function and $\epsilon_{i,t}$ is the error term.

The right-hand side of Equation 3 includes three sets of variables. The first set is traditional WOM variables, including the valence, volume, and variance of cumulative reviews, where we calculate the valence and variance using the numeric ratings of reviews. To separate the effect of similarity in topics from the effect of similarity in valence, we control for the similarity in ratings by including $\text{Abs(ValenceGap)}_{i,t-1}$, the absolute gap between the log-transformed valence of cumulative critic reviews and cumulative user reviews.

The second set of variables consists of WOM measures inferred from review texts (i.e., text-based measures), including the review topic consistency between critics and users ($\text{TopicConsistency(C, U)}_{i,t-1}$), our focal variable of interest. We add additional controls for review valence by including the mean and variance of the sentiment of review texts (i.e., text valence). We further control for the valence similarity using the overall consistency of the valence of topics denoted by $\text{ValConsistency(C, U)}_{i,t-1}$, which is calculated in several steps. For each topic, we first sample all the sentences that include the top 100 bigrams for this topic from critic reviews. We use these sentences to measure topic valence from the critics' perspective. In particular, we compute the topic valence using the proportion of positive sentences from the sample. Using a similar method, we can create the topic valence for user reviews. Given the vectors of topic valence for users and critics, we then use the cosine similarity to measure the consistency of the valence of topics.

Lastly, we control for advertising, the number of theaters in play, and competition. We control for seasonality by including dummy variables for national holidays, days since the release ($\text{AGE}_{i,t}$), the day of the week (since Fridays and Saturdays see far more theater traffic than, e.g., Mondays), and the month and year fixed effects (ω_t). We do not include lagged DV as an independent variable because we believe our current model specification adequately captures the dynamics in box office revenue over the release window by incorporating various fixed effects. In particular, we include release-date fixed effects (i.e., the first day after release, the second day after release, etc.) in all our models to flexibly capture the box office revenue pattern over time. Also, the high R^2 indicates

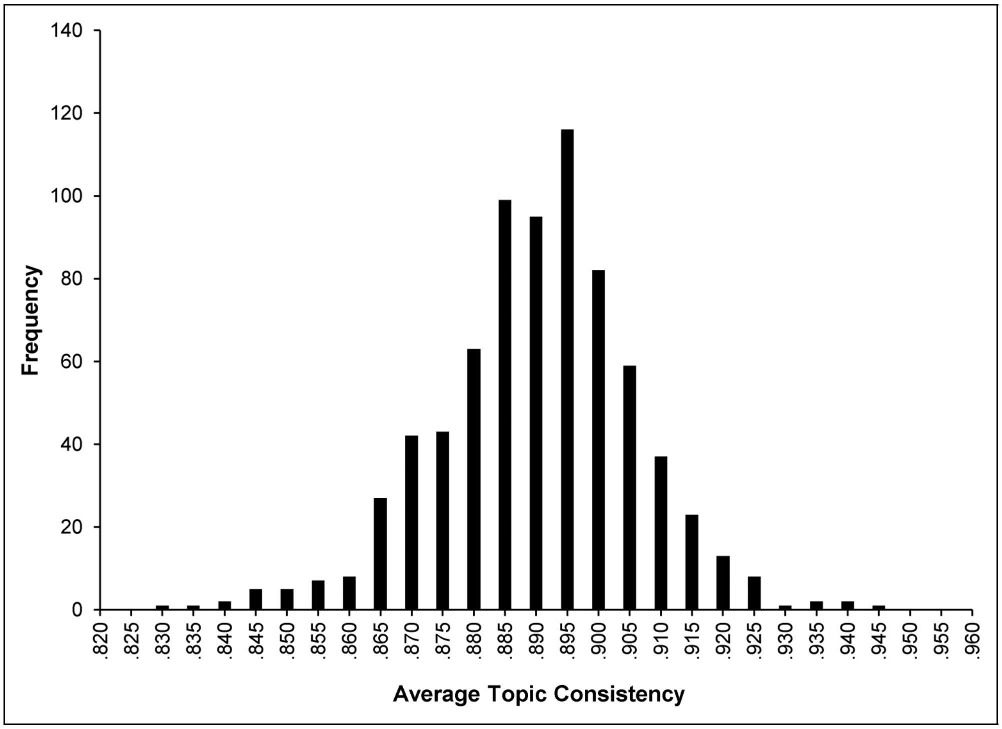

Figure 3. Histogram of Topic Consistency Between Critics and Users.

Table 5. Correlations Among Key WOM Measures.

	LnCumCriticVol	LnCumUserVol	LnCumCriticVal	LnCumUserVal	TopicConsistency
LnCumCriticVol	1	—	—	—	—
LnCumUserVol	.632	1	—	—	—
LnCumCriticVal	.406	.121	1	—	—
LnCumUserVal	.287	.248	.605	1	—
TopicConsistency	−.268	.100	−.186	−.041	1

Notes: All significant at $p < .01$.

that the model fit is not a primary concern. Because of the inclusion of movie fixed effects (θ_i), we do not include any time-invariant movie characteristics (e.g., genre, studio, budget, star power) in Equation 3 due to the lack of identification.

Endogeneity Issues

Despite the control for unobserved static movie characteristics through movie fixed effects, a potential endogeneity problem may arise if there still exist time-variant unobserved factors that correlate with certain covariates. For example, unobserved market information held by movie studios may lead to correlation between the error term and marketing-mix variables (i.e., LnAdvertising, LnTheaters). Following the literature, we instrument advertising and the number of theaters using the mean value of similar movies from the same genre as the focal movie i on the same day t after release (e.g., Chintagunta, Gopinath, and Venkataraman 2010; Lu, Wang, and Bendle 2020; Ryoo, Wang, and Lu 2021). The rationale for these instrumental variables (IVs) is that movies with similar traits are likely to share similar promotional strategies and release patterns, thus satisfying the relevance condition. These IVs are also likely to satisfy the exclusion restriction condition as the marketing mix set by other movies at different times is unlikely to correlate with the focal movie's demand shock.

WOM variables can also be endogenous because the number of potential user reviewers can be affected by the unobserved demand shock, affecting the number of reviews. We instrument user review volume (LnCumUserVol) using other previously released movies from the same genre as the focal movie i on the same day t after release. However, regarding review valence, the Hausman-type instrument based on similar movies is not strong, as indicated by the small F-statistic (3.383) from the first-stage regression. Because of the unavailability of a strong and valid IV for review valence, we follow the suggestion by Rossi (2014) to not use the weak IV for review

Table 6. Two-Stage Least Squares Fixed-Effect Estimation Results of Box Office Revenue Model.

	DV = $\mathbf{LnBoxRev_{i,t}}$									
	(0)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	
$LnCumCriticVal^{\dagger}_{i,t-1}$	.343*	.904***	.898***	.903***	.905***	.908***	.800***	.822***	.670***	+
	(.175)	(.232)	(.237)	(.232)	(.237)	(.241)	(.215)	(.215)	(.191)	
$LnCumUserVal^{\dagger}_{i,t-1}$	.578*	.967***	.987***	.949***	1.010***	1.025***	1.021***	1.009***	.379**	+
	(.303)	(.254)	(.258)	(.251)	(.260)	(.274)	(.266)	(.266)	(.152)	
$LnCumCriticVol_{i,t-1}$	−.155	−.252	−.269	−.246	−.283	−.291	−.244	−.247	−.019	+
	(.103)	(.203)	(.216)	(.203)	(.216)	(.244)	(.203)	(.203)	(.063)	
$LnCumUserVol_{i,t-1}$	−.271	−.239	−.255	−.242	−.244	−.304	−.242	−.241	.101***	+
	(.138)	(.194)	(.198)	(.195)	(.195)	(.197)	(.194)	(.194)	(.036)	
$VarCumCriticVal_{i,t-1}$	.008	−.010	−.025	−.008	−.023	−.009	−.016	−.015	−.046	−
	(.027)	(.035)	(.034)	(.035)	(.034)	(.035)	(.035)	(.035)	(.030)	
$VarCumUserVal_{i,t-1}$	−.060**	−.021	−.026	−.022	−.023	−.033	−.017	−.018	.043	−
	(.031)	(.072)	(.073)	(.072)	(.073)	(.075)	(.072)	(.072)	(.029)	
$Abs(ValenceGap)^{\dagger}_{i,t-1}$	−.001	.157	.142	.162	.142	.158	.080	.059	−.181	−
	(.168)	(.178)	(.177)	(.179)	(.178)	(.184)	(.151)	(.149)	(.122)	
$TopicConsistency(C, U)^{\dagger}_{i,t-1}$		2.724***			4.678**		4.344***	4.758***	3.037***	+
		(.899)			(1.960)		(1.304)	(1.322)	(1.143)	
$TopicConsistency(C, C)_{i,t-1}$			10.401**		8.051					+
			(4.061)		(4.290)					
$TopicConsistency(U, U)_{i,t-1}$				1.270***	−1.368					+
				(.505)	(1.069)					
$TopicConsistency(All)_{i,t-1}$						5.720**				+
						(2.794)				
$ValConsistency(C, U)_{i,t-1}$		−.170	−.203	−.185	−.163	−.099	−.177	−.172	−.105	+
		(.121)	(.119)	(.120)	(.121)	(.110)	(.121)	(.120)	(.093)	
$CumCriticTextVal_{i,t-1}$		−.044	.045	−.046	.022	−.078	−.027	−.025	.159	+
		(.254)	(.241)	(.254)	(.243)	(.268)	(.253)	(.251)	(.197)	
$CumUserTextVal_{i,t-1}$		.120	.089	.116	.111	.085	.123	.119	.164*	+
		(.133)	(.134)	(.135)	(.132)	(.139)	(.133)	(.133)	(.091)	
$VarCumCriticTextVal_{i,t-1}$		.199	.290	.219	.240	.207	.202	.194	.112	−
		(.281)	(.275)	(.281)	(.279)	(.286)	(.281)	(.280)	(.262)	
$VarCumUserTextVal_{i,t-1}$		.144	.138	.145	.139	.154	.151	.162	.131	−
		(.108)	(.108)	(.110)	(.105)	(.113)	(.107)	(.107)	(.091)	
$LnAdvertising_{i,t-1}$	.052***	.054***	.051***	.053***	.053***	.062***	.054***	.053***	.034***	+
	(.016)	(.018)	(.018)	(.018)	(.018)	(.019)	(.018)	(.018)	(.004)	
$LnTheaters_{i,t}$	.918***	.945***	.946***	.944***	.949***	.972***	.944***	.945***	.816***	+
	(.036)	(.045)	(.046)	(.045)	(.047)	(.059)	(.045)	(.045)	(.009)	
$TopicConsistency(C, U)^{\dagger}_{i,t-1} \times Abs(ValenceGap)^{\dagger}_{i,t-1}$							−5.777			−
							(3.525)			
$TopicConsistency(C, U)^{\dagger}_{i,t-1} \times Pos(ValenceGap)^{\dagger}_{i,t-1}$								−9.830**	−8.508**	−
								(4.255)	(3.745)	
$TopicConsistency(C, U)^{\dagger}_{i,t-1} \times Neg(ValenceGap)^{\dagger}_{i,t-1}$								−2.047	−3.881	−
								(3.735)	(3.150)	
Endogeneity correction	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	
Observations	37,762	37,762	37,762	37,762	37,762	37,762	37,762	37,762	37,762	
R^2	.973	.974	.974	.974	.974	.973	.974	.974	.977	

**p <.05, ***p <.01.
†Mean-centered in models with interaction terms.
Notes: Standard error clustered by movies in parentheses. All models include movie, movie age, year, month, day of week, and holiday fixed effects and competitors. The last column indicates the expected sign.

valence to avoid substantial finite sample bias and sampling errors.

Unlike most user reviews, critic reviews are often produced by invitation and take substantial time to write because they require independent thinking, extensive research, and multiple viewings of the movie. Thus, these reviews are less likely to be affected by contemporary demand shocks. We therefore treat critic review measures as exogenous in this research. The F-statistics of IVs from the first-stage regressions of Equation 3 are all above 10, suggesting that our IVs are not weak. We provide the first-stage regression results in Table W7 in Web Appendix B. The fixed-effects regression results are also provided in Table W8 in Web Appendix B.

Empirical Findings

Table 6 shows the results of the two-stage least squares fixed-effects regressions of Equation 3 with the proposed IVs, where standard errors are clustered at the movie level. Column 8 shows the results from the fixed-effects regression of box office revenue. Column 0 shows a baseline model without any text-related variables. Due to the relatively high model fit, we find only a small change in the R^2 between the baseline model and other models (Columns 1 to 7), suggesting that text-related variables only add limited predictive value.[4]

We begin with the estimation results in Column 1, Table 6. Most estimated parameters are in line with expectations. For example, box office revenue is higher for movies that spent more on advertising (.054, $p < .01$) and movies that played in a greater number of theaters (.945, $p < .01$). For movie reviews, the positive and significant coefficients of LnCumCriticVal (.904, $p < .01$) and LnCumUserVal (.967, $p < .05$) suggest a persuasive effect of reviews from both groups through review valence. We do not find any significant effect for the variance of numeric ratings or for the measures related to the text valence, including the consistency in text valence between critics and users (ValConsistency(C, U)).

Does Topic Consistency Matter?

We now examine our focal variable of interest, the topic consistency between critic and user reviews (TopicConsistency(C, U)). The relationship between topic consistency and movie demand is significant and positive (2.724, $p < .01$), suggesting a positive main effect of topic consistency as shown in Column 1 of Table 6. Regarding the effect size, an increase of one standard deviation in the topic consistency for a movie is associated with a 4.6% increase in box office revenue (i.e., $2.724 \times .017 = .046$), all else being equal.

To gauge the importance of the topic consistency between the different review sources (i.e., critics and users), we further construct two within-group topic consistency variables (i.e., TopicConsistency(C, C), TopicConsistency(U, U)) using movie reviews from one source only. We present the regression results of Equation 3 by replacing TopicConsistency(C, U) with either TopicConsistency(C, C) or TopicConsistency(U, U) in Columns 2 and 3 of Table 6.[5] We find significant effects of the two within-group topic consistency measures when they are inserted independently. However, once we include all three topic consistency measures into the regression (Column 4 of Table 6), only TopicConsistency(C, U) remains significant (4.678, $p < .05$). For the sake of completeness, we also test the effect of the mean pairwise weighted similarity of all reviews (i.e., TopicConsistency(All)) and find the effect to be positive (5.720, $p < .05$; Column 5 of Table 6). Combining results from Columns 1 to 5 of Table 6, we conclude that the positive effect of the topic consistency between critics and users is the main driving force. That is, the between-group rather than within-group topic consistency is a significant and positive predictor of movie demand.

Moderators of the Effect of Topic Consistency

Having demonstrated the positive effect of topic consistency on movie demand, we further investigate the behavioral mechanism that may drive this effect. We contend that a large overlap in topics between different sources (i.e., critic and user reviews) might help consumers receive more focused cues along with the more salient movie attributes associated with the topic. These cues can then facilitate information recall when consumers make moviegoing decisions (Crowder 2015), which leads to higher ticket sales.

We indirectly test the information-recall mechanism of topic consistency by considering two potential moderators. The first moderator is the absolute gap in the average numeric ratings between cumulative critic and user reviews (Abs(ValenceGap)), which measures the difference in review valence between the two sources. Because cognitive response and attitude formulations for repeated messages depend on the convincingness of the message (Cacioppo and Petty 1980), consumers are more likely to attend to the overlap in review texts by critics and users when they also see similar review valence. When users and critics provide divergent numeric ratings but similar review content, it will be difficult for consumers to justify the similarity in review texts, which in turn might lower the convincingness of the repeated topics and consumers' ability to recall information. Following this reasoning, we expect the positive effect of topic consistency to be stronger when critics and users provide more similar numeric ratings for a movie (i.e., a smaller Abs(ValenceGap)).

We examine the moderating effect of Abs(ValenceGap) by including its interaction with TopicConsistency(C, U) in Equation 3, where both variables are mean-centered. We find a

[4] We investigate the predictive value of text-related variables further in Web Appendix A and present evidence that the predictive value of text-related variables is similar to the predictive value of rating-related variables.

[5] TopicConsistency(C, C) and TopicConsistency(U, U) are constructed based on the weighted cosine similarity measures. The weights are w_{idC} and w_{idU}, the average of topic weight vectors from a cumulative critic reviews and user reviews respectively. We took a similar approach to TopicConsistency(All), using all the reviews to generate the relative topic importance.

negative interaction effect of TopicConsistency(C, U) and Abs(ValenceGap) without endogeneity correction (−6.102, p < .05; Column 6 of Table W8 in Web Appendix B). With endogeneity correction, however, the effect becomes no longer significant (−5.777, p > .05; Column 6 of Table 6) although the effect sizes are similar. Thus, we do not find converging evidence of the information-recall mechanism. To further explore, we allow the interaction effect to vary by the sign of ValenceGap to capture the potential asymmetric interaction effect when critics' average numeric rating is either higher or lower than that of users. Specifically, we interact TopicConsistency(C, U) with both Pos(ValenceGap) and Neg(ValenceGap) and reestimate the model. Here, Pos(ValenceGap) equals Abs(ValenceGap) when the numeric rating of cumulative critic reviews is above (and Neg(ValenceGap) equals Abs(ValenceGap) when the numeric rating of cumulative critic reviews is below) the numeric rating of cumulative user reviews, and zero otherwise. Columns 7 and 8 of Table 6 show that the interaction effect is statistically significant negative only if critics evaluate the movie more favorably than users do either with (−9.830, p < .05) or without (−8.508, p < .05) endogeneity correction. Thus, we find converging evidence of asymmetric interaction effects between the gap of numeric ratings and topic consistency.

We also examine the moderating role of average review ratings of a movie. For movies with extreme ratings (either extremely high or extremely low), the strong cues from the numeric ratings might dampen the strength of cues associated with the repeated topics in review texts, which in turn lowers the information-recall benefit of topic consistency. It is also possible that consumers spend less time reading through the review content for movies with extreme ratings as they have already formed their attitude about the movie from the ratings. If so, we expect the positive effect of topic consistency to be weaker or even absent for movies with extreme ratings. To empirically test this hypothesis, we divide all movies into four buckets based on the mean of average ratings of cumulative critic reviews and average ratings of cumulative user reviews. We redo the fixed-effect estimation of Equation 3 using each of the four subsamples and present the results in Table 7. Consistent with Table 6, we report the baseline models (Columns 1 to 4) without text-related variables and full models (Columns 5 to 8) with text-related variables in Table 7.

We focus on the results in full models with topic consistency. Although the coefficient of topic consistency remains positive, it is statistically significant at the .05 level only if the average review rating is in the middle range (i.e., the bucket of [2.0, 3.0)). The effect size is also greater when the average review rating falls into this middle range than when it is more extreme. These results suggest that the midtier review ratings (between 2 and 3 on a 0–5 scale) seem to be one boundary condition of the positive effect of topic consistency.

Robustness Checks

We first check the robustness of our findings to an alternative measure of topic consistency. We used the bigram LDA model to construct topic consistency because of its advantage of extracting topics without human intervention. However, it is still possible that the topics identified by the LDA model do not fully coincide with the key topics perceived by actual moviegoers. As a robustness check, we reran the same analysis using a new topic consistency measure constructed by a fixed set of prominent topics identified by the movie literature. Following Gelper, Peres, and Eliashberg (2018), we consider nine topics commonly appearing in movie reviews: actor, director, the movie itself (e.g., storyline or filmmaking), trailer, critics' reviews, the genre, another movie, the movie listing, and "other." We then recruited participants on Amazon Mechanical Turk to classify each movie review into one of the nine topics. To keep the classification task tractable, we focused on a random subsample of 100 movies in this robustness check and asked human participants to classify a random sample of 5,000 movie reviews (including both critic and user reviews). To apply text classification, we preprocessed the text by removing stop words, punctuation, and numbers, tokenizing each word using a standard stemming algorithm, and extracting text features using term frequency–inverse document frequency. Given the labeled training sample, we then used the support vector machine to predict the key topic for the remaining movie review data set (F1 score = .70). We constructed a new topic consistency measure similar to what we did in the main analysis; the only difference is that the topic weight used in this new measure has a unit mass on one topic either labeled by human participants or predicted by the support vector machine, whereas the topic weight from the bigram LDA is a discrete distribution. We report the regression results with this new topic consistency measure in Column 1 of Table 8. The positive and significant coefficient of the new topic consistency measure suggests that our findings are not sensitive to the employment of topic modeling.

We assumed that consumers read all reviews when constructing text-based WOM measures in our main analyses. However, readers may not have read all posted reviews, and moviegoers' reading behavior can be another boundary condition. On Rotten Tomatoes, reviews are listed in the order of their posting date, with newer reviews listed toward the top. Thus, if moviegoers read only reviews on the first few pages, the text of older reviews would not have influenced the moviegoers' decisions. To check the sensitivity of our findings, we reestimated Equation 3 and compared results using topic consistency created from the most recent ten days of reviews versus reviews that were posted at least ten days earlier. We chose a sliding window of ten days because the number of critic reviews on the first page approximately equals the average number of critic reviews posted in the past ten days. As Columns 2 and 3 of Table 8 show, there is a positive and significant effect of topic consistency on box office revenue. However, topic consistency derived from older reviews is statistically insignificant, supporting our prior assumption that readers mostly read reviews displayed in the first few pages. These results suggest that moviegoers' review reading is a boundary condition for the topic consistency effect.

Table 7. Effect of Topic Consistency by Average Numeric Ratings.

	DV = LnBoxRev$_{i,t}$							
	Baseline Models				Full Models			
	(1) Ratings ∈[0, 2)	(2) Ratings ∈[2, 3)	(3) Ratings ∈[3, 4)	(4) Ratings ∈[4, 5]	(5) Ratings ∈[0, 2)	(6) Ratings ∈[2, 3)	(7) Ratings ∈[3, 4)	(8) Ratings ∈[4, 5]
Percentage of movies in each bucket	20.00%	29.73%	33.73%	16.53%	20.00%	29.73%	33.73%	16.53%
LnCumCriticVal$_{i,t-1}$	2.412	.765**	.722	13.467	1.707	.680**	.491	13.838
	(1.379)	(.338)	(.519)	(12.238)	(1.329)	(.312)	(.499)	(12.994)
LnCumUserVal$_{i,t-1}$	.762**	.799***	.635	1.776	.712**	.739***	.591	1.275
	(.311)	(.261)	(.329)	(2.195)	(.292)	(.270)	(.347)	(2.179)
LnCumCriticVol$_{i,t-1}$	−.083	−.015	−.103	.219	−.08	.047	.090	.112
	(.115)	(.121)	(.150)	(.184)	(.072)	(.064)	(.049)	(.076)
LnCumUserVol$_{i,t-1}$	−.052	.061	.089	.126	−.133	.032	−.076	.225
	(.072)	(.067)	(.051)	(.077)	(.116)	(.121)	(.146)	(.190)
VarCumCriticVal$_{i,t-1}$	−.354	.029	.018	.457	−.212	.032	.028	.453
	(.228)	(.073)	(.045)	(.558)	(.220)	(.065)	(.039)	(.581)
VarCumUserVal$_{i,t-1}$	.109	.032	.007	.143	.103*	.029	.004	.125
	(.060)	(.052)	(.041)	(.097)	(.061)	(.049)	(.040)	(.095)
Abs(ValenceGap)$_{i,t-1}$	−.164	.227	.027	1.264	−.093	.192	.095	.725
	(.280)	(.269)	(.271)	(1.945)	(.269)	(.256)	(.283)	(1.912)
TopicConsistency(C, U)$_{i,t-1}$					1.295	3.004***	.218	.897
					(1.173)	(1.147)	(.894)	(1.494)
ValConsistency(C, U)$_{i,t-1}$					−.043	.021	−.050	.050
					(.174)	(.176)	(.153)	(.229)
CumCriticTextVal$_{i,t-1}$					−.940***	.383	.329	−.691
					(.314)	(.263)	(.479)	(.771)
CumUserTextVal$_{i,t-1}$					.004	.267**	.288*	−.446
					(.131)	(.134)	(.158)	(.367)
VarCumCriticTextVal$_{i,t-1}$					.449	.692	−.507	−.924
					(.414)	(.418)	(.602)	(.757)
VarCumUserTextVal$_{i,t-1}$					.006	.198	.289**	−.486
					(.151)	(.138)	(.128)	(.302)
LnAdv$_{i,t-1}$	.039***	.031***	.025***	.044***	.040***	.031***	.025***	.044***
	(.011)	(.009)	(.007)	(.009)	(.011)	(.008)	(.007)	(.008)
LnTheaters$_{i,t}$	.864***	.828***	.789***	.767***	.871***	.828***	.789***	.768***
	(.021)	(.023)	(.014)	(.017)	(.021)	(.022)	(.014)	(.017)
NComp$_{i,t}$	−.005	.003	.010	.004	−.003	.004	.011	.004
	(.012)	(.007)	(.006)	(.009)	(.012)	(.007)	(.006)	(.009)
Movie fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year, month, day of week, movie age, holiday fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,153	10,909	13,097	6,603	7,153	10,909	13,097	6,603
R^2	.975	.979	.976	.975	.976	.980	.976	.976

$^{**}p < .05$, $^{***}p < .01$.
Notes: Standard error clustered by movies in parentheses.

Experiments

To validate our prior findings and complement the empirical study, we conducted experiments aimed at explicating the potential mechanism that drives the positive effect of topic consistency.

Design

We use a between-subjects design with two conditions: high topic consistency versus low topic consistency between critic and user reviews, while controlling for the review valence by selecting neutral reviews. To simulate the moviegoing decision

Table 8. Robustness Check Results.

	DV = $LnBoxRev_{i,t}$		
	(1)	(2)	(3)
$LnCumCriticVal_{i,t-1}$	1.345**	.763***	−.272
	(.633)	(.238)	(1.086)
$LnCumUserVal_{i,t-1}$	1.348***	−.231	1.455
	(.438)	(.375)	(1.554)
$LnCumCriticVol_{i,t-1}$	−.288	−.073	−.697
	(.214)	(.155)	(.906)
$LnCumUserVol_{i,t-1}$	.202***	.119*	.302
	(.075)	(.071)	(.333)
$VarCumCriticVal_{i,t-1}$	−.287***	−.023	.028
	(.086)	(.030)	(.122)
$VarCumUserVal_{i,t-1}$	.125	.033	.117
	(.064)	(.038)	(.146)
$Abs(ValenceGap)_{i,t-1}$	−.984***	.256	.091
	(.322)	(.208)	(1.222)
$TopicConsistency(C, U)_{i,t-1}$	9.878**	.565**	2.529
	(4.072)	(.238)	(1.879)
$ValConsistency(C, U)_{i,t-1}$	.547**	−.198*	−.406
	(.237)	(.112)	(.441)
$CumCriticTextVal_{i,t-1}$	.017	.051	.294
	(.641)	(.248)	(.643)
$CumUserTextVal_{i,t-1}$	−.629**	.240**	−.328
	(.269)	(.109)	(.517)
$VarCumCriticTextVal_{i,t-1}$	.261	.204	.756
	(.653)	(.295)	(.859)
$VarCumUserTextVal_{i,t-1}$	−.755***	.150	−.507
	(.264)	(.116)	(.615)
$LnAdv_{i,t-1}$	.050***	.148***	.529*
	(.009)	(.047)	(.314)
$LnTheaters_{i,t}$	.835***	.848***	1.029***
	(.019)	(.046)	(.239)
$NComp_{i,t}$	.002	.009*	.016
	(.008)	(.005)	(.011)
Movie fixed effects	Yes	Yes	Yes
Year, month, day of week, movie age, holiday fixed effects	Yes	Yes	Yes
Observations	5,045	37,762	31,408
R^2	.981	.962	.928

$^{**}p < .05$, $^{***}p < .01$.

Notes: In Model 1, we constructed the topic consistency using a set of predetermined topics. The sample size changes for Model 1 because we took a random sample of 100 movies to manually label the topics of the reviews. In Model 2 and Model 3, we constructed the topic consistency measure using reviews posted in the past 10 days and reviews posted more than 10 days ago. The sample size changes for Model 3 because the first 10 days of data do not have reviews 10 days old. Standard error clustered by movies in parentheses.

for a new movie, we selected a moderate movie in terms of box office sales and WOM measures. In particular, we removed movies with either relatively low (below the 75th percentile) or high (above the 25th percentile) volume for both critic and user reviews. We applied the same filtering criteria on average ratings of critic and user reviews and on the box office revenue. Then we dropped the movies with the difference in numeric ratings between critics and users larger than 1. After this filtering process, we randomly chose a movie, *Jack Ryan: Shadow Recruit*, for the experiment. The selected movie received 54% in the Tomatometer for critics and 53% in audience score on Rotten Tomatoes in May 2021.

At the beginning of the survey, we presented participants from both conditions with some basic movie information such as the title, description, and poster, as shown in Figure W1 in Web Appendix B. Next, we presented one critic review and one user review from each of the two conditions. Table 9 presents the reviews used in the two conditions with different degrees of topic consistency. We kept the user review the same and displayed a different critic review between the high and low topic consistency conditions. We utilized the estimated topic weights from the bigram LDA model to select the critic reviews that were associated with either the same topic as the chosen user review (high topic consistency) or with a different topic (low topic consistency). Both the critic review and user review in the high topic consistency condition mainly focused on main characters, whereas the critic review in the low topic consistency condition comments on the original source of the movie (e.g., novel). To control for the review valence, we chose relatively neutral reviews on the basis of the text sentiment scores. The pretest later confirmed the neutral sentiment. We also slightly modified the reviews to ensure that the review length and sentiment were similar between the two conditions.

After showing them a pair of reviews, we asked participants to rate their likelihood of watching this movie on a scale from 0 to 100, which is our main dependent variable. Next, we asked questions on the suggested mediator, memorability, to understand the potential behavioral mechanism. After measuring the memorability, we asked participants to rate the extent to which they agreed with the following three statements using a seven-point scale (1 = "Strongly disagree," and 7 = "Strongly agree") to measure the perceived topic consistency: (1) "The critic and the user are highlighting the same topic of the movie," (2) "The critic and the user are focusing on the same topic of the movie," and (3) "The critic and the user are commenting on the same topic of the movie." We also asked participants to rate the valence of each sentence separately using the seven-point scale. At the end of the survey, we asked subjects whether they had watched this movie, along with some other demographic variables.

Using the online panel Prolific, we recruited participants who were age 21 and over and resided in the United States. Participants were randomly assigned to one of the two conditions (high vs. low topic consistency). We had 361 participants (M_{age} = 35.74 years, SD = 10.97; 68.98% female) after removing 31 participants who had seen the movie before and nine outliers who took more than 15 minutes to complete the survey or did not spend much time on the review page. Of these 361 participants, 178 were assigned to the high topic consistency condition and 183 were assigned to the low topic consistency condition.

Table 9. Review Texts Across Experiment Conditions.

Condition	Critic Review	User Review
High topic consistency	Average review valence: 3.93 "Chris Pine brings a welcome boyishness and idealism to Ryan. He is the one saving grace, believable as a young CIA recruit. Perhaps the script could have provided just a little more levity. He seems to be clobbered with cliché by his director and screenwriters sometimes. The evildoers, it turns out, were right in his backyard."	Average review valence: 3.84 "Not bad, not great. Chris Pine carried the movie overall, a very effective actor. However, didn't feel much chemistry between Jack and his leading lady until near the midpoint of the film. Kevin Costner was slightly underused. None of the characters really did much for me but Ryan as played by Pine. Not the worst for a Saturday evening but could have been better."
Low topic consistency	Average review valence: 3.83 "If you were to take dashes of Bourne and Bond, season with Mission: Impossible 4 and pour the lot into a mixer, the result would look a lot like this watchable but derivative spy movie. This is the first Ryan adventure that isn't directly based on a Clancy novel and perhaps that explains why it seems to be working as well as it does … at least until it doesn't."	Average review valence: 3.94 "Not bad, not great. Chris Pine carried the movie overall, a very effective actor. However, didn't feel much chemistry between Jack and his leading lady until near the midpoint of the film. Kevin Costner was slightly underused. None of the characters really did much for me but Ryan as played by Pine. Not the worst for a Saturday evening but could have been better."

Table 10. Experiment Results.

	Critic–User Reviews			User 1–User 2 Reviews
	DV = Likelihood of Watching the Movie	DV = Memorability	DV = Likelihood of Watching the Movie	DV = Likelihood of Watching the Movie
High topic consistency dummy	5.101**	.406***	1.392	−1.960
	(2.519)	(.119)	(2.310)	(4.725)
Critic (User 1) review valence	1.927	.115*	.872	.509
	(1.355)	(.064)	(1.229)	(3.08)
User (User 2) review valence	8.602***	.212***	6.660***	8.766***
	(1.720)	(.081)	(1.567)	(3.563)
Memorability			9.145***	
			(1.008)	
Intercept	−2.800	.895***	−10.980	6.827
	(6.613)	(.309)	(6.036)	(12.435)
Adjusted R^2	.106	.068	.271	.058

*p < .10, **p < .05, ***p < .01.
Notes: Standard errors in parentheses.

Manipulation Check

We ran a separate pretest on 50 participants using Prolific to examine whether the manipulated topic consistency and the perceived topic consistency were aligned, and whether consumers perceived the chosen reviews as neutral reviews. Each participant was randomly assigned to either the high or low topic consistency condition (25 participants in each condition). After showing the two reviews to participants without source information (critic review presented as Review 1, user review presented as Review 2), we measured the perceived level of topic consistency by asking participants to rate their agreement with the same three statements as described in the main experiment using the seven-point Likert scales. The three scales loaded in one factor and were reliable (Cronbach's $\alpha = .925$). We denote the mean of these three scales as PTopicConsistency (perceived topic consistency). We also asked all the participants to rate the overall sentiment for each of the four movie reviews in Table 10 using the seven-point Likert scale (1 = "Strongly negative," and 7 = "Strongly positive"). We denote the outcome of these questions by Val_{H_CR}, Val_{L_CR}, Val_{H_UR}, and Val_{L_UR}, respectively, where H and L indicate the reviews from the high and low topic consistency conditions and CR and UR indicate critic and user reviews.

We find evidence that supports the alignment between manipulated and perceived topic consistency (mean($PTopicConsistency_H$) = 4.787, mean($PTopicConsistency_L$ = 3.907), $t = 2.214$, $p < .05$). We also find that the chosen reviews were indeed perceived as neutral by consumers. The t-test shows that the valence of critic reviews in the high and low consistency conditions is not significantly different, and they are perceived as neutral, near four on a seven-point scale ($Val_{H\_CR} = 3.960$, $Val_{L\_CR} = 4.080$, $t = -.360$, $p > .1$). A similar observation applies to user-side reviews ($Val_{H\_UR} = 3.68$, $Val_{L\_UR} = 3.68$, $t = .000$, $p > .1$).

Findings

In line with our findings from the pretest, participants in the high topic consistency condition were indeed more likely to agree that the two reviews were on the same topic than participants in the low topic consistency condition (mean($PTopicConsistency_H$) = 5.041, mean($PTopicConsistency_L$) = 4.233, t = 6. 669, p < .001) and that the topic consistency condition and the perceived topic consistency were aligned. The valence rating of critic and user reviews from the two conditions were overall not statistically different from one another (Val_{H_CR} = 3.933, Val_{L_CR} = 3.836, t = .862, p > .1; Val_{H_UR} = 3.832, Val_{L_UR} = 3.945, t = −1.317, p > .1). To examine the main effect of topic consistency, we regressed the likelihood of watching the movie on the high topic consistency dummy, critic review valence ratings, and user review valence ratings. The coefficient of the high topic consistency dummy is significantly positive (5.101, t = 2.025, p < .05; Column 1, Table 10), indicating that participants were more likely to see the movie after reading a pair of neutral user and critic reviews that focused on the same topic rather than on different topics,[6] which lends support to the causal link between topic consistency and demand.

Behavioral Mechanism

We further explore the potential behavioral mechanism by investigating whether the improved memorability of review content drives the positive main effect of topic consistency. To test this explanation, we asked participants to rate the extent to which they agreed that the reviews of the movie shown were memorable using a seven-point Likert scale at the end of the survey. We then examined whether memorability mediated the positive effect of topic consistency on participants' likelihood of watching the movie.

We conducted a mediation analysis with Hayes's PROCESS Model 4 (Hayes 2018). We report the results in Columns 2 and 3 in Table 10. As predicted, topic consistency is a significant positive predictor of memorability (.406, t = 3.399, p < .01; Column 2). Also, there was a significant indirect effect of topic consistency on the likelihood of watching the movie via memorability (3.709, SE = 1.289), and the 95% confidence interval (CI) of the indirect effect did not include zero (95% CI = [1.419, 6.403]). Therefore, these findings combined suggest that memorability mediates the effect of the topic consistency on the likelihood of watching the movie.

Additional Experiment Using a Pair of User Reviews

We conducted an additional experiment using a pair of reviews by User 1 and User 2. Similar to the main experiment design, we use a between-subject design with two conditions: high topic consistency (same topic) versus low topic consistency (different topic), and we control for review valence by using neutral reviews.

Table 11 shows the selected reviews. These reviews were pretested on 49 participants using Prolific (n_H = 24, n_L = 25). According to the same perceived topic consistency measure as used in the critic–user experiment, the participants in the high topic consistency condition showed a significantly higher perceived topic consistency than those in the low topic consistency condition (mean($PTopicConsistency_H$) = 5.57, mean($PTopicConsistency_L$) = 4.64, t = 2.67, p < .01). Also, these reviews are perceived as neutral (Val_{H_UR1} = 4.25, Val_{L_UR1} = 4.23, Val_{UR2} = 4.28).

We recruited 149 participants who were aged 21 and over and resided in the United States using Prolific. After removing 19 participants who had seen the movie before and three outliers, 127 participants remained (M_{age} = 37.75 years, SD = 11.19; 51.97% female). Of these participants, 66 were assigned to the high topic consistency condition and 61 were assigned to the low topic consistency condition. The participants in the high topic consistency condition showed a significantly higher perceived topic consistency than those in the low topic consistency condition (mean ($PTopicConsistency_H$) = 5.76, mean ($PTopicConsistency_L$) = 4.53, t = 6.29, p < .001). Also, the valences of these reviews are perceived as neutral, and the user reviews are not significantly different from each other (Val_{H_UR1} = 4.24, Val_{L_UR1} = 4.03, t = 1.294, p > .1; Val_{H_UR2} = 4.06, Val_{L_UR2} = 3.97, t = .621, p > .1). Consistent with the previous experiment using critic and user reviews, we regress the likelihood of watching the movie on the high topic consistency dummy, User 1 review valence ratings, and User 2 review valence ratings. The coefficient of the high topic consistency dummy was not significant (−1.960, t = −.415, p > .1; Column 4 in Table 10). Thus, our results indicate that the positive impact of topic consistency between critics and users matters more than the within-user topic consistency for movie demand.

Conclusions

In a new era of "perpetual contact," the interaction among data of different sources, channels, and modalities deserves further attention from marketing scholars (Berger et al. 2020), and this research contributes by deepening our current understanding of cross-message implications. By analyzing movie reviews from Rotten Tomatoes, we find that not only review valence and volume, but also the overlap in content that critics and users are discussing, are positively associated with movie demand. Moreover, this association is more prominent for movies with mediocre review ratings than for movies with extreme ratings. It is also stronger when the review ratings from critics are closer to those of users.

Managerial Implications

Our findings carry several valuable implications for managers. First, movie producers and marketing agencies should actively

[6] We also replicate the main effect with positive and negative review valence in Web Appendix C.

Table 11. Review Texts Across Experiment Conditions with User 1 and User 2.

Condition	User 1 Review	User 2 Review
High topic consistency	Average review valence: 4.24 "Not a bad attempt to reboot the franchise with original material. Pine does good job as Ryan. Costner plays a good supporting role. However, I don't think Knightley was the right actor for the role of Cathy. The chemistry between them in the book doesn't quite match this reboot. The storyline was ok to keep you engaged. Overall, not bad, not good, just there."	Average review valence: 4.06 "Not bad, not great. Chris Pine carried the movie overall, a very effective actor. However, didn't feel much chemistry between Jack and his leading lady until near the midpoint of the film. Kevin Costner was slightly underused. None of the characters really did much for me but Ryan as played by Pine. Not the worst for a Saturday evening but could have been better."
Low topic consistency	Average review valence: 4.03 "It might have been better to leave it as the original movie and as the book. This could have been named something else, and it would have been ok. I felt like the characters were somewhat bland and not that exciting. The whole movie was ok; some spectacular scenes stood out in the movie. Don't get me wrong, it wasn't the worst movie I've seen this year, but it just wasn't in my top 5. Still enjoyable."	Average review valence: 3.97 "Not bad, not great. Chris Pine carried the movie overall, a very effective actor. However, didn't feel much chemistry between Jack and his leading lady until near the midpoint of the film. Kevin Costner was slightly underused. None of the characters really did much for me but Ryan as played by Pine. Not the worst for a Saturday evening but could have been better."

listen to both professional critics' and general consumers' voices. To take advantage of the topic consistency effect, producers should invest in understanding the similarities and differences between critics' and everyday moviegoers' responses, engaging with both types of reviewers to see if there is some common ground between them. Successfully identifying such common ground can help producers better forecast performance and revenues. In addition, movie theaters should also actively monitor reviews generated by both critics and users and use our proposed topic consistency measure in predicting box office sales when optimizing the showing window for each movie.

Second, movie producers and advertisers should consider inducing a common topic or theme for critics and users to discuss, as we find that high topic consistency can boost movie revenues. Our empirical results show that an increase of one standard deviation in the topic consistency leads to a 4.63% ($.017 \times 2.724$) increase in box office revenue, all else being equal. The effect size is not negligible considering that an increase of one standard deviation in advertising leads to a 16.3% ($139/46 \times .054$) increase in box office revenue. Given that the effect size of topic consistency is comparable to about 30% of advertising ($4.63/16.3 = .284$), movie producers and advertisers should actively leverage the commonality of review content from critics and users as a part of their promotion strategy.

Although movie studios cannot directly affect users' and critics' conversations, they can influence topics that might appear in movie reviews through the narratives highlighted in their promotional activities, such as movie trailers, posters, blogs, and TV and online commercials. For example, a studio that intends to trigger more conversation about the special effects of the movie (Topic 16) can include more visually appealing scenes with special effects in their trailers to steer users' and critics' attention to this topic. Similarly, a studio that intends to trigger more conversation about the award-winning potential of the movie (Topic 14) can emphasize the cast's past recognition in posters, official blogs, and sponsored stories in the media. When considering possible common topics, producers and advertisers should also bear in mind that users and critics do not necessarily have to align with each other in terms of their opinions of the movie. Our experiment results suggest that as long as potential moviegoers perceive users and critics discussing the same topic, their willingness to watch the movie increases. To be clear, we do not expect potential moviegoers to read every single review, especially as consumers nowadays face vast quantities of such information. However, similar to the idea that the number and average ratings of movie reviews, respectively, constitute measurable proxies for WOM volume and valence (Dellarocas, Awad, and Zhang 2004), topic consistency can serve as a proxy for the overlap in the information that moviegoers gather from critics and users. As such, our findings are still meaningful even if consumers do not read all available reviews.

Finally, although we focus on two groups (i.e., critics and users), the concept of topic consistency can be readily generalized to multiple groups (e.g., experts, reviewers with status, general reviewers). Further, despite our focus on the movie industry in this research, the concept and the measurement of topic consistency can be extended to any other markets and products that both experts and general users commonly review. In this sense, managers in other industries (e.g., cosmetics, book publishing, hotel industry) should be mindful of the interactions between the professionals and the general users in their marketing promotion plan.

Limitations and Future Research

Like any study, our work is subject to certain limitations that offer promising avenues for future research. First, to stay relevant to actual movie reviews, we did not remove movie titles, actors, genres, or rating-related information in the topic modeling analysis. As a result, some of the identified topics are partially driven by textual references to the franchise, genres, or actor names. We admit that the topic consistency in this study is partially driven by these references, given the nature of identified topics. It would be interesting to explore the role of topic consistency further with these textual references removed. However, as the majority of identified topics are not specific to individual movies or actors, we believe our main findings are likely to hold. Second, our research focuses on movies, which are experiential products with high uncertainty about quality. Users often find it difficult to judge the quality of experiential products before purchase, and therefore tend to rely on others' opinions to make purchase decisions. It would be interesting to investigate whether the effect of topic consistency still holds for less experiential product categories (e.g., utilitarian products, the quality of which can be judged by consumers in a relatively objective manner). Third, we only considered critic and user reviews from Rotten Tomatoes, but it is likely that moviegoers visit other platforms (e.g., IMDb and YouTube) and have access to other information sources, such as an offline network of friends and families. It would be interesting to test whether our findings are generalizable to other platforms and information sources. Fourth, our mediation analysis in the experiment shows a considerable increase in the adjusted R^2 after incorporating memorability into the full equation. Aside from topic consistency, other aspects in the review or individual characteristics may affect the memorability of online reviews, which further increases movie demand. It would be interesting to explore what other factors could affect memorability in the online review setting, which we leave to future study. Finally, we operationalized the topic consistency metrics at the daily level without exhaustive knowledge of individual review-viewing behavior. Future research on review-viewing behavior would be an excellent complement to the current study.

Acknowledgments

The authors thank the *JM* review team for valuable feedback, as well as attendees of the 2020 Marketing Science Conference and NBS marketing seminar for comments.

Associate Editor

David Schweidel

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this work was supported by a grant from Nanyang Technological University (to E. Kim).

ORCID iDs

Xin (Shane) Wang https://orcid.org/0000-0003-1257-6931
Shijie Lu https://orcid.org/0000-0002-4180-6022

References

Alba, Joseph W., Wesley J. Hutchinson, and John G. Lynch (1991), "Memory and Decision Making," in *Handbook of Consumer Theory and Research*, Harold H. Kassarjian and Thomas S. Robertson, eds. Prentice Hall, 1–49.

Amblee, Naveen and Tung Bui (2007), "Freeware Downloads: An Empirical Investigation into the Impact of Expert and User Reviews on Demand for Digital Goods," *AMCIS 2007 Proceedings*, 21, https://aisel.aisnet.org/amcis2007/21.

Babić Rosario, Ana, Francesca Sotgiu, Kristine De Valck, and Tammo H.A. Bijmolt (2016), "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors," *Journal of Marketing Research*, 53 (3), 297–318.

Basuroy, Suman, Subimal Chatterjee, and Abraham S. Ravid (2003), "How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets," *Journal of Marketing*, 67 (4), 103–17.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1–25.

Blei, David M. and John D. Lafferty (2009), "Topic Models," in Text Mining: Classification, Clustering, and Applications, Ashok N. Srivastava and Mehran Sahami, eds. Chapman & Hall/CRC, 71–94.

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

Boatwright, Peter H., Wagner Kamakura, and Suman Basuroy (2007), "Reviewing the Reviewers: The Impact of Individual Film Critics on Box Office Performance," *Quantitative Marketing and Economics*, 5 (4), 401–25.

Burmester, Alexa B., Jan U. Becker, Harald J. van Heerde, and Michel Clement (2015), "Pre- and Post-Launch Effects of Publicity and Advertising on Sales," *International Journal of Research in Marketing*, 32 (4), 408–17.

Burtch, Gordon, Yili Hong, Ravi Bapna, and Vladas Griskevicius (2018), "Stimulating Online Reviews by Combining Financial Incentives and Social Norms," *Management Science*, 64 (5), 2065–82.

Buschken, Joachim and Greg M. Allenby (2016), "Sentence-Based Text Analysis for Customer Reviews," *Marketing Science*, 35 (6), 953–75.

Cacioppo, John T. and Richard E. Petty (1980), "Persuasiveness of Communications Is Affected by Exposure Frequency and Message Quality: A Theoretical and Empirical Analysis of Persisting Attitude Change," *Current Issues and Research in Advertising*, 3 (1), 97–122.

Chakravarty, Anindita, Yong Liu, and Tridib Mazumdar (2010), "The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-Release Movie Evaluation," *Journal of Interactive Marketing*, 24 (3), 185–97.

Chamary, J.V. (2018), "Why You Hated 'Star Wars: The Last Jedi' but Critics Loved It," *Forbes* (March 16), https://www.forbes.com/sites/jvchamary/2018/03/16/star-wars-last-jedi-science-movie-reviews/.

Chen, Yubo, Yong Liu, and Jurui Zhang (2012), "When Do Third-Party Product Reviews Affect Firm Value and What Can Firms Do? The Case of Media Critics and Professional Movie Reviews," *Journal of Marketing*, 76 (2), 116–34.

Chen, Yubo and Jinhong Xie (2008), "Online Consumer Reviews: Word-of-Mouth as a New Element of Marketing Communication Mix," *Management Science*, 54 (3), 477–91.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345–54.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29 (5), 944–57.

Clemons, Eric K., Guodong Gordon Gao, and Lorin M. Hitt (2006), "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems*, 23 (2), 149–71.

Cones, John W. (1992), *Film Finance & Distribution: A Dictionary of Terms*. Silman-James Press.

Crowder, Robert G. (2015), *Principles of Learning and Memory: Classic Edition*. Psychology Press.

Dellarocas, Chrysanthos, Neveen Awad, and Xiaoquan Zhang (2004), "Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning," *ICIS 2004 Proceedings*, 30, https://aisel.aisnet.org/icis2004/30.

Dellarocas, Chrysanthos and Ritu Narayan (2006), "A Statistical Measure of a Population's Propensity to Engage in Post-Purchase Online Word-of-Mouth," *Statistical Science*, 21 (2), 277–85.

Dellarocas, Chrysanthos, Xiaoquan Michael Zhang, and Neveen F. Awad (2007), "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, 21 (4), 23–45.

Deng, Tianjie (2020), "Investigating the Effects of Textual Reviews from Consumers and Critics on Movie Sales," *Online Information Review*, 44 (6), 1245–65.

Deng, Yipu, Jinyang Zheng, Warut Khern-am-nuai, and Karthik Natarajan Kannan (2021), "More Than the Quantity: The Value of Editorial Reviews for a UGC Platform," *Management Science*, 68 (9), 6355–7064.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do Online Reviews Matter?—An Empirical Investigation of Panel Data," *Decision Support Systems*, 45 (4), 1007–16.

Eliashberg, Jehoshua and Steven Shugan (1997), "Film Critics: Influencers or Predictors?" *Journal of Marketing*, 61 (2), 68–78.

Faber, Ronald J. and Thomas C. O'Guinn (1984), "Effect of Media Advertising and Other Sources on Movie Selection," *Journalism Quarterly*, 61 (2), 371–77.

Gardial, Sarah Fisher and Gabriel J. Biehal (1985), "Memory Accessibility and Task Involvement as Factors in Choice," in *Advances in Consumer Research*, Vol. 12, Elizabeth C. Hirschman and Moris B. Holbrook, eds. Association for Consumer Research, 414–19.

Gelper, Sarah, Renana Peres, and Jehoshua Eliashberg (2018), "Talk Bursts: The Role of Spikes in Prerelease Word-of-Mouth Dynamics," *Journal of Marketing Research*, 55 (6), 801–17.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.

Godes, David and José C. Silva (2012), "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science*, 31 (3), 448–73.

Handel, Leo A. (1950), *Hollywood Looks at Its Audience*. University of Illinois Press.

Hayes, Andrew F. (2018), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, 2nd ed. Guilford Press.

Hennig-Thurau, Thorsten and Mark B. Houston (2019), *Entertainment Science: Data Analytics and Practical Theory for Movies, Games, Books, and Music*. Springer.

Hennig-Thurau, Thorsten, Caroline Wiertz, and Fabian Feldhaus (2015), "Does Twitter Matter? The Impact of Microblogging Word of Mouth on Consumers' Adoption of New Movies," *Journal of the Academy of Marketing Science*, 43 (3), 375–94.

Herhausen, Dennis, Stephan Ludwig, Dhruv Grewal, Jochen Wulf, and Marcus Schoegel (2019), "Detecting, Preventing, and Mitigating Online Firestorms in Brand Communities," *Journal of Marketing*, 83 (3), 1–21.

Holbrook, Morris B. (1999), "Popular Appeal Versus Expert Judgments of Motion Pictures," *Journal of Consumer Research*, 26 (2), 144–15.

Holbrook, Morris B. and Michela Addis (2007), "Taste Versus the Market: An Extension of Research on the Consumption of Popular Culture," *Journal of Consumer Research*, 34 (3), 415–24.

Homburg, Christian, Laura Ehm, and Martin Artz (2015), "Measuring and Managing Consumer Sentiment in an Online Community Environment," *Journal of Marketing Research*, 52 (5), 629–41.

Jansen, Bernard J., Mimi Zhang, Kate Sobel, and Abdur Chowdury (2009), "Twitter Power: Tweets as Electronic Word of Mouth," *Journal of the American Society for Information Science and Technology*, 60 (11), 2169–88.

Lau, Jey Han, Timothy Baldwin, and David Newman (2013), "On Collocations and Topic Models," *ACM Transactions on Speech and Language Processing*, 10 (3), 1–14.

Lee, Thomas Y. and Eric T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48 (5), 881–94.

Li, Mengxiang, Liqiang Huang, Chuan-Hoo Tan, and Kwok-Kee Wei (2013), "Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features," *International Journal of Electronic Commerce*, 17 (4), 101–36.

Litwak, Mark (1986), *Reel Power: The Struggle for Influence and Success in the New Hollywood*. William Morrow.

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (3), 74–89.

Liu, Xiao, Dokyun Lee, and Kannan Srinivasan (2019), "Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning," *Journal of Marketing Research*, 56 (6), 918–43.

Liu, Angela Xia, Jan-Benedict E.M. Steenkamp, and Jurui Zhang (2018), "Agglomeration as a Driver of the Volume of Electronic Word of Mouth in the Restaurant Industry," *Journal of Marketing Research*, 55 (4), 507–23.

Liu, Xiao, Param Vir Singh, and Kannan Srinivasan (2016), "A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing," *Marketing Science*, 35 (3), 363–88.

Lord, Kenneth R. and Sanjay Putrevu (1998), "Communicating in Print: A Comparison of Consumer Responses to Different Promotional Formats," *Journal of Current Issues & Research in Advertising*, 20 (2), 1–18.

Lu, Shijie, Xin (Shane) Wang, and Neil Bendle (2020), "Does Piracy Create Online Word of Mouth? An Empirical Analysis in the Movie Industry," *Management Science*, 66 (5), 2140–62.

Ma, Liye, Baohong Sun, and Sunder Kekre (2015), "The Squeaky Wheel Gets the Grease: An Empirical Analysis of Customer Voice and Firm Intervention on Twitter," *Marketing Science*, 34 (5), 627–45.

Mankad, Shawn, Hyunjeong "Spring" Han, Joel Goh, and Srinagesh Gavirneni (2016), "Understanding Online Hotel Reviews Through Automated Text Analysis," *Service Science*, 8 (2), 124–38.

Moe, Wendy W. and David A. Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science*, 31 (3), 372–86.

Moe, Wendy W. and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (3), 444–56.

Moon, Sangkil, Paul K. Bergey, and Dawn Iacobucci (2010), "Dynamic Effects Among Movie Ratings, Movie Revenues, and Viewer Satisfaction," *Journal of Marketing*, 74 (1), 108–21.

Puranam, Dinesh, Vishal Narayan, and Vrinda Kadiyali (2017), "The Effect of Calorie Posting Regulation on Consumer Opinion: A Flexible Latent Dirichlet Allocation Model with Informative Priors," *Marketing Science*, 36 (5), 726–46.

Reinstein, David A. and Christopher M. Snyder (2005), "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics," *Journal of Industrial Economics*, 53 (1), 27–51.

Rossi, Peter E. and (2014), "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," *Marketing Science*, 33 (5), 655–72.

Ryoo, Jun Hyun (Joseph), Xin (Shane) Wang, and Shijie Lu (2021), "Do Spoilers Really Spoil? Using Topic Modeling to Measure the Effect of Spoiler Reviews on Box Office Revenue," *Journal of Marketing*, 85 (2), 70–88.

Schofield, Alexandra, Måns Magnusson, and David Mimno (2017), "Pulling Out the Stops: Rethinking Stopword Removal for Topic Models," in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Association for Computational Linguistics, 432–36.

Schweidel, David A. and Wendy W. Moe (2014), "Listening In on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research*, 51 (4), 387–402.

Sievert, Carson and Kenneth Shirley (2014), "LDAvis: A Method for Visualizing and Interpreting Topics," in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. Association for Computational Linguistics, 63–70.

Stammerjohan, Claire, Charles M. Wood, Yuhmiin Chang, and Esther Thorson (2005), "An Empirical Investigation of the Interaction Between Publicity, Advertising, and Previous Brand Attitudes and Knowledge," *Journal of Advertising*, 34 (4), 55–67.

Statista (2018), "Share of Adults Who Read Reviews Before Watching a Movie in the United States as of August 2018," https://www.statista.com/statistics/898999/reading-reviews-before-viewing-movies-united-states/.

Sun, Monic (2012), "How Does the Variance of Product Ratings Matter?" *Management Science*, 58 (4), 696–707.

Timoshenko, Artem and John R. Hauser (2019), "Identifying Customer Needs from User-Generated Content," *Marketing Science*, 38 (1), 1–20.

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), 463–79.

Toubia, Olivier, Garud Iyengar, Renée Bunnell, and Alain Lemaire (2019), "Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption," *Journal of Marketing Research*, 56 (1), 18–36.

Wang, Yang and Alexander Chaudhry (2018), "When and How Managers' Responses to Online Reviews Affect Subsequent Reviews," *Journal of Marketing Research*, 55 (2), 163–77.

Wang, Feng, Xuefeng Liu, and Eric Er Fang (2015), "User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth Effects," *Journal of Retailing*, 91 (3), 372–89.

Youngs, Ian (2017), "Star Wars: The Last Jedi—the Most Divisive Film Ever?" BBC News (December 20), https://www.bbc.com/news/entertainment-arts-42424445.

Zhang, Xiaoquan and Chrysanthos Dellarocas (2006), "The Lord of the Ratings: Is a Movie's Fate Is Influenced by Reviews?" *ICIS 2006 Proceedings*, 117, https://aisel.aisnet.org/icis2006/117.

Zhong, Ning and David A. Schweidel (2020), "Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model," *Marketing Science*, 39 (4), 827–46.

Zhou, Wenqi and Wenjing Duan (2016), "Do Professional Reviews Affect Online User Choices Through User Reviews? An Empirical Study," *Journal of Management Information Systems*, 33 (1), 202–28.