

What Reviews Foretell about Opening Weekend Box Office Revenue: The Harbinger of Failure Effect in the Movie Industry

Pantelis Loupos^{1*}, Yvette Peng¹, Sute Li¹ and Hao Hao¹

^{1*}Graduate School of Management, University of California Davis, 540 Alumni Ln, Davis, 95616, CA, United States.

*Corresponding author(s). E-mail(s): ploupos@ucdavis.edu;
Contributing authors: ympeng@ucdavis.edu; sueli@ucdavis.edu;
hahao@ucdavis.edu;

Abstract

We empirically investigate the harbinger of failure phenomenon in the motion picture industry by analyzing the pre-release reviews written on movies by film critics. We find that harbingers of failure do exist. Their positive (negative) pre-release movie reviews provide a strong predictive signal that the movie will turn out to be a flop (success). This signal persists even for the top critic category, which usually consists of professional critics, indicating that having expertise in a professional domain does not necessarily lead to correct predictions. Our findings challenge the current belief that positive reviews always help enhance box office revenue and shed new light on the influencer-predictor hypothesis. We further analyze the writing style of harbingers and provide new insights into their personality traits and cognitive biases.

Keywords: Movies, Reviews, Harbingers of Failure, Text Analytics, Preference Heterogeneity

1 Introduction

Reviews play a significant role in various industries and business settings. In retailing, for example, reviews can be directly related to the ultimate success or failure of a product (Ho-Dac et al, 2013; Chevalier and Mayzlin, 2006). In the motion picture industry, it is hypothesized that movie critics can play a dual role. They can act as influencers when their reviews influence consumers' movie-viewing behavior, and/or they can act as predictors when they merely predict it. Previous research on this question produced mixed conclusions. Eliashberg and Shugan (1997) find that critics can correctly predict box office performance but do not influence it. Basuroy et al (2003), on the other hand, find that critics can also influence the outcome of a movie during the early stages of a film's run. In this work, we offer new findings to the influencer-predictor hypothesis by empirically investigating this question in a pre-release movie setting and demonstrate that a distinction should be made among film critics.

Pre-release screenings are a common practice in the motion picture industry. Producers and distributors invite film critics in order to test the wider audience appeal of the film. The film critics then publish their pre-release reviews in mediums like Rotten Tomatoes so people can read the critics' opinions about the quality of the upcoming movie. However, critics' opinion is not always in alignment with the public taste as anecdotal evidence suggests¹. The movies *Baywatch* and *Tomb Raider*, for example, received positive pre-release reviews but generated dismal opening box office results. On a seemingly unrelated research topic, Anderson et al (2015) demonstrated the existence of certain customers in the retail industry, called harbingers of failure, whose early adoption of a new product is a strong predictor that the product will fail in the market. Motivated by this finding, we empirically investigate in this work whether the harbinger phenomenon extends to the motion picture industry. Does the positive feedback given by some film critics before the release of a movie signal its financial failure instead of its future success as currently believed?

To address this question, we use pre-release movie reviews from Rotten Tomatoes in combination with other movie-related data collected from different data sources. Our empirical findings document the existence of harbingers of failure in the motion picture industry. We identify these harbingers and show that they systematically misclassify flop movies as successful. Their positive pre-release reviews of a movie provide a strong predictive signal that the movie will fail financially in its opening weekend box office revenue. This finding persists even when we account for sleeper movies. Sleeper movies are critically acclaimed movies, such as *The Shawshank Redemption* and *Children of Men*, that tank in their opening weekend box office revenue but eventually become successful. Moreover, we find the harbinger effect to be symmetric. That is, there are harbinger critics that review a movie negatively and then the movie

¹<https://vtropes.org/pmwiki/pmwiki.php/Main/AcclaimedFlop>

turns out to be successful. We then investigate if the harbinger phenomenon still holds true for the top critic category, which usually consists of professional and experienced critics. Surprisingly, we find a strong harbinger effect indicating that having expertise in a professional domain does not necessarily lead to correct predictions.

To further gain insights into the personality traits of harbinger film critics, we use text analytics to analyze their writing style and compare it with that of non-harbinger critics. We find that harbinger critics employ a formal and analytical style of writing and have a lower rate of self-reference pronouns compared to non-harbingers. This indicates that harbinger critics show overconfidence in their abilities, while non-harbinger critics are more self-reflective about the audience's opinion and convey more confidence in their reviews. Interestingly, we find that these differences persist when we stratify the analysis by the critic's status category. In addition, we find that top critic harbingers are also using more adverbs than their counterparts in their positive reviews, a trait that is associated with extraversion and over-optimism (Tausczik and Pennebaker, 2010a).

To sum up, the main contributions of this work are as follows. First, we document the harbinger effect in the motion picture industry. This documentation sheds new light on the influencer-predictor hypothesis (Eliashberg and Shugan, 1997). We find that, in a pre-release setting, a distinction should be made among film critics. In particular, there exist harbinger critics who do not influence the outcome of a movie but rather mispredict it. Second, we add to the existing literature on harbingers with the fact that the harbinger phenomenon is not only practiced by everyday consumers, but also by professionals with expertise domain knowledge. Last, we analyze the writing style of harbingers and provide insights into their personality traits. Previous research has used transactional data to provide correlational evidence that the mechanism behind the harbinger phenomenon is preference minorities (Waldfogel, 1999). In our case, we have text data, a much richer source of data that we use to offer evidence about the intrinsic personality characteristics of harbingers that drive their choices.

2 Conceptual Framework

2.1 Harbingers of Failure in the Motion Picture Industry

Several works have explored the effect that movie reviews have on box office revenue (Reddy et al, 1998; Basuroy et al, 2006; Boatwright et al, 2007; Delarocas et al, 2007; Duan et al, 2008; Chintagunta et al, 2010; Song et al, 2019). Litman (1983) suggested that movie critics' positive reviews can influence the popularity of a movie in the early weeks of its release, but never tested this hypothesis due to the absence of opening box office data. Eliashberg and Shugan (1997) first proposed and tested the two different roles that movie critics can play, namely influencers and predictors. They find that movie critics can predict the outcome of a movie but do not influence it. Basuroy

et al (2003) break down this hypothesis across the stages of a movie's lifetime and find that movie critics' opinions can have an influencer effect in the early stages of a film's run before the word-of-mouth mechanism emerges. Moreover, movie critics' reviews are predictive of the box office outcome on the entire run of the movie. Both their positive and negative reviews are significantly correlated with box office revenue, with the impact of negative reviews (but not positive reviews) diminishing over time.

Anderson et al (2015) first documented the existence of harbingers of failure. These are consumers whose early endorsement of a new product is a strong signal that mainstream adoption of the product will not follow. They also find the harbinger effect to be symmetric. That is, there are customers who tend to avoid buying successful products. Simester et al (2019) later documented the existence of the harbinger phenomenon in political campaign donations and zip codes. Both of these works study groups of customers who do not influence the decisions of others. Our work diverges and expands on this subject by investigating how the critics' reviews could influence the general audience in a pre-release movie setting.

Combining the above literature, we expect to find harbingers of failure in the motion picture industry, such that their positive pre-release movie reviews will signal the movie's failure at the opening box office, while their negative pre-release movie reviews will signal the movie's success at the opening box office.

2.2 Writing Style and Personality Traits

Previous research has shown that the writing style of a person reflects the person's personality (Pennebaker and King, 1999; Küfner et al, 2010). There are three different writing styles, namely formal, analytic, and narrative, each associated with different personality traits and cognitive biases. In what follows, we develop our hypotheses on the potential writing style of harbinger and non-harbinger critics in the motion picture industry and link each writing style with their potential cognitive biases.

2.2.1 Extraversion

Extraversion is a major trait in the most prevalent theories of personality, such as Eysenck's three-factor model (Eysenck, 1991) and Costa and McCrae's five factor model (Costa Jr and McCrae, 1992). Optimism bias is linked to extraversion and is the tendency of people to place a greater probability in experiencing positive events while underestimating the probability of negative events. Studies report that a large majority of the population display optimism bias across many different settings (Sharot, 2011). Moreover, people maintain their overly positive expectations even in the face of disconfirming evidence because they tend to update their beliefs more in response to positive than negative information about the future. On the basis of these findings, we hypothesize that harbinger movie critics will display optimism bias in their

positive pre-release movie reviews. In terms of writing style, extravert people tend to use fewer adjectives and more adverbs (Tausczik and Pennebaker, 2010b). Hence, we expect harbinger critics to use fewer adjectives and more adverbs in their positive reviews compared to non-harbinger critics in order to express their aspects of optimism. We propose the following hypothesis:

Hypothesis 1: *Harbinger movie critics will use fewer adjectives and more adverbs in their positive reviews compared to non-harbingers to express their aspects of overoptimism.*

2.2.2 Conscientiousness

Conscientiousness is another major personality trait. Conscientious people tend to be self-reflective, efficient, and organized (Barrick and Mount, 1991). Being a conscientious movie critic can be rewarding since the critic can put aside personal biases and reflect on how the audience will perceive a movie. Hence, we expect non-harbinger critics to be conscientious. In contrast, we hypothesize that harbinger critics will be more concerned with status and power, and be less self-reflective. Writers with these attributes tend to employ a formal writing style (Pennebaker et al, 2003). Thus, we expect harbinger critics to employ a more formal writing style compared to non-harbingers. The formal writing style is characterized by an impersonal, objective, and precise use of language, and is associated with a lower rate of self-reference pronouns, fewer fillers, and fewer nonfluencies (Davis and Brock, 1975). It also includes a higher rate of uncommon words and hyphens. We propose the following hypothesis:

Hypothesis 2: *Non-harbinger movie critics are more conscientious and self-reflective about general audience preferences compared to harbinger critics.*

2.2.3 Overconfidence

Overconfidence bias is the tendency of people to overestimate their abilities relative to others (West and Stanovich, 1997). In the case of movie reviewers, this means that harbingers will place much more weight on their own opinion and believe that the general audience will agree with them. People who tend to be overconfident generally use I-words at very low levels. Hence, we expect harbinger critics to have a lower rate of self-reference pronouns, such as I, me, and my. In contrast, we expect non-harbingers to have a higher rate of first-person singular pronouns to convey more confidence in their reviews and a lower rate of first-person plural pronouns such as we, us, and our (Slatcher et al, 2007). High usage of we-words in movie reviews can be perceived by the audience as cold and not authentic. In summary, we propose the following set of hypotheses:

Hypothesis 3: *Harbinger movie critics are more overconfident in their positive reviews than non-harbinger critics.*

Hypothesis 4: *Non-harbinger movie critics are conveying more confidence in their reviews compared to harbinger critics.*

3 Overview of Data and Methods

3.1 Data Overview

To conduct this research, we combined data sets from three different data sources. First, we collected movie-related variables including movie title, genre, studio, whether the movie is a sequel, release date, opening weekend revenue, and gross domestic revenue from Box Office Mojo ². Second, we augmented our movie-related information by collecting the movie budgets from The Numbers ³. We use this information to define our movie success metrics. Last, we obtained our review-related information from Rotten Tomatoes⁴.

Rotten Tomatoes is an aggregator review website for movies and television shows. It aggregates reviews across different mediums (e.g., blog posts, websites, etc.) in order to increase their usefulness. The company was launched in 1998 by three undergraduate students at the University of California, Berkeley, and it is now owned by the American ticketing company Fandango. It attracts around 26 million unique monthly visitors globally, out of which 14 million are in the U.S. To become a critic for Rotten Tomatoes you have to go through a vetting process that occurs twice a year, and you need to produce consistent review output for at least two years. The critics must display journalistic integrity and ethical behavior in their reviews. Rotten Tomatoes also has a top critic designation that is given to selected candidates after careful evaluation by their advisory committee. The requirements to become a top critic are stringent. For example, you need to have written reviews professionally for a minimum of five years and you also need to have a verifiable social media profile with a large audience. The long-term tenure of critics in Rotten Tomatoes is crucial for our investigation of harbingers, since in order to show their repeated tendency to misclassify flop movies we need access to their long-term review data. Our review data include the critic's name, the date that the review was posted on Rotten Tomatoes, whether the critic is a top critic or not, the review's text, and the Fresh Or Rotten rating. The latter variable is worth further explaining. Since Rotten Tomatoes aggregates reviews across different mediums, these reviews come with different grading scales. For example, some critics use a 4-point scale, some use a 10-point scale and others provide a percentage. To avoid comparing grades from across different scales, a team of curators at Rotten Tomatoes reads the reviews and classifies them as Fresh if they are positive, or Rotten if they are negative. Therefore, we also use the Fresh Or Rotten rating to classify our reviews into positive or negative in our analysis. After merging all the aforementioned data, we end up with 448 movies released from 2015 to 2019, and 28,884 pre-release reviews written about them. A detailed overview of our final dataset and of all the variables used in our models can be found in Appendix A.

²<https://www.boxofficemojo.com>

³<https://www.the-numbers.com>

⁴<https://www.rottentomatoes.com>

3.2 Methodology

3.2.1 Defining whether a movie is a success or a flop

We first need to classify a movie as being successful or not. After consulting with movie industry experts, we classify a movie as successful if its Return on Investment (ROI) is greater than 0.5, where the ROI is calculated by dividing the opening weekend revenue of the movie over its production budget. In other words, a successful movie earns more than half of its production budget in the opening weekend. More formally, we have that:

$$\text{Successful Movie} = \begin{cases} 1 & \text{if ROI} \geq 0.5 \\ 0 & \text{o.w.} \end{cases}$$

where

$$\text{ROI} = \frac{\text{Opening Weekend Revenue}}{\text{Production Budget}}$$

Overall, 44.86% of the movies in our dataset are successful.

3.2.2 Defining whether a movie critic is a harbinger

Following the methodology of [Anderson et al \(2015\)](#), we adopt the flop affinity ratio to measure the propensity of a movie critic to review a flop movie positively. For each movie critic i , we calculate the proportion of positively reviewed movies that turned out to be flops over their total number of positively reviewed movies. That is:

$$\text{Flop Affinity}_i = \frac{\text{Number of Flop Movies Reviewed Positively}}{\text{Number of Movies Reviewed Positively}}$$

For example, if a movie critic gave positive reviews to ten movies before their release and eight of them turned out to be flops, then his flop affinity ratio is 0.8. Critics with a low flop affinity rate indicate that their movie opinion coincides with public taste. On the other hand, a high flop affinity ratio implies that the critic is likely to be a harbinger of failure. Note here that the flop affinity ratio does not give a higher weight to critics with a higher total number of reviews. For example, there might be two critics with 0.5 flop affinity ratio, but one critic has written four reviews while the other critic has written twenty reviews. To this end, we also use an alternative grouping method as a robustness check. In this approach, we weigh the critic's number of positive/negative reviews with regard to their overall number of positive/negative reviews. The pattern of the results under this new grouping method remains unchanged. For more details, see [Appendix B](#).

We then proceed to classify movie critics into five groups based on their flop affinity score. Group 1 represents the cohort with the lowest flop affinity scores, while Group 4 is on the other end of the spectrum. Last, critics who reviewed no more than two movies before their release are placed in the Other group. Since we cannot judge the harbinger effect of critics with such a limited

number of written reviews, we exclude them from our analysis. Table 1 shows all the relevant information on the critics' grouping.

Table 1 Movie critics grouping based on their flop affinity score

Group	Critics Distribution		
	Flop Affinity Range	Percentile Cutoff	No. of Critics
Group 1	[0, 40%]	25th percentile	162
Group 2	(40%, 50%]	50th percentile	145
Group 3	(50%, 62%]	75th percentile	152
Group 4	(62%, 100%]		153
Other (≤ 2 positive reviews)			226

3.3 Modeling

To investigate whether the harbinger effect exists in the movie industry, we estimate two competing logistic regression models. More specifically, we have the following models:

$$\text{Model 1: } \log(\text{Odds of Success of Movie } i) = \alpha + \beta_0 \text{TotalPositiveReviews}_i \quad (1)$$

$$\begin{aligned} \text{Model 2: } \log(\text{Odds of Success of Movie } i) = & \alpha + \beta_1 \text{Group1PositiveReviews}_i + \\ & \beta_2 \text{Group2PositiveReviews}_i + \beta_3 \text{Group3PositiveReviews}_i \\ & + \beta_4 \text{Group4PositiveReviews}_i \end{aligned} \quad (2)$$

The unit of analysis for both models is a movie denoted by i . The dependent variable is the log odds of success of movie i as defined in 3.2.1. The key difference between the two competing models is in their independent variables. The *TotalPositiveReviews* in Model 1 refers to the total number of positive reviews on movie i , and under this specification, all critics are treated equally. However, Model 2 incorporates the harbinger effect by including all four groups of critics. For example, the variable *Group1PositiveReviews_i* counts the number of positive reviews posted by the critics in Group 1 for movie i .

3.4 Model Evaluation

To evaluate our models, we split our data into two random non-overlapping sets across movies: the classification and the prediction set. The classification set contains 286 movies released from 2015 to 2017, out of which 111 (38.8%) are successful, while the prediction set contains 162 movies released from 2018 to 2019, out of which 90 (55.5%) are successful. Note here that both sets are

balanced. We use the classification set to get our model estimates. Holding the movie critic groups stable, we then use the aforementioned model estimates to examine whether we can accurately predict the outcome of movies in the prediction set. The fact that our movie critic groups do not change over time ensures that if a harbinger effect indeed exists, then that will mean that harbinger critics repeatedly get it wrong.

To measure the performance of our models in the prediction set, we report the AUCPR score. The primary reason we chose to report this metric is that AUCPR does not care about true negatives, and is much more sensitive to true positives, false positives, and false negatives than AUC. In our case, a true negative corresponds to the case where a movie critic predicted that a movie would be a flop and turned out to be a flop; hence, it is not relevant to the harbinger effect. We also report the F1 score to evaluate the robustness of the findings.

4 Empirical Results

4.1 Harbinger Effect

We present the results of our models in Table 2. The coefficients of interest in equations 1 and 2 are estimated using the 285 movies in the classification set. We observe that Model 2 has a significant improvement over Model 1 in their goodness of fit, as shown by the chi-square test on the likelihood ratio of the two models. Moreover, the coefficient estimates of Group 1 and Group 4 in Model 2 turn out to be statistically significant. This means that reviews from these two critic groups are informative regarding the financial outcome of the movie. However, group 1 has positive marginal effects, while group 4 has negative, indicating the existence of harbingers of failure among the movie critics. More specifically, we see that the expected change in log odds for a one-unit increase in the reviews of Group 1 is 0.59. Equivalently, a one-unit increase in the reviews of Group 1 increases the odds of a movie being successful by 80%. On the other hand, the expected change in log odds for a one-unit increase in the reviews of Group 4 is -0.47. In other words, a one-unit increase in the reviews of Group 4 decreases the odds of success by about 37%. We should note here that the addition of control variables as shown in Appendix A does not change the results.

Table 2 also shows the performance of our models in the prediction set. The AUCPR of Model 2 is .70 compared to 0.56 of Model 1, a relative improvement of 25%. Moreover, we also get an outstanding boost on the F1 score from 0.22 to 0.57, a relative improvement of 159%. As a robustness check, we replicate the analysis to account for sleeper movies and report the complete findings in Appendix C. The pattern of findings is unchanged. Furthermore, the coefficients of Group 1 and Group 4 are now larger in absolute size indicating a stronger harbinger effect. The expected change in log odds for a one-unit increase in the reviews of Group 4 is now -0.57, or equivalently it decreases the odds of success by about 43%.

Last, we also investigate whether the harbinger effect is symmetric in the case of movies and find strong evidence that this is the case. More specifically, there are film critics that review a movie negatively and then the movie turns out to be successful. We report the complete findings in Appendix D. We observe in Table D7 that both the coefficient estimates of Group 1 and Group 4 in Model D2 turn out to be statistically significant. Group 1 in this case has negative marginal effects, while group 4 has positive. The latter indicates that the more negative reviews given by high-success avoidance critics, the more likely it is that the movie will be a success. Specifically, we see that the expected change in log odds for a one-unit increase in the negative reviews of Group 4 is 0.75. In other words, a one-unit increase in the reviews of Group 4 increases the odds by 111% for the movie being successful.

Table 2 The table reports the coefficients of interest from Equations 1 and 2. Standard errors are in parentheses.

Logistic Models Results Using Flop Affinity Grouping Method		
	Model 1	Model 2
Total positive reviews	-0.008 (.002)	
Group 1 positive reviews		.59 (.104)***
Group 2 positive reviews		.04 (.057)
Group 3 positive reviews		-.05 (.041)
Group 4 positive reviews		-.47 (.089)***
Log-likelihood	-197.4	-111.8
Likelihood ratio test, chi-square (d.f.=3)		171.3***
AUCPR	.56	.70
F1	.22	.57

* p-value $\leq .1$, ** p-value $\leq .05$, *** p-value $\leq .01$

4.2 Writing Style of Harbingers

We use here the Linguistic Inquiry and Word Count (LIWC) text analytics software (Pennebaker et al, 2007) to investigate if there are any differences in the writing style between harbinger and non-harbinger movie critics. The comparison revealed some interesting differences in their writing styles of positive and negative reviews, which are summarized in Table 3.

First, with regard to positive reviews, we find that harbinger movie critics exhibit a much lower rate of self-reference pronouns compared to non-harbinger critics. This indicates that harbinger critics show overconfidence in their abilities, while non-harbinger critics convey more confidence and leadership through their reviews providing support for our Hypotheses 3 and 4, respectively. Moreover, harbinger critics have a very low rate of fillers, which in combination with the low rate of self-reference pronouns, indicates that they tend to employ a formal writing style. This provides support for our hypothesis 2 that harbinger critics are more concerned with status and power, and are less self-reflective about the audience’s opinion. With respect to negative reviews, we observe an additional emerging pattern. Harbinger critics tend to

employ a combination of analytical and formal types of writing. Analytical writing seeks to go beyond the mere presentation of facts to the reader by providing substantive analysis and evaluation of a topic. This type of writing, however, increases the cognitive complexity and might backfire when writing a negative movie review as it can potentially create confirmation bias among the harbinger critics. More specifically, they might use analytical thinking to list all the potential factors of failure of a movie, and write their movie review by sticking to information that confirms their preconceptions (Nickerson, 1998).

Table 3 This table reports the average % of writing style metrics and the t-test results for harbinger and non-harbinger movie critics stratified by review sentiment. Standard deviations are in parentheses.

Writing Style Comparison of Group 1 and Group 4 – Positive Reviews				
Metrics	Mean of Group 1 (SD)	Mean of Group 4 (SD)	T-value	P-value
Tone	64.6% (.178)	65.1% (.185)	0.25	0.4
Adverb	3.9% (.019)	4.1% (.018)	0.94	0.17
Adjectives	7.2% (.026)	7% (.024)	0.59	0.27
Analytical	78.6% (.111)	77.9% (.114)	0.58	0.28
Conjunctions	6.4% (.018)	6.7% (.017)	1.40	0.08*
Past Focus	1.25% (.013)	1.10% (.008)	1.22	0.11
Third Person	0.56% (.006)	0.70% (.008)	1.59	0.06*
Third Person Plural	0.28% (.004)	0.33% (.004)	0.99	0.16
Anxiety	0.51% (.005)	0.47% (.005)	0.77	0.22
Filler	0.01% (.0002)	0.0019% (.001)	1.41	0.07*
NonFluency	0.16% (.003)	0.15% (.002)	0.20	0.42
Hyphens	2.11% (.016)	1.96% (.013)	0.91	0.18
Uncommon Words	17% (.03)	17.2% (.03)	0.52	0.29
1st Person Singular PRO	0.30% (.006)	0.18% (.004)	2.08	0.02**
1st Person Plural PRO	0.40% (.005)	0.39% (.006)	0.18	0.42
Clout	56% (.111)	56.6% (.094)	0.49	0.31
Writing Style Comparison of Group 1 and Group 4 – Negative Reviews				
Metrics	Mean of Group 1 (SD)	Mean of Group 4 (SD)	T-value	P-value
Tone	47.1% (.248)	48.7% (.248)	0.54	0.3
Adverb	5.1% (.035)	4.6% (.029)	1.17	0.12
Adjectives	6.4% (.032)	6.9% (.033)	1.33	0.09*
Analytical	73.4% (.192)	76.6% (.179)	1.51	0.06*
Conjunctions	6.4% (.031)	6.3% (.024)	0.22	0.41
Past Focus	1.8% (.02)	1.7% (.02)	0.57	0.28
Third Person	0.54% (.009)	0.45% (.007)	0.96	0.16
Third Person Plural	0.36% (.007)	0.33% (.005)	0.55	0.29
Anxiety	0.47% (.008)	0.52% (.009)	0.51	0.31
Filler	0.02% (.002)	0.007% (.0004)	0.76	0.22
NonFluency	0.11% (.003)	0.09% (.003)	0.38	0.35
Hyphens	1.86% (.02)	2.13% (.02)	1.12	0.13
Uncommon Words	17% (.06)	17.9% (.05)	1.29	0.09*
1st Person Singular PRO	0.29% (.006)	0.26% (.005)	0.41	0.34
1st Person Plural PRO	0.36% (.007)	0.32% (.005)	0.55	0.29
Clout	49.8% (.16)	48.9% (.14)	0.55	0.29

Motivated by the above findings, we investigate further to see if they still hold true when we stratify across critic status category. The results for positive reviews are presented in Table 4 and for negative reviews in Table E8. Surprisingly, we find that non-harbinger critics use more self-reference pronouns in their positive reviews than non-harbinger critics regardless of status category, but this difference becomes even stronger in the case of top critics. This

indicates that non-harbinger critics are more self-reflective overall compared to harbinger critics who are overconfident in their abilities when writing a positive review. What is more is that in the top-critic category, harbinger critics use fewer adjectives and more adverbs. These two features have been shown to be associated with extraversion (Dewaele and Furnham, 1999). Hence, we find support for our Hypothesis 1 that top harbinger critics are over-optimistic in their positive reviews. With respect to negative reviews, we find that harbinger critics write in an analytical and formal style, whereas non-harbinger critics engage in a more narrative style of writing, but this difference becomes much stronger for the top-critic category.

Table 4 This table reports the average % of writing style metrics and the t-test results for harbinger and non-harbinger movie critics stratified by status category and review sentiment. Standard deviations are in parentheses.

Writing Style Comparison of Group 1 and Group 4 Top Critics – Positive Reviews				
Metrics	Mean of Group 1 (SD)	Mean of Group 4 (SD)	T-value	P-value
Tone	64.3% (.184)	58.4% (.191)	1.50	0.06*
Adverb	3.7% (.021)	4.3% (.018)	1.46	0.07*
Adjectives	6.4% (.022)	6.3% (.019)	0.33	0.37
Analytical	78.4% (.121)	79.1% (.106)	0.31	0.37
Conjunctions	6.5% (.016)	6.8% (.019)	0.81	0.20
Past Focus	0.9% (.01)	1.1% (.01)	1.33	0.09*
Third Person	0.71% (.009)	0.82% (.007)	0.61	0.27
Third Person Plural	0.36% (.004)	0.39% (.004)	0.43	0.33
Anxiety	0.54% (.006)	0.6% (.006)	0.48	0.31
Filler	0.02% (.0001)	0.0042% (.0003)	0.79	0.21
NonFluency	0.14% (.002)	0.16% (.002)	0.43	0.33
Hyphens	2.26% (.02)	2.20% (.02)	0.15	0.43
Uncommon Words	16.6% (.02)	16.9% (.02)	0.53	0.29
1st Person Singular PRO	0.32% (.007)	0.11% (.002)	1.71	0.04**
1st Person Plural PRO	0.23% (.004)	0.34% (.007)	0.93	0.18
Clout	57.7% (.11)	56.9% (.07)	0.38	0.35
Writing Style Comparison of Group 4 and Group 1 Non-Top Critics – Positive Reviews				
Metrics	Mean of Group 1 (SD)	Mean of Group 4 (SD)	T-value	P-value
Tone	64.7% (.178)	68.1% (.175)	1.4	0.06*
Adverb	3.9% (.018)	4% (.019)	0.13	0.44
Adjectives	7.4% (.027)	7.3% (.025)	0.23	0.41
Analytical	78.6% (.11)	77.3% (.11)	0.92	0.17
Conjunctions	6.4% (.019)	6.6% (.016)	1.06	0.14
Past Focus	1.36% (.014)	1.07% (.009)	1.86	0.03**
Third Person	0.52% (.005)	0.64% (.008)	1.32	0.09*
Third Person Plural	0.26% (.004)	0.29% (.004)	0.72	0.23
Anxiety	0.51% (.005)	0.41% (.004)	1.44	0.07*
Filler	0.009% (.0006)	0.0009% (.00001)	1.32	0.09*
NonFluency	0.17% (.003)	0.16% (.002)	0.38	0.35
Hyphens	2.2% (.021)	1.9% (.015)	1.42	0.07*
Uncommon Words	17.1% (.03)	17.3% (.03)	0.49	0.31
1st Person Singular PRO	0.30% (.005)	0.21% (.004)	1.25	0.1*
1st Person Plural PRO	0.46% (.006)	0.41% (.006)	0.62	0.26
Clout	55.5% (.11)	56.4% (.10)	0.69	0.24

5 Conclusions and Discussion

Predicting a movie's box office revenue is one of the most fundamental needs in the motion picture industry. This becomes even more challenging when someone tries to predict the opening weekend box office revenue since there is little to no information available about the wider audience's reaction. As [Cabral and Natividad \(2016\)](#) show in their work, doing well at the box office during the opening weekend has an economically and statistically significant effect on the movie's eventual performance. Prior research has explored how different attributes of a movie, such as star inclusion ([Elberse, 2007](#); [Karniouchina, 2011](#); [Liu et al, 2014](#)), the activity level of editors and viewers of the movie's corresponding entry in Wikipedia ([Mestyán et al, 2013](#)), and competition among movies that are released at the same time ([Ainslie et al, 2005](#); [Hennig-Thurau et al, 2007](#); [Delre et al, 2016](#)), could predict the box office performance. With regards to online movie reviews, [Moon et al \(2010\)](#) show that movie ratings from professional critics and viewers' communities are predictive of total box office revenue. Moreover, [Basuroy et al \(2003\)](#) find that both positive and negative reviews are significantly correlated with box office revenue with the impact of negative reviews (but not positive reviews) diminishing over time. We diverge from this research and propose that a distinction should be made among movie critics because as we demonstrate not all positive (negative) reviews are a signal for success (failure).

More specifically, we combine three different data sources to empirically investigate the harbinger of failure phenomenon in the motion picture industry. We analyze the pre-release reviews written on movies by film critics and find that harbingers of failure do exist. Their positive pre-release reviews provide a strong predictive signal that the movie will turn out to be a flop. Moreover, we find the harbinger effect to be symmetric. That is, there are harbinger critics who give negative reviews and the movie turns out to be successful. These findings shed new light on the influencer-predictor hypothesis. We document that in a pre-release setting, there is a portion of movie critics that neither influence nor predict the right outcome of a movie. On the contrary, the outcome of the movie turns out to be the exact opposite of their prediction.

We further analyze the writing style of film critics and connect them to potential cognitive biases that might give rise to the harbinger phenomenon. We find that harbinger critics engage in an analytical and formal style of writing and have a lower rate of self-reference pronouns compared to non-harbingers. These differences indicate that harbinger critics are less self-reflective about the audience's opinion compared to non-harbingers and tend to show overconfidence in their abilities. When we stratify the analysis across the critic status category, we find the aforementioned differences between harbingers and non-harbingers to be even more significant. In the cohort of top critics, which usually consists of professional and experienced reviewers, we find that top critic harbingers are also using more adverbs than their counterparts in their positive reviews indicating that they are over-optimistic in

their assessments of movies. In the case of negative reviews, we find that the top critic harbingers are significantly more analytical than their counterparts.

Our findings have important managerial implications for the motion picture industry and its key channel entities: movie studios, distribution companies, and movie theaters. First, our research provides a methodology based on pre-release film reviews that allows movie producers and distributors to identify early on which movies are going to perform badly. This will in turn allow them to make better pre-launch marketing decisions and save significant marketing costs on flop movies. It will also allow theaters to better allocate their theater space, a finite resource that is crucial to the success of theaters. Second, film studios can greatly benefit from identifying the harbinger critics and using them during the early stages of production. More specifically, movie studios can use harbinger critics to select the scripts that will maximize their box office revenue instead of using mere guesswork. Our approach complements that of [Eliashberg et al \(2007\)](#), which uses natural-language processing to select the winning scripts. Third, our approach can potentially serve as a diagnostic for reviewers across fields. It is crucial for companies that employ reviewers to know whether their reviewers' opinions can be used as a diagnostic tool to determine success or failure. Based on that, companies might want to reclassify who they designate as a "top reviewer" or create a new class of reviewers.

We should acknowledge here the limitations of this work and present potential future research avenues. The main limitation of our research, similar to [Anderson et al \(2015\)](#) and [Simester et al \(2019\)](#), is that we do not provide a causal explanation about where the harbingers' preferences are coming from. However, we do provide insights into their personality traits and cognitive biases based on their writing style. This paves new avenues for further experimental behavioral research about the underlying mechanisms of the harbinger phenomenon. A second potential limitation of our research is that movie producers and distributors could potentially strategically pick which critics to invite to their pre-release screenings. However, there is no substantial evidence that this is happening in the movie industry, as it would jeopardize the reputation of, and confidence in, movie studios and movie critics alike. Last, another possibility, unobserved to the researcher, is that critics are getting influenced by other critics before they submit their review by either reading their reviews or talking to them. Future research might include further contextual variables, such as the choice of movies to review by harbingers, the timing of the reviews, and the lack of learning, to further the theoretical understanding of the harbinger phenomenon in the motion picture industry.

To conclude, does the positive (negative) feedback given by film critics before the release of a movie signal its financial failure (success) instead of its success (failure) as currently believed? Our findings document that this is not always the case. At least in a pre-release movie setting, a distinction should be made among film critics because of the existence of harbinger critics; their endorsement of a movie is a signal of the opposite outcome.

Declarations

Conflict of interest The authors declare no competing interests. **Funding** No funding was received for conducting this study. **Ethical approval** The authors declare compliance with the ethical standards. This research does not involve Human Participants and/or Animals.

References

- Ainslie A, Drèze X, Zufryden F (2005) Modeling movie life cycles and market share. *Marketing science* 24(3):508–517
- Anderson E, Lin S, Simester D, et al (2015) Harbingers of failure. *Journal of Marketing Research* 52(5):580–592
- Barrick MR, Mount MK (1991) The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology* 44(1):1–26
- Basuroy S, Chatterjee S, Ravid SA (2003) How critical are critical reviews? the box office effects of film critics, star power, and budgets. *Journal of marketing* 67(4):103–117
- Basuroy S, Desai KK, Talukdar D (2006) An empirical investigation of signaling in the motion picture industry. *Journal of marketing research* 43(2):287–295
- Boatwright P, Basuroy S, Kamakura W (2007) Reviewing the reviewers: The impact of individual film critics on box office performance. *Quantitative marketing and economics* 5(4):401–425
- Cabral L, Natividad G (2016) Box-office demand: The importance of being# 1. *The Journal of Industrial Economics* 64(2):277–294
- Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *Journal of marketing research* 43(3):345–354
- Chintagunta PK, Gopinath S, Venkataraman S (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):944–957
- Costa Jr PT, McCrae RR (1992) Four ways five factors are basic. *Personality and individual differences* 13(6):653–665
- Davis D, Brock TC (1975) Use of first person pronouns as a function of increased objective self-awareness and performance feedback. *Journal of Experimental Social Psychology* 11(4):381–388

- Dellarocas C, Zhang XM, Awad NF (2007) Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing* 21(4):23–45
- Delre SA, Broekhuizen TL, Bijmolt TH (2016) The effects of shared consumption on product life cycles and advertising effectiveness: the case of the motion picture market. *Journal of Marketing Research* 53(4):608–627
- Dewaele JM, Furnham A (1999) Extraversion: The unloved variable in applied linguistic research. *Language Learning* 49(3):509–544
- Duan W, Gu B, Whinston AB (2008) Do online reviews matter?—an empirical investigation of panel data. *Decision support systems* 45(4):1007–1016
- Elberse A (2007) The power of stars: Do star actors drive the success of movies? *Journal of marketing* 71(4):102–120
- Eliashberg J, Shugan SM (1997) Film critics: Influencers or predictors? *Journal of marketing* 61(2):68–78
- Eliashberg J, Hui SK, Zhang ZJ (2007) From story line to box office: A new approach for green-lighting movie scripts. *Management Science* 53(6):881–893
- Eysenck HJ (1991) Dimensions of personality: 16, 5 or 3?—criteria for a taxonomic paradigm. *Personality and individual differences* 12(8):773–790
- Hennig-Thurau T, Henning V, Sattler H, et al (2007) The last picture show? timing and order of movie distribution channels. *Journal of Marketing* 71(4):63–83
- Ho-Dac NN, Carson SJ, Moore WL (2013) The effects of positive and negative online customer reviews: do brand strength and category maturity matter? *Journal of marketing* 77(6):37–53
- Karniouchina EV (2011) Impact of star and movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing* 28(1):62–74
- Küfner AC, Back MD, Nestler S, et al (2010) Tell me a story and i will tell you who you are! lens model analyses of personality and creative writing. *Journal of Research in Personality* 44(4):427–435
- Litman BR (1983) Predicting success of theatrical movies: An empirical study. *Journal of popular culture* 16(4):159
- Liu A, Liu Y, Mazumdar T (2014) Star power in the eye of the beholder: A study of the influence of stars in the movie industry. *Marketing letters*

25(4):385–396

- Mestyán M, Yasseri T, Kertész J (2013) Early prediction of movie box office success based on wikipedia activity big data. *PloS one* 8(8):e71,226
- Moon S, Bergey PK, Iacobucci D (2010) Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. *Journal of marketing* 74(1):108–121
- Nickerson RS (1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology* 2(2):175–220
- Pennebaker JW, King LA (1999) Linguistic styles: language use as an individual difference. *Journal of personality and social psychology* 77(6):1296
- Pennebaker JW, Mehl MR, Niederhoffer KG (2003) Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54(1):547–577
- Pennebaker JW, Booth RJ, Francis ME (2007) Linguistic inquiry and word count: Liwc [computer software]. Austin, TX: liwc net 135
- Reddy SK, Swaminathan V, Motley CM (1998) Exploring the determinants of Broadway show success. *Journal of Marketing Research* 35(3):370–383
- Sharot T (2011) The optimism bias. *Current biology* 21(23):R941–R945
- Simester DI, Tucker CE, Yang C (2019) The surprising breadth of harbingers of failure. *Journal of Marketing Research* 56(6):1034–1049
- Slatcher RB, Chung CK, Pennebaker JW, et al (2007) Winning words: Individual differences in linguistic style among US presidential and vice presidential candidates. *Journal of Research in Personality* 41(1):63–75
- Song T, Huang J, Tan Y, et al (2019) Using user-and marketer-generated content for box office revenue prediction: Differences between microblogging and third-party platforms. *Information Systems Research* 30(1):191–203
- Tausczik YR, Pennebaker JW (2010a) The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology* 29(1):24–54
- Tausczik YR, Pennebaker JW (2010b) The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1):24–54
- Waldfogel J (1999) Preference externalities: An empirical study of who benefits whom in differentiated product markets

West RF, Stanovich KE (1997) The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin & Review* 4(3):387–392

Appendix A Data Overview and Variable Description

Table A1 Final dataset overview

Dataset Overview	
Time frame	2015 - 2019
Number of Reviews	28884
Number of Movies	448
Calibration/Training sample size	286
Holdout/Testing sample size	162

Table A2 Variable Description

Source	Variable	Description	Data Type
Box Office Mojo	Movie Title	Title of the movie.	String
	Release Date	Movie release date to the public.	Date
	Opening Weekend Revenue	Total revenue during the opening weekend of the movie.	Integer
	Gross Domestic Revenue	Gross domestic revenue of the movie in US\$.	Integer
	Genre	Genre of the movie.	String
Numbers	Movie Budget	Budget of the movie.	Integer
Rotten Tomato	Critic Name	Name of the critic.	String
	Review Date	Date when the critic posted the review.	Date
	Top Critic	Whether the critic is a top critic or not.	Boolean
	Review Text	The review text as shown in Rotten Tomatoes.	String
	Fresh or Rotten	The review is flagged as "Fresh" if it's a positive one, and "Rotten" if it's negative.	String
	Genre	The genre of the movie.	String
	Studio	A movie's production/distribution company.	String
	Sequel	Whether the movie is a sequel or not.	Boolean

Appendix B F1 Score Grouping Approach

In addition to the flop affinity ratio, we group critics based on their F1 score. F1 score originates from machine learning and measures the model accuracy by taking the harmonic mean of precision and recall. That is:

$$F_1 = 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}}$$

In our setting F1 score can serve as a measurement of the movie critics' tastes relative to the public taste. Specifically, precision is the number of correctly predicted positive cases, divided by the total number of positive cases identified by the model, including both correct and incorrect predictions. Recall has the same numerator as precision, but its denominator is the number of true positive cases in reality. If a movie critic can be construed as a predictive model, one's precision score will be the number of successful movies positively reviewed divided by the number of movies positively reviewed, and the recall score will be the number of successful movies positively reviewed divided by the number of actual successful movies the critic commented on. Therefore, the formula of F1 score can be derived using the following equation:

$$F_1 = 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}} = \frac{TP}{TP + \frac{1}{2}(FP + FN)},$$

where TP is the number of successful movies reviewed positively, FP is the number of flop movies reviewed positively and FN is the number of successful movies reviewed negatively.

After computing the F1 score for each critic, we group the critics using F1 score quartile cutoff. Under this approach, the groups are distributed more evenly, and Group 1 has the largest number of harbingers of failure effect (as the incorrect predictions are in the denominator), leading to a lower value of F1 score. Then, we estimate Models 1 and 2 similarly to the flop affinity approach. The chi-square test on the likelihood ratio of Model 1 and Model 2 shows that the improvement of Model 2 fitness is significant. The model results are summarized in the table below. The coefficients of the harbinger groups, Group 1 and Group 2, are significant and have negative signs as expected.

Critics Distribution Using F1 Score			
Group	F1 Range	Percentile Cutoff	No. of Critics
Group 1	[0, 48%]	25th percentile	153
Group 2	(48%, 57%]	50th percentile	158
Group 3	(57%, 67%]	75th percentile	154
Group 4	(67%, 100%]		147
Other (≤ 2 positive reviews)			226

Results of Logistic Models Using F1 Score Grouping Method		
	Model 1	Model 2
Total positive reviews	-.0008 (.002)	
Group 1 positive reviews		-.35 (.073)***
Group 2 positive reviews		-.12 (.045)***
Group 3 positive reviews		.01 (.042)**
Group 4 positive reviews		.43 (.088)***
Log-likelihood	-197.45	-127.18
Likelihood ratio test, chi-square (d.f.=3)		140.53***
AUCPR	.56	.65
F1	.22	.44

* p-value $\leq .1$, ** p-value $\leq .05$, *** p-value $\leq .01$

Table B3 The table reports the coefficients of interest from Equations 1 and 2 using the F1 Grouping Method. Standard errors are in parentheses.

Appendix C Sleeper Movies

To evaluate the robustness of our results, we rerun the analysis to account for sleeper movies. Sleeper movies are movies that tanked in their opening box office but became a success later on. Using overall domestic box-office revenue data, we find that 20 out of the 448 movies in our dataset are sleeper movies. We re-classify movie critics into five groups based on their flop affinity score. Table C4 shows all the relevant information on the critics' grouping.

Table C4 Movie critics grouping based on their flop affinity score when accounting for sleeper movies.

Critics Distribution Accounting For Sleeper Movies			
Group	Flop Affinity Range	Percentile Cutoff	No. of Critics
Group 1	[0, 33%]	25th percentile	160
Group 2	(33%, 47%]	50th percentile	146
Group 3	(47%, 58%]	75th percentile	155
Group 4	(58%, 100%]		151
Other (≤ 2 positive reviews)			226

We re-estimate models 1 and 2, and present the results in Table C5. First, we observe that the pattern of findings remains unchanged. Moreover, we find that the coefficients of Group 1 and Group 4 are now larger in absolute size indicating a stronger harbinger effect. More specifically, we see that the expected change in log odds for a one-unit increase in the reviews of Group 1 and Group 4 are now 0.66 and -0.57, respectively. Equivalently, a one-unit increase in the reviews of Group 1 increases the odds of a movie being successful by 93%, while a one-unit increase in the reviews of Group 4 decreases the odds of success by about 43%.

Table C5 The table reports the coefficients of interest from Equations 1 and 2 when accounting for sleeper movies. Standard errors are in parentheses.

Logistic Models Results		
	Model 1	Model 2
Total positive reviews	.001 (.002)	
Group 1 positive reviews		.66 (.121)***
Group 2 positive reviews		.14 (.055)***
Group 3 positive reviews		-.07 (.046)
Group 4 positive reviews		-.57 (.01)***
Log-likelihood	-197.4	-102.8
Likelihood ratio test, chi-square (d.f.=3)		171.3***
AUCPR	.57	.75
F1	.27	.62

* p-value $\leq .1$, ** p-value $\leq .05$, *** p-value $\leq .01$

Appendix D Success Avoidance

We investigate here whether the harbinger effect is symmetric. To do this, we group critics based on their success avoidance score. For each movie critic i , we calculate the proportion of negatively reviewed movies that turned out to be successful over their total number of negatively reviewed movies. That is:

$$\text{Success Avoidance}_i = \frac{\text{Number of Successful Movies Reviewed Negatively}}{\text{Number of Movies Reviewed Negatively}}$$

Following this definition, we get the following grouping of critics shown in Table D6. Our two competing models are modified as follows:

$$\text{Model 1: } \log(\text{Odds of Success of Movie } i) = \alpha + \beta_0 \text{TotalNegativeReviews}_i \quad (\text{D1})$$

$$\begin{aligned} \text{Model 2: } \log(\text{Odds of Success of Movie } i) = & \alpha + \beta_1 \text{Group1NegativeReviews}_i + \\ & \beta_2 \text{Group2NegativeReviews}_i + \beta_3 \text{Group3NegativeReviews}_i \\ & + \beta_4 \text{Group4NegativeReviews}_i \end{aligned} \quad (\text{D2})$$

Observing the results in Table D7, we can conclude that the harbinger effect is symmetric.

Table D6 Movie critics grouping based on their success avoidance score

Critics Distribution			
Group	Success Avoidance Range	Percentile Cutoff	No. of Critics
Group 1	[0, 20%]	25th percentile	122
Group 2	(20%, 31%]	50th percentile	107
Group 3	(31%, 41%]	75th percentile	108
Group 4	(41%, 100%]		112
Other (≤ 2 negative reviews)			389

Table D7 The table reports the coefficients of interest from Equations D1 and D2 in the case of success avoidance. Standard errors are in parentheses.

Logistic Models Results Using Success Avoidance Grouping Method		
	Model 1	Model 2
Total negative reviews	-.0019 (.004)***	
Group 1 negative reviews		-1.37 (.211)***
Group 2 negative reviews		-.27 (.099)***
Group 3 negative reviews		-.30 (.086)***
Group 4 negative reviews		.75 (.144)***
Log-likelihood	-185.06	-190.53
Likelihood ratio test, chi-square (d.f.=3)		225***
AUCPR	.63	.64
F1	.26	.4

* p-value $\leq .1$, ** p-value $\leq .05$, *** p-value $\leq .01$

Appendix E Writing Style Differences for Negative Reviews

Table E8 This table reports the average % of writing style metrics and the t-test results for harbinger and non-harbinger movie critics stratified by status category and review sentiment. Standard deviations are in parentheses.

Writing Style Comparison of Group 1 and Group 4 Top Critics – Negative Reviews				
Metrics	Mean of Group 1 (SD)	Mean of Group 4 (SD)	T-value	P-value
Tone	48.6% (.25)	47.4% (.19)	.25	0.40
Adverb	5% (.02)	4.6% (.02)	0.91	0.18
Adjectives	6.1% (.02)	6.6% (.03)	0.87	0.19
Analytical	73.4% (.13)	78.7% (.12)	1.98	0.02**
Conjunctions	6.5% (.02)	6.4% (.02)	0.24	0.40
Past Focus	1.9% (.02)	1.5% (.01)	1.06	0.14
Third Person	0.77% (.01)	0.42% (.006)	1.81	0.04**
Third Person Plural	0.36% (.007)	0.28% (.004)	0.54	0.29
Anxiety	0.45% (.007)	0.54% (.008)	0.49	0.31
Filler	0% (.0)	0% (.0001)	1	0.16
NonFluency	0.10% (.003)	0.15% (.003)	0.74	0.23
Hyphens	2.3% (.02)	2.2% (.01)	0.3	0.38
Uncommon Words	15.7% (.03)	17.1% (.03)	1.72	0.04**
1st Person Singular PRO	0.13% (.003)	0.24% (.004)	1.44	0.07*
1st Person Plural PRO	0.25% (.006)	0.20% (.004)	0.42	0.33
Clout	49.4% (.14)	48% (.09)	0.56	0.28
Writing Style Comparison of Group 4 and Group 1 Non-Top Critics – Negative Reviews				
Metrics	Mean of Group 1 (SD)	Mean of Group 4 (SD)	T-value	P-value
Tone	46.6% (.24)	49.2% (.27)	0.75	0.22
Adverb	5.1% (.04)	4.6% (.03)	0.89	0.18
Adjectives	6.5% (.03)	7% (.03)	1.13	0.13
Analytical	73.4% (.2)	75.5% (.2)	0.77	0.22
Conjunctions	6.4% (.03)	6.3% (.02)	0.17	0.43
Past Focus	1.7% (.02)	1.7% (.02)	0.05	0.48
Third Person	0.46% (.008)	0.46% (.008)	0.01	0.49
Third Person Plural	0.36% (.008)	0.33% (.006)	0.29	0.38
Anxiety	0.48% (.008)	0.51% (.009)	0.30	0.38
Filler	0.02% (.002)	0.006% (.0003)	0.91	0.18
NonFluency	0.12% (.003)	0.07% (.003)	0.87	0.19
Hyphens	1.7% (.02)	2.1% (.02)	1.59	0.06*
Uncommon Words	17.3% (.06)	18.2% (.06)	1.03	0.15
1st Person Singular PRO	0.35% (.005)	0.27% (.005)	0.77	0.22
1st Person Plural PRO	0.33% (.006)	0.27% (.005)	0.71	0.24
Clout	50% (.17)	49.4% (.15)	0.28	0.38