

Rossi, Michelangelo; Schleef, Felix

**Working Paper**

## Quality disclosures and disappointment: Evidence from the academy nominations

CESifo Working Paper, No. 11573

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Rossi, Michelangelo; Schleef, Felix (2026) : Quality disclosures and disappointment: Evidence from the academy nominations, CESifo Working Paper, No. 11573, Munich Society for the Promotion of Economic Research - CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/338294>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**CES ifo**

**11573**  
**2026**

Original Version: December 2024  
This Version: Januar 2026

# Working Papers

## **Quality Disclosures and Disappointment: Evidence from the Academy Nominations**

Michelangelo Rossi, Felix Schleef

**CES ifo**

Imprint:

**CESifo Working Papers**

ISSN 2364-1428 (digital)

Publisher and distributor: Munich Society for the Promotion  
of Economic Research - CESifo GmbH

Poschingerstr. 5, 81679 Munich, Germany  
Telephone +49 (0)89 2180-2740

Email [office@cesifo.de](mailto:office@cesifo.de)  
<https://www.cesifo.org>

Editor: Clemens Fuest

An electronic version of the paper may be downloaded free of charge

- from the CESifo website: [www.ifo.de/en/cesifo/publications/cesifo-working-papers](http://www.ifo.de/en/cesifo/publications/cesifo-working-papers)
- from the SSRN website: [www.ssrn.com/index.cfm/en/cesifo/](http://www.ssrn.com/index.cfm/en/cesifo/)
- from the RePEc website: <https://ideas.repec.org/s/ces/ceswps.html>

# Quality Disclosures and Disappointment: Evidence from the Academy Nominations\*

Michelangelo Rossi and Felix Schleeff<sup>†</sup>

October 20, 2025

## Abstract

This study examines the unintended consequences of quality disclosures, focusing on how Academy Award nominations impact consumer satisfaction in the movie industry. Awards and certifications typically signal high quality and increase consumer expectations. Yet, if the experience falls short of the expectation, they may also lead to disappointment. Using a novel dataset from MovieLens, we analyze user ratings for movies surrounding Academy Award nominations from 1995 to 2019. We first implement a difference-in-differences strategy comparing nominated and non-nominated films, and then introduce a novel recommendation-based matching approach that leverages vector representations of user preferences trained prior to the nominations. Our analysis removes taste-based selection and isolates changes in user experience: users who rate a movie after its nomination assign significantly lower ratings than similar users who rated the same film earlier. This effect accounts for more than 7% of the pre-nomination rating gap between nominated and non-nominated films and is most pronounced among less experienced users. Our findings are validated with data from IMDb, where the effect is even more pronounced, likely reflecting differences in the composition of the user base across platforms. Additional textual analysis of user-generated content on both platforms provides further evidence that the post-nomination decline in ratings is driven by disappointment, rather than disinterest, deteriorating viewing conditions, or snob effects associated with mainstream popularity.

---

\*We would like to thank Heski Bar-Isaac, Tommaso Bondi, Rahul Deb, Xavier D’Haultfoeuille, Natalia Fabra, Alina Ferracatu, Roxana Fernandez, Duarte Gonçalvez, Matilde Machado, Leonardo Madio, Sridhar Moorthy, Yeşim Orhun, Pedro Rey Biel, Navdeep Sahni, Avner Strulov-Shlain, Kevin Ducbao Tran, Mark J. Tremblay, and Thibaud Vergé. Helpful feedback was received at the internal seminars at Universidad Carlos III de Madrid, CREST, University of Toronto, and University of Nevada; the 19th International Industrial Organization Conference, the 22nd ACM Conference on Economics and Computation (EC’21), the 50th European Association Research Industrial Economics Conference, and the Behavioral Industrial Organization & Marketing Symposium. All errors are ours.

<sup>†</sup>M.Rossi: HEC Paris, and CESifo. E-mail: [rossim@hec.fr](mailto:rossim@hec.fr). F.Schleeff: HEC Paris. E-mail: [schleeff@hec.fr](mailto:schleeff@hec.fr).

# 1 Introduction

Third-party quality disclosures are often used in markets where consumers are imperfectly informed about sellers' quality. In the hospitality industry, hotels and restaurants advertise using Michelin stars, and in the film industry, movies are promoted based on awards like the Academy Awards (colloquially known as "Oscars"). Used car sellers might show prospective buyers certificates of the car's latest inspections. The positive effect of quality disclosure has already been studied in many contexts (Chen and Xie, 2005; Hilger, Rafert and Villas-Boas, 2011; Friberg and Grönqvist, 2012). When a third-party is credible, its certification or awards can effectively signal sellers' quality and attract buyers towards high-quality products.

Yet, the effect of certifications and awards on consumers' satisfaction depends on several factors (Li, Deng, Manchanda and De Reyck, 2024). Awarded sellers may have incentives to change either the quality of the product they offer or its price (Ham, Koch, Lim and Wu, 2021; Farronato and Zervas, 2022). Such adjustments may affect the consumer experience and mitigate the positive demand shock generated by the certification itself.

Even when quality and prices are held constant, quality disclosures may affect consumer satisfaction through different mechanisms. A positive certification or recommendation may attract a different type of consumers (Li and Hitt, 2008; Aridor, Gonçalves, Kluver, Kong and Konstan, 2024a). For example, a movie that receives an Oscar nomination may draw more inexperienced viewers. As a result, changes in audience composition may shift overall satisfaction due to the differing preferences of the consumers attracted by the award (Bondi, 2023; Lagios and Méon, 2023). In addition, quality disclosures can alter how consumers experience the product (Kovács and Sharkey, 2014). Awards often raise expectations, which may increase the likelihood of disappointment if the actual experience does not match them. Other concurrent channels may also be at play: some users who watch a movie after its nomination may be less intrinsically interested in it, but feel social pressure to consume a "prestigious" product; others may face changes in the conditions under which they consume it (for example, watching the film in more crowded theaters). Furthermore, consumers may exhibit conformity bias, deriving more enjoyment from a positively certified product (Gergaud, Storchmann and Verardi, 2015), or experience a reverse-bandwagon (snob) effect, where the increased popularity due to an award nomination diminishes their satisfaction (Leibenstein, 1950). All of these mechanisms can be understood as channels that affect utility or satisfaction, even for users with the very same underlying tastes and preferences.

In this paper, we isolate one of these channels from the others. We provide empirical evidence on how certifications, and in particular the disappointment they may generate, can alter consumer experience, shed-

ding light on the unintended downside of positive quality signals in the movie industry. We examine the causal impact of Academy of Motion Picture Arts and Sciences (AMPAS) award nominations on movie ratings. Using data from the MovieLens platform, we focus on an 80-day window around AMPAS nominations (40 days before and after) from 1995 to 2019.

We start by presenting a difference-in-differences (DiD) estimate of the effect of nominations on ratings, comparing ratings for nominated and non-nominated movies before and after the AMPAS nominations. This analysis shows that nominated films have higher ratings compared to non-nominated movies, but they experience a significant drop in their ratings relative to non-nominated films after the nominations. This approach captures both selection effects and the range of channels that can change user experience, including expectation-based disappointment, disinterest, altered viewing conditions, and social-influence mechanisms.

To separate these channels from pure taste-based selection, we propose a novel recommendation-based matching approach. Our method involves two steps. First, we train a recommender system using only ratings posted prior to the AMPAS nominations, generating a vector representation of consumer preferences. We then use this vector representation of tastes in a propensity score matching strategy, enabling us to account for selection based on preferences. By comparing the ratings of these matched users posted before and after the nominations, we estimate the extent to which award nominations change users' reported satisfaction, holding constant underlying tastes. Our results show that users who rate a nominated movie after its nomination give significantly lower ratings than users with similar tastes who rated the movie before. This short-run effect accounts for more than 7% of the pre-nominations difference in ratings between nominated and non-nominated films. The effect is primarily driven by users who have posted relatively few ratings on the platform, suggesting that quality disclosures such as AMPAS nominations mainly influence less experienced users, who are more likely to rely on awards as quality signals. As a placebo analysis, we repeat the matching approach for non-nominated movies and find no comparable effect.

To assess external validity, we also scrape all reviews posted for the nominated movies in our sample on IMDb, along with all reviews from users who rated these films. This dataset does not allow us to apply the same recommendation-based matching approach. Yet, we control for several observable user characteristics, helping to reduce potential confounding due to selection. The significant drop in ratings after nominations is confirmed, and its magnitude is approximately three times larger than the effect observed on MovieLens. This suggests that our findings extend to other platforms and that our analysis on MovieLens may underestimate the true extent of changes in user experience. MovieLens users are in general experienced moviegoers, who are less likely to be disappointed or to assign AMPAS nominations a major role in shaping their opinions. In many other contexts, however, such as IMDb (where users tend to have lower expertise), consumers

may be more sensitive to nominations and other quality disclosures.

In this context, the quality of a film is fixed after its release, and prices remain relatively stable: nominated films are no more expensive to watch than non-nominated films. Therefore, aside from potential selection effects, our analysis should primarily capture changes in how consumers experience movies following an award nomination.

We interpret the changes in users' satisfaction observed after the nominations as being primarily driven by disappointment. The stronger decline observed among less experienced users is consistent with this interpretation, as they are more reliant on external signals in forming expectations. A change in disinterest for nominated movies could in principle play a role, but this would likely be captured by our matching approach if users had displayed a similar pattern in previous years for comparable films. Changes in viewing conditions may also matter, although our focus on a short window of days implies that substantial shifts in theater environments would be required. Moreover, we do not find an effect for non-nominated movies, which could suffer the opposite treatment (being demoted to smaller screens or watched in less crowded theaters), suggesting that this channel is not first order. With respect to the reverse-bandwagon effect, MovieLens users are less exposed to social signaling pressures, since ratings are not tied to identities and individual reviews are not visible.<sup>1</sup> While some users may become more critical once a film becomes too popular, this explanation is difficult to reconcile with our evidence: more experienced users are less, rather than more, affected by nominations, even though such users would be the most plausible candidates for "snob" behavior by rejecting mainstream movies.<sup>2</sup>

To further reinforce the interpretation that the observed decline in ratings is driven by disappointment, we examine the textual content generated by users on both platforms. On MovieLens, we analyze user-generated tags, which provide preliminary suggestive evidence consistent with the disappointment hypothesis. We then turn to a richer textual dataset: written reviews on IMDb. Leveraging the more detailed and nuanced nature of this content, we find that IMDb users are significantly more likely to express disappointment in their reviews following the nominations. In contrast, we find no comparable evidence supporting alternative explanations for a decline in ratings, such as disinterest, deteriorated viewing conditions, or the reverse-bandwagon effect.

Our analysis sheds light on the welfare consequences of quality disclosures. Most of the literature has

---

<sup>1</sup>When clicking on a movie's page on MovieLens, users see a personalized predicted rating, the average rating from all users, and a histogram of all ratings. However, these ratings are not associated with individual users, and user-level data remain private.

<sup>2</sup>Snob and bandwagon effects are often linked to status-seeking in conspicuous and luxury consumption. Independent self-concept, a view of the self as a unique and autonomous individual, is a key mediator of such behavior (Kastanakis and Balabanis, 2012, 2014). This psychological trait makes individuals more likely to exhibit snob effects, as they seek to differentiate themselves from others. Since experienced MovieLens users may plausibly have developed a stronger independent self-concept through their platform engagement, they should be more likely to engage in snob behavior, and yet we observe the opposite.

focused on their positive impact on sellers' performance and market outcomes (Dranove and Jin, 2010). In digital contexts, Elfenbein, Fisman and McManus (2015), and Hui, Saeedi, Shen and Sundaresan (2016) find certification programs on eBay improve seller quality, while Farronato, Fradkin, Larsen and Brynjolfsson (2024) report that licensing status in an online platform for home services does not increase transactions or prices. Research on how quality disclosures influence consumer experience is still relatively limited. Kovács and Sharkey (2014) and Lagios and Méon (2023) examine the impact of book awards, while Gergaud et al. (2015) and Li et al. (2024) study the effect of Michelin stars on restaurants. However, no study has yet examined how awards influence consumer ratings in the movie industry. Moreover, previous research has documented a decline in ratings for books and restaurants post-award, but these studies have not fully disentangled disappointment from user selection or potential strategic responses by sellers. Our study seeks to credibly isolate the disappointment effect, which is essential to assess the true welfare impact of quality disclosures and, more broadly, to understand how expectations shape consumers' satisfaction.

Empirical evidence of reference-based preferences is scarce outside the realms of finance and choice under uncertainty (Barberis, 2013). The work closest to ours is by Rey-Biel, Gneezy, Gneezy, Lauga and Llull (2024), who study theatergoers attending a "pay-what-you-want" performance. Using pre- and post-show surveys, they relate individuals' payments to the gap between expected and realized enjoyment. They find that this expectation gap best predicts how much individuals choose to pay. Unlike our approach, their study does not exploit external shocks to expectations, but documents how deviations from expectations shape willingness to pay in a natural setting. We build on this by demonstrating how reference points interact with quality disclosures in a quasi-experimental setting. Similarly, Backus, Blake, Masterov and Tadelis (2022) examine the role of expectations in an online context, finding that eBay bidders who lose to a "Buy-It-Now" option are more likely to leave the platform, attributing this behavior to disappointment.

We also contribute to the growing body of literature on potential biases in online reviews. Several studies highlight that user reviewing behavior does not necessarily reflect only product quality. Reviews can be directly manipulated by sellers (Anderson and Simester, 2014; Luca and Zervas, 2016; He, Hollenbeck and Proserpio, 2022), reviewers may act strategically (Moe and Schweidel, 2012; Nosko and Tadelis, 2015) or be influenced by social comparison (De Langhe, Fernbach and Lichtenstein, 2016; Wang, Zhang and Hann, 2018; Sunder, Kim and Yorkston, 2019) and user reciprocity (Proserpio, Xu and Zervas, 2018; Fradkin, Grewal and Holtz, 2021). In the context of movie reviews, Lee, Hosanagar and Tan (2015) demonstrates that online reviewers' behavior is influenced by prior ratings, particularly when their friends have rated a movie. Our paper provides additional evidence on the role of "external" factors influencing reviewers' behavior, as well as the impact of users' expectations.

Previous research has recognized the potential of recommendation algorithms in addressing identification challenges in panel data settings (Athey and Imbens, 2019; Dell, 2024). However, to the best of our knowledge, we are the first to integrate embeddings derived from a recommendation algorithm into a matching strategy. By treating the user embeddings generated by the algorithm as representations of individual preferences, we effectively address selection concerns arising from quality disclosures. This method holds promise for numerous applications involving large datasets, particularly those characterized by sparsity. For instance, our approach could be applied to policy evaluation questions using bank transaction data, or labor economics studies involving firm-to-firm transitions. Additionally, it is particularly well-suited for studying platform-related challenges, such as evaluating the effectiveness of platform design choices, or analyzing the impact of user-generated content on user engagement in social networks.

The paper proceeds as follows. Section 2 provides background on MovieLens and describes the dataset. Section 3 outlines the theoretical framework. Section 4 presents the initial difference-in-differences (DiD) analysis, comparing ratings of nominated and non-nominated movies before and after the AMPAS nominations. Section 5 introduces the recommendation-based matching approach and reports the main results. Section 6 extends the analysis using IMDb data. Section 7 turns to a textual analysis to reinforce the interpretation of the post-nomination decline in ratings as a disappointment effect. Section 8 concludes.

## 2 Empirical Setting and Dataset

In this Section, we introduce the movie recommendation platform, MovieLens, and outline the timing and functionality of its rating process. We then describe the “MovieLens 25M” dataset and explain how we enhance it with information about AMPAS nominations and awards. Finally, we present descriptive statistics for nominated and non-nominated movies: we compare these statistics before and after the nomination announcements, along with an analysis of the types of users who watch and rate nominated movies before and after the nominations.

### 2.1 MovieLens and the 25M Dataset

MovieLens is an online movie recommendation platform managed by GroupLens, a research group in the Department of Computer Science and Engineering at the University of Minnesota. Registered users rate movies, and these ratings inform a personalized list of recommended films generated by the MovieLens algorithm. Launched in 1997, MovieLens has become a widely used recommendation system. Its data has supported research in computer science, economics, and management, contributing to numerous studies on

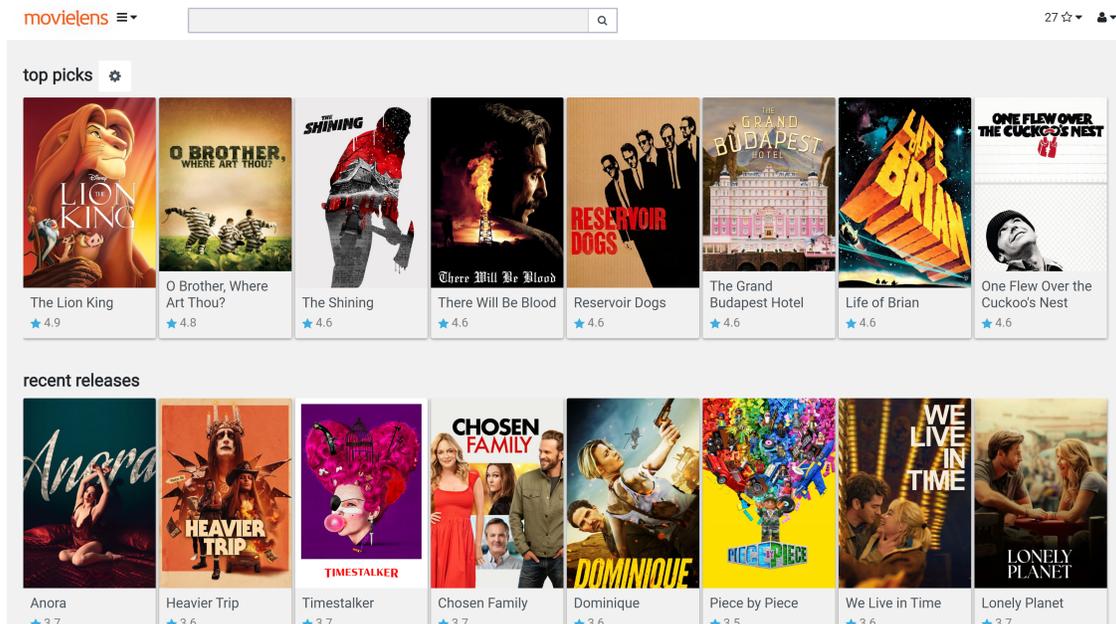


Figure 1. Snapshot of MovieLens.org: Main Page

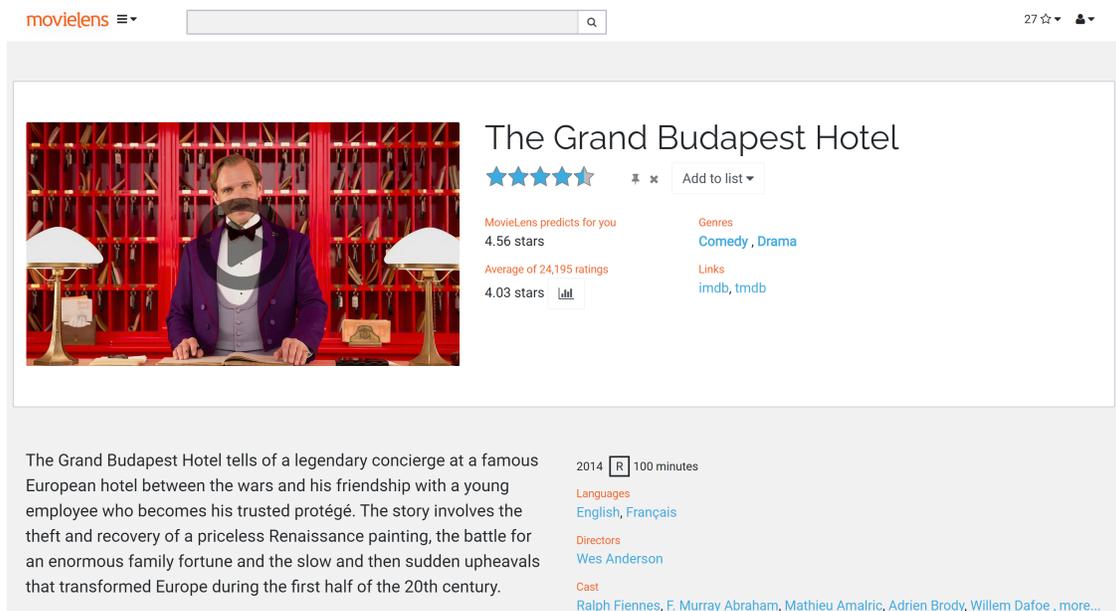


Figure 2. Snapshot of MovieLens.org: Movie Page

recommender systems and user rating behavior (Chen, Harper, Konstan and Li, 2010; Aridor et al., 2024a; Bondi, Rossi and Stevens, 2024). Although the platform’s design has evolved, the core rating mechanism has remained largely unchanged.<sup>3</sup>

The current rating process works as follows. Figure 1 shows a snapshot of the MovieLens homepage, where a selection of movies is listed. Users can search for specific films via the search bar. Clicking a movie title redirects the user to the film’s detailed page (Figure 2), where further information is available. Users rate movies on a 0.5 to 5-star scale. Unlike platforms like IMDb, Rotten Tomatoes, or Letterboxd, where user ratings are publicly visible, MovieLens keeps individual ratings private. These ratings are used to improve the platform’s recommendation algorithm, rather than for public display or social interaction.

For our analysis, we use the “MovieLens 25M” dataset (publicly available), released in December 2019. It includes 25 million ratings for 62,423 movies, given by 162,541 users between January 1995 and November 2019. Each rating is linked to both the movie’s and the user’s ID, along with the date of the rating. The dataset represents a random sample of users who rated at least 20 movies with no user-specific details. The dataset includes some movie attributes (e.g., genre, release year, IMDb ID). Additionally, we enriched the dataset by scraping IMDb for additional details, including the director, lead actors, language, and release date.

## 2.2 AMPAS Nominations and Awards

Since 1929, the Academy of Motion Picture Arts and Sciences (AMPAS) has awarded excellence in cinematic achievements across various categories.<sup>4</sup> However, the so-called “Big Five” categories typically receive the most public attention: Best Picture, Best Director, Best Leading Actor, Best Leading Actress, and Best Screenplay.

The award process involves two main steps each year. First, a set of nominated movies, often five per category, is selected from those that qualify. To qualify, a movie must have premiered in Los Angeles County in the previous calendar year.<sup>5</sup> Since 2004, nominations have been announced in mid-January, whereas prior to 2004, the announcement took place in early February. The awards ceremony follows six weeks after the nominations, held at a major theater in Hollywood.<sup>6</sup>

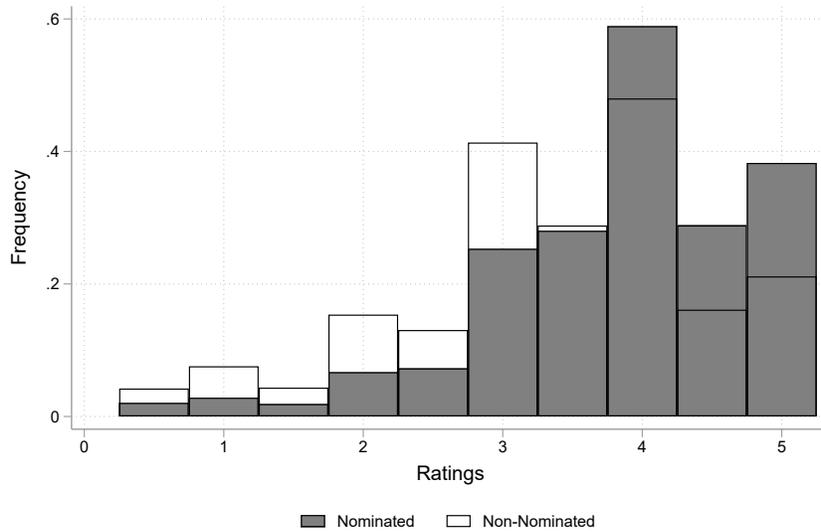
---

<sup>3</sup>Harper and Konstan (2015) provides a detailed account of the changes in the recommendation algorithm and rating process over time. Notably, the most significant change to the ratings interface occurred with the launch in 2003, when the system shifted from whole-star to half-star ratings. Despite this change, no visible global shifts in rating distributions were observed.

<sup>4</sup>Some categories have been discontinued, but since 2001, the awards have consistently featured twenty-four categories.

<sup>5</sup>More details on the selection process can be found on the official AMPAS website (<https://www.oscars.org/oscars/rules-eligibility>).

<sup>6</sup>In this analysis, we focus on AMPAS nominations rather than the awards themselves, as we show that users’ expectations are influenced by the nominations, causing a negative shift in their ratings. Concentrating on the awards would be more problematic,



**Figure 3.** Difference in Ratings Between Nominated and Non-Nominated Movies

*Note:* The figure shows the distributions of ratings for nominated and non-nominated movies. Nominated movies (in gray) have higher ratings than non-nominated movies.

To assess the impact of AMPAS nominations, we analyze all nominated and awarded movies from 1995 to 2019, focusing on the “Big Five” categories and the “Best Animated Feature Film” award. This yields 428 nominated films, of which 427 are present in the MovieLens dataset.<sup>7</sup> We exclude two animated films that only received reviews after their nomination and none prior. For each year, we use the nomination date to distinguish whether a movie’s ratings on MovieLens occurred before or after its nomination.<sup>8</sup> To be able to use non-nominated movies for our analysis, we assign each a reference nomination date: we use the date of the first rating shown on MovieLens to select the first nomination ceremony after that rating as the reference year. Finally, we exclude any movies whose first rating on the platform appears more than two years after their production year, ensuring older films are not compared with more recent ones.

### 2.3 Descriptive Statistics

Before analyzing the impact of nominations and isolating the disappointment effect in user ratings, we first present some summary statistics. These include a comparison of ratings for nominated and non-nominated movies, and how these ratings change over time, focusing on the period before and after the nominations. Finally, we examine the profiles of users who rate movies before and after the nominations.

as it would involve studying users who have already been impacted by a disclosure of the film’s perceived quality.

<sup>7</sup>The missing film is the animated movie *Ernest & Celestine*.

<sup>8</sup>Nomination dates are sourced from the official AMPAS website and related Wikipedia entries.

**Table 1.** Summary Statistics: Nominated Movies

	Mean	SE	Min	Max
<i>Average Ratings</i>				
All	3.704	.011	2.753	4.284
Before Nom.	3.841	.0158	2.286	4.833
After Nom.	3.697	.0111	2.747	4.284
<i>Number of Ratings</i>				
All	8193	519.47	59	57,309
Before Nom.	374.3	31.2	1	4,986
After Nom.	7819	503.529	53	57,303
<i>Average Ratings (80-days window)</i>				
All	3.826	.015	2.732	4.75
Before Nom.	3.832	.0166	2.286	4.833
After Nom.	3.823	.0157	2.5	4.857
<i>Number of Ratings (80-days window)</i>				
All	264	13.71	2	1,832
Before Nom.	126.5	7.523	0	1,032
After Nom.	137.4	6.82	2	927

*Note:* The table presents information on all 426 movies nominated for AMPAS awards included in our sample. We report the average ratings and the number of ratings both before and after the nominations. The first two panels include all available ratings without any time restrictions, while the second two panels focus on ratings within an 80-day window surrounding the nominations.

Figure 3 shows the relative frequency with which nominated and non-nominated movies receive each rating level. We observe that a large proportion of ratings for both categories fall between 3 and 4. However, nominated movies have a 5-10 percentage point higher share of ratings at 4, 4.5, and 5 compared to non-nominated movies, which are more likely to be rated 3 or below.

For both groups, the ratings, rather than exhibiting a j-shape, follow a distribution that is roughly normal but with a slight rightward skew. The absence of j-shaped distributions in our dataset supports the notion that *extremity* bias, where users often rate products only when they are either highly satisfied or dissatisfied (Dellarocas and Narayan, 2006; Hu, Pavlou and Zhang, 2009, 2017), is less likely to occur on information platforms with a large scale of ratings, such as MovieLens (Schoenmueller, Netzer and Stahl, 2020).

Movie ratings are not static; they can vary over time, particularly in response to significant events like AMPAS nominations. In Table 1, we show that nominated films tend to receive slightly higher average ratings prior to their nomination, suggesting more favorable initial audience perceptions. This pattern may reflect a sense of disappointment after the nominations, but it could also result from selection effects: different types of users rating films before versus after the announcement.

We also observe an increase in the number of ratings following the nominations, indicating heightened audience engagement. However, the top panel of Table 1 compares ratings across full pre- and post-nomination periods, and the longer post-nomination window likely inflates this increase. To account for this, the bottom panel reports the same statistics using symmetric 80-day windows before and after the nomination. In this restricted window, we still observe a modest drop in average ratings and a clear rise in the number of ratings, consistent with a surge in popularity immediately following the nominations.

Appendix Table A1 presents analogous statistics for non-nominated movies. These films, on average, receive lower ratings than nominated ones. Their average ratings remain relatively stable around the reference date, with no marked change following the nominations. While the top panel shows a clear post-nomination increase in the number of ratings, this again reflects the longer duration of the post-period. In the balanced 80-day comparison, the average rating remains roughly constant, or even slightly increases, while the number of ratings declines.

Appendix Figure A1 shows the daily arrival of ratings around the nomination period for both nominated and non-nominated films. For nominated movies, we observe a positive time trend after nominations, characterized by an increase in the number of ratings, especially around weekends. In contrast, non-nominated movies do not exhibit a similar trend, although we do observe a noticeable influx of ratings in the weeks leading up to the nominations.

## 2.4 User Selection Before and After the Nominations

As the number of ratings varies over time, the set of users who post ratings for nominated films may change before and after the AMPAS nominations. To explore the self-selection of users into watching and rating nominated movies, we focus on four dimensions that might describe users’ characteristics and the determinants of their rating behavior, namely their average rating, their experience on the platform, the types of movies they rate, and their rating stringency.

For each user  $i$ , we define the average user rating  $\bar{r}_i^j$  before rating movie  $j$  as the average of the ratings that user  $i$  posted before rating movie  $j$ . We also record the total number of ratings user  $i$  posted before rating movie  $j$ , denoted as  $N_i^j$ . These two variables captures users’ experience at the time they rated the movies and the types of movies they rated. In line with Bondi et al. (2024), we measure each user’s stringency relative to others. To do so, we calculate the rating stringency as  $s_{ij} = \bar{r}_j - r_{ij}$ , where  $\bar{r}_j$  is the average rating for movie  $j$  across all users, and  $r_{ij}$  is the rating given by user  $i$  for that movie. We define the stringency of user  $i$  in rating movie  $j$ , denoted as  $\bar{s}_i^j$ , as the moving average of the stringency of each user over their previous 10 ratings. Positive values of  $\bar{s}_i^j$  indicate lower than average ratings, suggesting greater stringency.

**Table 2.** User Selection Before and After the Nominations: 80-days window

	Mean	SE	Min	Max
<i>Average User Ratings</i>				
All	3.451	.0026	1.207	5
Before Nom.	3.45	.0032	1.038	5
After Nom.	3.463	.0032	1.14	5
<i>Number of User Ratings</i>				
All	752.9	7.26394	2	11,354.3
Before Nom.	725.3	7.715	0	14,959
After Nom.	764.4	11.3181	1	25,034
<i>Average User Stringency</i>				
All	.0915	.0033	-2.204	3.161
Before Nom.	.0943	.0039	-2.204	3.204
After Nom.	.077	.0042	-2.855	3.161
<i>Average Share of User Ratings for Current-Year Releases</i>				
All	.052	.00039	0	.6667
Before Nom.	.0585	.000499	0	1
After Nom.	.0451	.000447	0	.6667

*Note:* The table presents descriptive statistics of users who rated movies nominated for AMPAS Awards included in our sample. We only included users who submitted a rating within an 80-day window surrounding the nominations. We report the average user rating, the average number of ratings per user, the average rating stringency (a measure of how strictly users rate, based on how their ratings compare to the average rating of the movie at the time), and the average share of ratings for movies produced in the same year as the rating. The stringency of a user is defined as a moving average of the stringency of their past five ratings. These statistics are presented separately for users who rated the movies before and after the nominations.

Finally, we compute the share of movies released in the same calendar year as movie  $j$  that user  $i$  had already rated before rating movie  $j$ . This share provides information about users' interest in recent movies and their likelihood to rate movies that have been or are expected to be nominated for AMPAS awards.

In Table 2 we present a summary of these variables for users who rated nominated movies within an 80-day window surrounding the nominations.<sup>9</sup> In terms of average user ratings, we observe a slight increase after nominations, suggesting that the quality signal provided by AMPAS nominations may attract users who are more likely to give higher ratings on average. This could be because these users are generally more positive in their ratings, or because they tend to select higher quality movies.

Additionally, the total number of ratings a user has posted before reviewing a nominated movie is higher after the nominations.<sup>10</sup> We also observe a decrease in user stringency after the nominations: those users who rate nominated movies after the nominations are less critical and generally post higher-than-average

<sup>9</sup>Appendix Table A2 presents a similar analysis, but considering the full history of ratings.

<sup>10</sup>This variable is computed as the average number of ratings users posted prior to each nominated movie, which is why its maximum value may not be an integer.

ratings. Finally, the share of ratings for current year releases decreases after the nominations, suggesting that users who rate nominated films after the nominations are generally less inclined to rate recent films.

All of this evidence points to a shift in the profile of users before and after the nominations. Those who rate nominated movies before they receive AMPAS nominations rate fewer movies, tend to be more stringent, and are more active in rating recent releases.

It is important to note that we observe the timing of the rating, not the actual moment when a user watches a movie. Therefore, ratings posted before the nomination are necessarily unaffected by the nomination itself. Conversely, for ratings posted after the nomination, we cannot rule out the possibility that some users watched the movie prior to the nomination. Nonetheless, even if some users watched the movie before the nomination, the nomination could still influence their decision to rate the movie and the rating they assign, for example by increasing the salience of the movie or shifting their evaluation criteria.<sup>11</sup> To understand how this selection may affect our identification strategy, in the next Section we present a conceptual framework in which we model users self-selecting into movie viewing.

### 3 Conceptual Framework

To guide the empirical analysis of how users experience disappointment following AMPAS nominations, we propose a conceptual framework. This framework clarifies our assumptions regarding the movie rating process and the mechanisms through which quality signals influence user ratings. It also helps illustrate how different empirical strategies can isolate the distinct effects of nominations on user evaluations. Throughout this framework and in the empirical Sections that follow, we interpret any change in users’ utility after a nomination as disappointment. In Section 7, we return to this assumption and use textual evidence to show that disappointment is indeed the primary channel at play, rather than alternative mechanisms.

#### 3.1 The Decision to Watch: User Utility, Quality Signals, and Ratings

Each movie is defined by a quality parameter  $\theta_j \in \mathbb{R}$  and a vector  $\mu_j \in \mathbb{R}^M$ , which describes the movie’s “position” in an  $M$ -dimensional feature space (e.g., the movie’s genres, actors, length, etc.). Similarly, each user is characterized by  $v_i \in \mathbb{R}$ , which captures user-specific baseline enjoyment of movies or their individual use of the rating scale (e.g., some users may rate more stringently or generously), and a vector  $\pi_i \in \mathbb{R}^M$ , which describes user  $i$ ’s preferences in the feature space (e.g., their likes and dislikes for specific genres, actors, screenwriters, etc.).

---

<sup>11</sup>In Appendix B, we discuss and articulate two competing hypotheses regarding how such signals might affect evaluations when they arrive after viewing but prior to rating.

We assume that users are aware of their own position in the feature space as well as the positions of the movies. However, before watching movie  $j$ , users do not perfectly observe its quality. Thus, the expected utility for user  $i$  when watching movie  $j$  is defined as follows:

$$E(u_{ij}) = E(\theta_j) + v_i - d(\boldsymbol{\pi}_i, \boldsymbol{\mu}_j), \quad (1)$$

where the function  $d(\cdot)$  is a strictly increasing, continuous norm that describes the distance between the position of movie  $j$  and the preferences of user  $i$ .<sup>12</sup> Accordingly, we can rewrite the expected utility as  $E(u_{ij}) = E(\theta_j) + \alpha_{ij}$ , where  $\alpha_{ij} = v_i - d(\boldsymbol{\pi}_i, \boldsymbol{\mu}_j)$  represents the sum of all observed parameters: the utility from going to the theater, and the utility “match” value between user  $i$  and movie  $j$ . Thus, we can interpret  $E(\theta_j)$  as the component of the experience attributes that cannot be observed prior to watching the movie, such as the quality of acting, cinematography, or direction, while  $\alpha_{ij}$  captures the utility derived from search attributes that are observable beforehand, such as the identity of the director and actors or the movie’s genre (Moorthy, Ratchford and Talukdar, 1997).

After watching movie  $j$ , user  $i$  observes its quality. Ex-ante expectations may enter into the ex-post utility of users for various reasons. We believe that MovieLens users are unlikely to incorporate biases related to groupthink or social signaling into their ratings, as individual user ratings are not shared with others, and there is no visible link between a user’s identity and their ratings on the platform. Yet, the expected utility  $E(u_{ij})$  may serve as a reference point and affect user  $i$  ex-post satisfaction. Thus, we define the ex-post utility for user  $i$  after watching movie  $j$  as follows:

$$u_{ij} = \theta_j + \alpha_{ij} + \gamma(\theta_j + \alpha_{ij} - E(u_{ij})) = \theta_j + \alpha_{ij} + \gamma(\theta_j - E(\theta_j)), \quad (2)$$

where  $\gamma(\cdot)$  represents the user’s gain-loss utility factor as in Kőszegi and Rabin (2006). We assume  $\gamma(\cdot)$  to be linear with the parameter  $\gamma > 0$  multiplying the gain/loss term. Consequently, the higher user  $i$ ’s expectation about the quality of movie  $j$ ,  $E(\theta_j)$ , the lower the gain-loss utility term, and the lower user  $i$ ’s ex-post utility.

For each movie  $j$ , the following timing describes the process through which users are informed about the movie, they decide to watch it and, in such a case, rate it:

1. Movie  $j$  is released and a unit measure of potential users associated with movie  $j$  is formed;
2. A signal  $s$  appears and a proportion of users becomes aware of movie  $j$ . The distribution of  $\alpha_{ij}$  among

---

<sup>12</sup>We use a distance-based formulation to capture the dissimilarity between user preferences and movie attributes. Yet, this is not essential to the model. Any function that maps preferences and attributes into utility could be used. We adopt this distance formulation for consistency with the recommender system presented in Section 5, which relies on proximity in a shared embedding space.

aware users after signal  $s$  at time  $t$  is  $F_{s,t}(\alpha)$ ;

3. Aware users form expectations about the quality of the movie  $E(\theta_j|s)$  and watch movie  $j$  if  $E(u_{ij}|s) = E(\theta_j|s) + \alpha_{ij} > 0$ ;
4. Finally, users who rate movie  $j$  are randomly drawn from the set of users who watched it, and they report their ex-post utility in their ratings:  $r_{ij} = u_{ij}$ .<sup>13</sup>

### 3.2 The Impact of the Nominations for the Academy Awards

A positive quality signal like the AMPAS nominations alters the profile of users aware of the movie (i.e.,  $F_{s,t}(\alpha)$  depends on  $s$ ) and influences the quality expectations of all users aware of the movie. These two effects impact the ratings before and after the nominations. Specifically, the expected rating of movie  $i$ , given its quality  $\theta_i$  and signal  $s$ , is as follows:

$$E(r_{ij}|\theta_j, s) = \theta_j + \int_{\alpha > -E(\theta_j|s)} \alpha, dF_{s,t}(\alpha) + \gamma(\theta_j - E(\theta_j|s)). \quad (3)$$

To better understand how a shift in expected movie quality affects ratings, let us assume the signal changes from  $s$  to  $s^+$ , where  $E(\theta_j|s) < E(\theta_j|s^+)$ . The difference in expected ratings for the same movie  $j$  but different signals is given by:

$$\begin{aligned} E(r_{ij}|\theta_j, s^+) - E(r_{ij}|\theta_j, s) &= -\gamma(E(\theta_j|s^+) - E(\theta_j|s)) \\ &\quad + \int_{\alpha > -E(\theta_j|s^+)} \alpha, dF_{s^+,t}(\alpha) - \int_{\alpha > -E(\theta_j|s)} \alpha, dF_{s,t}(\alpha). \end{aligned} \quad (4)$$

The first term,  $-\gamma(E(\theta_j|s^+) - E(\theta_j|s))$ , represents the *disappointment effect*, as it captures the negative impact on ratings due to increased expectations when  $\gamma > 0$ . This effect is present across all users and its magnitude depends on the size of  $\gamma$  and the variation of quality expectations due to the change in signals.

The second term,  $\int_{\alpha > -E(\theta_j|s^+)} \alpha, dF_{s^+,t}(\alpha) - \int_{\alpha > -E(\theta_j|s)} \alpha, dF_{s,t}(\alpha)$ , represents the *selection effect*. It reflects the change in the profile of users who decide to watch movie  $i$  due to the shift in the signal. While disappointment always drives ratings downward (for  $\gamma > 0$ ), the effect of selection is more ambiguous. If  $F_{s,t}(\alpha) = F_{s^+,t}(\alpha)$ , selection will reduce the average ratings: users who are willing to watch the movie after observing  $s^+$  are either those who derive less utility from the movie's observable search attributes (e.g., they

<sup>13</sup>Users' decision to rate may depend on their ex-post utility levels. Specifically, users who rate products online may be more likely to have experienced either extremely positive or extremely negative utility levels from the transaction (Dellarocas and Narayan, 2006; Hu et al., 2017; Schoenmueller et al., 2020). In the next Subsections, we show how changes in quality signals affect aggregate ratings and discuss the role of this assumption (no selection into reviewing).

like the director or the genre less), or those who place relatively less weight on search attributes compared to experiential quality.

If  $F_{s,t}(\alpha) \neq F_{s^+,t}(\alpha)$ , the selection effect could potentially drive ratings upward. The sign of the selection effect becomes even more challenging to predict if we assume some degree of selection into reviewing. In fact, if different types of users choose to rate a movie depending on the signal they receive and the ex-post utility they experience (Dellarocas and Narayan, 2006), it becomes much more likely that  $F_{s,t}(\alpha) \neq F_{s^+,t}(\alpha)$ , further complicating the prediction of the selection effect. However, the selection into reviewing does not directly affect the disappointment effect, which primarily depends on the gap between users’ expectations and the actual quality of the movie, regardless of the profile of users who post ratings.

### 3.3 Isolating the Disappointment Effect

A comparison of average ratings for the same movie under different quality signals does not allow us to isolate disappointment from selection, as shown in Equation 4. Nevertheless, this approach offers valuable insights into the overall effect of quality signals on user satisfaction. In Section 4, we begin by exploring the empirical counterpart of this analysis with a DiD approach.

To isolate the disappointment effect from selection, we must observe the user-movie features  $\alpha_{ij}$  and hold them constant when comparing ratings with and without nominations:

$$E(r_{ij}|\theta_j, \alpha_{ij}, s^+) - E(r_{ij}|\theta_j, \alpha_{ij}, s) = -\gamma(E(\theta_j|\alpha_{ij}, s^+) - E(\theta_j|\alpha_{ij}, s)). \quad (5)$$

By comparing users with identical  $\alpha_{ij}$  values, we can fully isolate disappointment. Moreover, matching on  $\alpha_{ij}$  also addresses selection into reviewing, as it controls for the same user-movie fit that drives both viewing and reviewing decisions. The empirical analog of this method, based on recommendation-based matching, is presented in Section 5.

Our framework assumes that quality signals arrive either before both the watching and rating decisions or after both, but never in between. In practice, users may receive new signals, such as award nominations, after watching but before rating a movie. We explore the implications of this intermediate timing in Appendix B.

## 4 Comparing Ratings for Nominated and Non-Nominated Movies

To assess how a shift in expected movie quality influences users’ satisfaction, we can compare the ratings of users who rated nominated movies before and after the nomination date. This approach mirrors the idea in

Equation 4 from the previous Section, where we aim to compare ratings of the same movies under different quality signals. In practice, however, comparing average ratings before and after nomination simultaneously captures both the impact of the AMPAS nomination as a quality signal and any temporal differences.<sup>14</sup> Specifically, we compare ratings posted around January and February, before the nominations, with those posted later in February and March, after the nominations.

As our theoretical framework shows, the type of users who is aware of a movie might change over time. To account for the fact that fans of the actor, director, or genre might be among the first to watch a movie, we implement a DiD design, comparing ratings for nominated and non-nominated movies before and after the nomination date.

Each movie appears in the estimation sample only once, either as a treated movie (if nominated) or as a control (if not nominated). We assign each movie an 80-day window centered around the AMPAS nomination date. For non-nominated films, we assign a reference “nomination date” by identifying the date of the first rating on MovieLens and matching it to the first AMPAS nomination ceremony that occurs after that date. Accordingly, both nominated and non-nominated films are included in the sample only for the ratings posted in the 80-day window of the year associated with their assigned AMPAS nomination date, ensuring a clear separation between treated and control groups.

To capture the impact of the AMPAS nominations on nominated movies, we use the following DiD specification:

$$r_{jt} = \gamma_1 + \gamma_2 \text{Nominated}_j + \gamma_3 \text{post}_t^{\text{Nom}} + \delta \text{Nominated}_j \times \text{post}_t^{\text{Nom}} + \beta \mathbf{X}_{jt} + \varepsilon_{it}, \quad (6)$$

where  $r_{jt}$  is the rating of movie  $j$  on day  $t$ .  $\text{Nominated}_j$  equals 1 if movie  $j$  has received at least one AMPAS nomination; and  $\text{post}_t^{\text{Nom}}$  equals 1 for all ratings posted after the nominations. The set of controls  $\mathbf{X}_{jt}$  includes genre, year, movie, and day fixed effects, where the day fixed effects capture all variations in ratings that posted on specific days relative to the nomination date.

Table 3 shows four specifications for the DiD estimates in Equation 6. The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. We cluster standard errors at the movie level.

In Column (1), we do not use any fixed effects, while we add fixed effects for the genre and the production year of the movies in Column (2). In Column (3), we add movie fixed effects that absorb all time-invariant movie characteristics. Finally, Column (4) presents the results with movie and day fixed effects,

---

<sup>14</sup>Recall that the distribution of  $\alpha_{ij}$ ,  $F_{s,t}(\alpha)$ , is a function of signal and time.

accounting for seasonality and time-varying confounders that affect all movies uniformly. The estimates of the coefficient  $\delta$  in Equation 6 are negative and significant across all specifications. The drop in ratings posted for nominated movies after the nominations is approximately 0.02 stars. This drop may seem small compared to the average rating (around 3.5); however, in Columns (1) and (2), we observe that the difference in ratings between nominated and non-nominated movies (before nominations) is less than 0.6 stars. Thus, the short-term negative effect linked to the quality signal from nominations is approximately 5% of the “rating premium” experienced by nominated movies, which may reflect either the higher quality of these movies or the characteristics of users who rate them.

The coefficient  $\delta$  captures the effect of the AMPAS nominations on nominated movies under the assumption that non-nominated movies provide a valid counterfactual group for the evolution of ratings after nomination. To check for potential pre-trends between treated and control groups, we illustrate the evolution of  $r_{jt}$  over time with an event-study approach. We specify the following lead-lag model, in which  $r_{jt}$  is regressed over the interaction between the dummy  $Nominated_j$  and a set of dummy variables for each 5-day window around the AMPAS nomination dates. The model controls for movie and day fixed effects:

$$r_{jt} = \gamma_j + \rho_t + \sum_{\tau=-40 \text{ days}}^{40 \text{ days}} \delta_{\tau} Nominated_j \times 1(t = \tau) + \varepsilon_{jt}. \quad (7)$$

We present the estimates of Equation 7 in Figure 4, plotting the estimated  $\delta_{\tau}$  from 40 days before to 40 days after the AMPAS nominations, excluding ratings posted after the AMPAS award ceremony. The coefficients for days prior to the nominations are close to zero and show no clear trend, suggesting that the evolution of ratings for nominated and non-nominated movies was similar prior to the nominations. This finding supports the parallel trend assumption, which is essential for our analysis. In contrast, after the AMPAS nominations, particularly around 15 days after, the estimated  $\delta_{\tau}$  turns negative, indicating a relative decline in ratings for nominated movies compared to non-nominated ones.

The observed lag in the ratings decline may be due to the delay between when users watch a movie and when they post their ratings. Since we observe only the rating date, but not the viewing date, it is likely that some users who rate the movie shortly after the nominations had actually watched it before the AMPAS announcement and were therefore not influenced by it.<sup>15</sup> To account for this lag, Appendix Table C1 presents the same specification as in Column (4) of Table 3, excluding ratings posted 5, 10, and 15 days after the nominations. When we omit these ratings (as they may not fully reflect treatment effects), the coef-

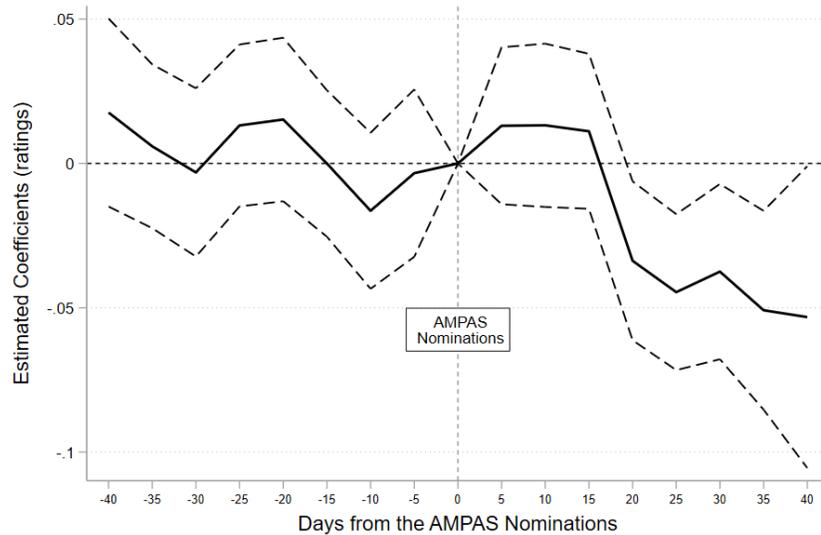
---

<sup>15</sup>In our model in Section 3, we do not account for the possibility that the nomination signal may arrive between the viewing and the rating decisions. Appendix B discusses how users’ ratings may differ depending on whether they learn about the nomination before or after watching the movie.

**Table 3.** Difference-in-Differences: Comparing Nominated and Non-Nominated Movies

	(1)	(2)	(3)	(4)
$Nominated_j$	0.590*** (0.023)	0.575*** (0.024)		
$post_t^{Nom}$	0.007 (0.005)	0.006 (0.005)	-0.005 (0.003)	
$Nominated_j \times post_t^{Nom}$	-0.034*** (0.011)	-0.037*** (0.011)	-0.019*** (0.007)	-0.018** (0.007)
Genre FEs		✓		
Year FEs		✓		
Movie FEs			✓	✓
Date FEs				✓
$R^2$	0.05	0.06	0.20	0.21
N	497,266	497,266	497,266	497,266
Mean Dep. Var.	3.503	3.503	3.503	3.503

Note: The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. Standard errors, clustered by movie, are shown in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Figure 4.** Event study: Comparing Nominated and Non-Nominated Movies

Notes: The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. MovieLens ratings are regressed on movie fixed effects and on interactions between a dummy variable (indicating whether a movie is nominated for an AMPAS award) and a set of five-day interval dummies centered on the AMPAS nomination date. The graphs display the estimated coefficients of these interactions, with the coefficient for the five-day interval of the nomination date normalized to zero. Standard errors, clustered by movie, are reported at the 10% level.

ficient  $\delta$  doubles in magnitude, consistent with the lag observed in the event study.

The parallel trends in Figure 4 suggest that the ratings for nominated and non-nominated films evolved similarly prior to the AMPAS nominations. However, not all non-nominated films may provide an appropriate counterfactual for nominated films, which tend to be of higher quality and generally receive higher ratings (as discussed in Section 2.3). To examine whether our results are sensitive to the selection of non-nominated films, Appendix Table C2 presents three alternative specifications that select the pool of non-nominated films. We restrict the analysis to movies with at least one professional review on the Metacritic website,<sup>16</sup> those with an average Metacritic score above 50 (out of 100), and those with an average IMDb rating above 5. By refining the control group to exclude non-nominated films with lower visibility and quality, we improve the comparability of nominated and non-nominated films. All specifications confirm a significant negative effect of nominations on the ratings of nominated films, suggesting that the observed ratings drop is not highly sensitive to the selection criteria for non-nominated films in the control group.

In Appendix Table C3, we examine how the results vary across different groups of nominated movies to identify which groups drive the observed effects. Our findings indicate that the negative effect of nominations is particularly pronounced for movies that receive more than one Oscar nomination, regardless of whether they ultimately win an AMPAS award. This suggests that disappointment is more likely to be triggered when a certification signals exceptionally high quality, raising consumers' expectations substantially, but the real quality of the product may not fully live up to those elevated expectations. For a nomination to significantly alter users' expectations and induce disappointment, the signal must be both informative and widely recognized. By contrast, movies with only one nomination are less likely to experience this effect, as a single nomination may not substantially influence users' expectations.

As discussed above, this approach does not disentangle disappointment from the user selection effects associated with nominations; however, it provides valuable insights into the combined effects of selection and disappointment in this context. In the next Section, we introduce a recommendation-based matching approach to compare users with similar tastes and preferences for movies before and after the nominations. Complementary to this approach, we can control for observable user characteristics within the DiD framework used in this Section. This may not fully remove all user self-selection, but it allows us to assess whether the results in Table 3 are robust to the inclusion of user characteristics. This analysis is presented in Appendix Table C4, where we replicate the specification from Column (4) of Table 3 but add controls for users' average rating prior to rating movie  $j$ , the total number of ratings posted by user  $i$  before rating movie  $j$ , user  $i$ 's stringency and the proportion of current-year releases rated by user  $i$  prior to rating  $j$ . We also

---

<sup>16</sup>For more information, see: <https://www.metacritic.com/>.

include user fixed effects in the final specifications. Notably, the significance and magnitude of the effect remain stable if not slightly larger in the most restrictive specification.

This result may seem counterintuitive at first; we might expect users who rate nominated movies after the nominations to be drawn in by the advertised high quality, potentially making them more critical of the nominated movies. However, as shown in Section 2.4, users who rate movies after the nominations tend to be less stringent and less attentive to new releases. This may partly explain why, when controlling for various user characteristics, we observe an even larger negative effect of nominations on ratings.

## 5 A Recommendation-based Matching Design

As shown in Equation 5, to identify the disappointment effect associated with the AMPAS nominations for a nominated movie  $j$ , we must account for taste-based selection, i.e. the fact that the nominations can attract moviegoers with different preferences and characteristics. To implement this approach and condition on  $\alpha_{ij}$ , we use a recommendation-based matching design. Specifically, for each year, we train a recommendation system using only ratings posted before the AMPAS nominations in order to obtain a vector representation of users’ “tastes” (i.e., their positions  $\pi_i$  in the feature space). We then integrate these results into a matching approach, comparing ratings for the same nominated movie before and after the nomination date, matching on users’ tastes and other characteristics that capture variation in baseline utility  $v_i$ .

In the remainder of this Section, we first describe the recommendation system and its characteristics. Next, we present the matching design and the main results of our analysis, which reveal a negative and significant disappointment effect following the AMPAS nominations. As a placebo test, we use the predicted ratings generated by the recommendation system as the outcome instead of the actual ratings; in this case, the matching yields null results, suggesting that the matched users indeed have similar preferences for the nominated movies. Then, we apply the same matching analysis to non-nominated movies and find no evidence of disappointment. Finally, we examine the evolution of the difference between actual and predicted ratings over time.

### 5.1 Training the Recommendation System

To compute a representation of user tastes, we rely on the literature on collaborative filtering and matrix factorization recommender systems (Koren, Bell and Volinsky, 2009; Billsus, Pazzani et al., 1998; Sarwar, Karypis, Konstan and Riedl, 2000; Goldberg, Nichols, Oki and Terry, 1992; Paterek, 2007). Matrix factorization techniques aim to approximate the matrix of ratings as the product of two smaller matrices, one

representing users and the other representing items, in our case, movies. The two smaller matrices,  $P$  and  $Q$ , are of dimensions  $I \times M$  and  $J \times M$ , respectively, where  $I$  is the number of users,  $J$  is the number of movies, and  $M$  is the number of latent factors, which we set to 50. We employ a singular value decomposition algorithm as described in [Koren et al. \(2009\)](#), where the embedding matrices are optimized via stochastic gradient descent.<sup>17</sup>

The embedding matrices computed in this manner capture certain aspects of the items and users. The latent factors represent a computerized alternative to intuitive categories that might influence movie recommendations, such as a user’s preference for action, drama, or romance movies, or their affinity for specific actors. However, these embeddings are not directly interpretable ([Koren et al., 2009](#)). Importantly, two users with identical embeddings can be said to share the same tastes since they will have the same *predicted* ratings for all movies and receive the same recommendations from the algorithm.

This approach allows us to match users based on their embeddings, thus accounting for user tastes and isolating the disappointment effect from selection. For implementation, we use the singular value decomposition as implemented in the `scikit-surprise` python library.<sup>18</sup> With this approach, ratings are modeled as follows:

$$\hat{r}_{ij} = \mu + b_i + b_j + q_j^T p_i,$$

where  $\mu$  is the mean rating, and  $b_i$  and  $b_j$  are user- and movie-specific bias terms. Here,  $p_i$  represents a vector of latent factors (or “embeddings”) for user  $i$ , while  $q_j$  represents a vector of latent factors for movie  $j$ . These latent factor vectors form the columns in the embedding matrices  $P$  and  $Q$ .

For each nomination year, we train the recommender system on all ratings up to 40 days before the nomination date, obtaining the user embeddings and predicted ratings. We then merge the user embeddings and the predicted ratings with the dataset containing user ratings and other user characteristics.

In Appendix D.1 we provide a detailed step-by-step explanation of how the user embeddings are constructed through a looping procedure that updates embeddings across years. To evaluate the performance of the recommender algorithm used to generate embeddings and predicted ratings, we conduct a series of vali-

---

<sup>17</sup>The same approach discovered by Simon Funk during the Netflix challenge. See: <https://sifter.org/simon/journal/20061211.html>.

<sup>18</sup>The `scikit-surprise` library is a well-known and widely used package for building and evaluating recommender systems, and has been applied in various empirical domains, including economics, management, and marketing. Its popularity and integration within the broader `scikit` ecosystem make it a practical choice for ensuring usability and reproducibility. Yet, `scikit-surprise` is not the only library available for implementing recommendation algorithms. In particular, the `LensKit` package offers a suite of algorithms explicitly designed and tuned for the MovieLens dataset ([Ekstrand, Riedl, Konstan et al., 2011a](#); [Ekstrand, Ludwig, Konstan and Riedl, 2011b](#)). In the remainder of this Section, we rely on the singular value decomposition algorithm as implemented in `scikit-surprise` to compute user and item embeddings. However, to ensure the robustness of our findings, we replicate our key analyses using alternative algorithms provided by `LensKit`, including matrix factorization (which provides embeddings) and user-user and item-item collaborative filtering (which do not produce embeddings, but do yield predicted ratings).

dation exercises detailed in Appendix D.2. In Appendix D.3, we further assess the quality of the embeddings by examining whether users who rate the same movies similarly also have more similar embeddings.

## 5.2 Matching Users Before and After the AMPAS Nominations

We now present the results of our matching analysis, comparing ratings for nominated movies posted 40 days before and after the AMPAS nominations. To isolate the effect of potential disappointment from selection, we match users based on their 50-dimensional embedding vectors and other attributes reflecting their preferences. These embeddings are designed to capture a comprehensive representation of users' tastes and behaviors. By matching on embeddings instead of predicted ratings, we ensure that the users being compared share not only similar preferences for the specific movies of interest but also for a broad range of genres and types of movies. Thus, the matched users have comparable likelihoods of watching and rating the movie after its nomination.

Because treatment is defined by the timing of each rating relative to the nomination date of the specific movie (rather than by the user alone) a single user can contribute both treated and control observations by rating different nominated movies at different times. We implicitly assume that watching a nominated movie before the announcement does not spill over to affect the user's utility or potential disappointment when rating a different nominated movie after nominations are revealed. This assumption is reasonable in our setting, where movies vary substantially in content and user expectations are primarily shaped at the title level. In contrast, for the placebo analysis involving non-nominated movies, we adopt a more conservative strategy: we exclude users who rated any nominated movie during the same nomination period. This ensures that we are analyzing both untreated movies and untreated users within that year, thereby providing a cleaner counterfactual by avoiding any potential exposure to the AMPAS nominations.

Table 4 presents OLS and propensity score matching estimates using various sets of variables to match users. Column 1 (OLS) reports estimates with movie fixed effects, comparing ratings of users who watched the same nominated movies without adjusting for user characteristics. Columns 2 to 6 (M1–M5) provide propensity score matching estimates, incorporating user embeddings (50 dimensions) and exact matching on movie identifiers, along with quintiles of different user attributes: average user ratings, the number of movies watched, stringency levels (as defined in Section 2.3), and the share of movies watched by the user in the year of their release.<sup>19</sup>

---

<sup>19</sup>The sample in Column 1 is substantially smaller than the sample used in the DiD analysis in Table 3 due to two main restrictions: we focus only on ratings of nominated movies, which reduces the sample from 497,266 to 106,090 observations; and, to construct meaningful user embeddings, we include only users who posted at least one rating prior to the 40-day pre-nomination window, further reducing the sample to 63,849 observations.

Across all specifications, the estimates reveal a statistically significant drop in ratings, with matching estimates being of a slightly larger magnitude compared to the OLS estimates. This difference may stem from the fact that more stringent users are disproportionately likely to watch nominated movies before the nominations.

In Appendix Tables D2, D3, D4, and D5, we show that the results are robust to different choices of embedding dimensionality, ranging from 10 to 200. In Appendix Table D6, we further demonstrate the robustness of our findings by replicating the analysis in Table 4 using user embeddings obtained from an alternative recommender system: the biased matrix factorization algorithm implemented in the `LensKit-TF` package.<sup>20</sup> The estimates in Appendix Table D6 closely mirror those in Table 4, reinforcing the conclusion that our results are not sensitive to the specific recommender system used to construct the embeddings.

The negative estimates are also not driven by users who rate movies shortly after the nominations or by users with few posted ratings, whose embeddings and predictions may be less accurate. As shown in Appendix Table D7, when we exclude ratings posted within 5, 10, and 15 days following the nominations, the estimates remain negative, significant, and of similar magnitude. In the same Table, we also address the possibility that many users post multiple ratings in a single session. To account for this behavior, we classify ratings into sessions based on calendar days (such that all ratings posted on the same date are considered part of the same session) and discard all ratings from the first session occurring after the nomination. The results of this exercise have very similar magnitude to the previous ones, further suggesting that this pattern of rating behavior does not confound our findings.

Finally, in the sample used for the matching analysis, only 1% of the ratings come from users who posted fewer than 44 ratings in total on the platform or fewer than 22 ratings prior to the rating for the nominated movie. This suggests that users with sparse rating histories are unlikely to meaningfully affect our estimates. To reinforce this point, Appendix Tables D8 and D9 show that the results remain qualitatively unchanged when excluding users who posted fewer than 50 ratings before the rating for the nominated movie or fewer than 50 ratings in total on the platform, respectively.

### **Placebo Tests: Predicted Ratings and Non-Nominated Movies**

We also conduct a series of placebo tests to reinforce these findings. Thus far, we have used embeddings to match users with a similar likelihood of being treated depending on their preferences. However, the recommender system also enables us to compute predicted ratings for each user, based on ratings posted pri-

---

<sup>20</sup>We use the Biased Matrix Factorization implementation described at <https://lkpy.readthedocs.io/projects/lenskit-tf/en/latest/biased-mf.html>. This method extends standard matrix factorization by incorporating user and item bias terms and learns parameters using gradient descent.

**Table 4.** Matching Users Before and After the AMPAS Nominations: Nominated Movies

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	-0.039*** (0.007)	-0.042*** (0.008)	-0.043*** (0.007)	-0.046*** (0.007)	-0.040*** (0.010)	-0.047*** (0.015)
User Embeddings		✓	✓	✓	✓	✓
$\bar{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\bar{s}_i^j$					✓	✓
$Share_i^j$						✓
N	63,849	59,937	59,914	56,098	40,395	18,218
Mean Dep. Var.	3.919	3.919	3.919	3.919	3.921	3.921

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M5 report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table 5.** Matching Users Before and After the AMPAS Nominations: Predicted Ratings

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	0.002 (0.003)	0.002 (0.004)	-0.003 (0.003)	-0.003 (0.003)	-0.003 (0.003)	-0.001 (0.005)
User Embeddings		✓	✓	✓	✓	✓
$\bar{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\bar{s}_i^j$					✓	✓
$Share_i^j$						✓
N	63,849	59,937	59,914	56,098	40,395	18,218
Mean Dep. Var.	3.919	3.919	3.919	3.919	3.921	3.921

*Note:* The sample includes predicted ratings for movies rated from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of predicted ratings on movie fixed effects and a dummy equal to one if a predicted rating is related to a rating posted after the nomination date. Columns M1 to M5 report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

or to the nominations. These predicted ratings, by design, are unaffected by disappointment or nomination-related effects.

In Table 5, we replicate the analysis from Table 4 using predicted ratings as the outcome variable. As expected, the predicted ratings remain stable over time, indicating that users with similar preferences (as captured by embeddings and other observable characteristics) have consistent predicted ratings before and after the reference nomination date. In Appendix Table D10, we further show that the null effect on predicted ratings is robust to the use of different recommendation algorithms. Specifically, we replicate the last specification (M5) from Table 5 using predicted ratings generated from three alternative algorithms available in the `LensKit` package: biased matrix factorization (MF), user-user collaborative filtering (UU),<sup>21</sup> and item-item collaborative filtering (II).<sup>22</sup> The user-user and item-item algorithms do not yield user embeddings and thus could not be used in the main matching specifications, but they do generate predicted ratings based on observed past behavior. Across all three approaches, we find no evidence of a shift in predicted ratings around the nomination date, reinforcing the interpretation that the drop in observed ratings is not driven by selection on preferences or exposure, but by disappointment induced by the nominations themselves.

Additionally, in Appendix Table D11, we repeat the analysis for non-nominated movies and find no significant effects, except for one specification with a negative effect statistically significant only at the 10 percent level. For these movies, ratings remain stable over time, and users with similar preferences provide consistent ratings before and after the reference nomination date.

## Movie and User Heterogeneity

Are some movies more affected by nomination-induced disappointment? Is the effect particularly strong for certain users? We address these questions in Appendix Tables D12, D13, D14, and D15, where we replicate the matching specification from Column (M5) in Table 4 across different subsets of movies and users. Appendix Table D12 shows that the disappointment effect is primarily driven by movies nominated for more than one AMPAS award. This finding aligns with the heterogeneity analysis discussed in the previous Section, suggesting that the disappointment effect is stronger when the signal is more likely to significantly influence users' expectations, that is, when movies receive multiple nominations. The negative effect is significant regardless of whether the movie ultimately wins an AMPAS award, but it appears to be larger for movies that are later awarded. This could be explained by the fact that movies that win awards

---

<sup>21</sup>A user-user collaborative filtering algorithm uses the ratings of a user to find other users that are similar. It then recommends items that similar users have rated highly. See <https://java.lenskit.org/documentation/algorithms/user-user/>.

<sup>22</sup>An item-item collaborative filtering algorithm uses all ratings of an item to determine similar items. It then recommends items that are similar to those that the user has rated highly. See <https://java.lenskit.org/documentation/algorithms/item-item/>.

tend to receive more nominations, and the importance of these nominations is particularly pronounced for such films. The heightened expectations created by multiple nominations, and the hype for the eventual awarding of some of these films, may amplify users' disappointment when their post-nomination viewing experience does not meet these elevated expectations.

The quality signal provided by AMPAS nominations and awards has evolved over time. In Appendix Table 13, we examine the heterogeneity of the disappointment effect across different years. The disappointment effect is the strongest for movies nominated from 1995 to 2004. For movies nominated between 2004 and 2011, the effect is smaller but still significant, whereas there is no significant effect for more recent movies from 2011 to 2019. This heterogeneity may be partly explained by the decreasing cultural relevance of the awards over time. In particular, various sources have noted a decline in the public interest in these awards.<sup>23</sup> However, additional factors likely contribute to the observed pattern.

One key factor is the evolution of the MovieLens platform itself. Its popularity, rating system, and user base have changed over time, potentially inducing selection effects. For example, early users (1995–2003) may have been less experienced and more prone to disappointment, whereas later users may differ in preferences, expectations, and viewing habits. The introduction of half-star ratings in 2003, while not causing visible global shifts in rating distributions (Harper and Konstan, 2015), may also have played a minor role. Importantly, we find a similar time-dependent pattern in IMDb reviews (Section 6), which mitigates concerns that the results are specific to MovieLens.

Another factor is the broader transformation in how consumers access and engage with movies. In the earlier years (1995–2003), users were more likely to rely on traditional distribution channels, such as theaters, cable TV, and video rental services, which often limited the number of movies they could watch. During this period, AMPAS nominations may have served as particularly salient quality signals, shaping expectations more strongly. In contrast, more recent periods are characterized by widespread access to streaming platforms and algorithmic recommendation systems. This shift has not only increased the volume of consumption but also diversified the sources of information users rely on when choosing movies. In this context, AMPAS nominations may play a less central role in shaping user expectations, contributing to the attenuated disappointment effect observed for more recent films.

We investigate user heterogeneity in Appendix Tables D14 and D15, where we show that the effect is primarily driven by users who have posted fewer than 400 ratings before rating the nominated movie, and by those who are less stringent (as defined in Section 2.3). In contrast, the average ratings users have posted

---

<sup>23</sup>See: <https://bunewsservice.com/the-death-of-awards-shows-do-people-no-longer-care/> and <https://movieweb.com/the-oscar-changed-lost-viewership/>.

and the share of movies they rated in the same year are not significant drivers of heterogeneity.

These findings suggest that users’ disappointment is especially pronounced for less experienced users, and by those who, on average, tend to rate movies more leniently. Previous research has shown that these features are typically associated with less expert users (Bondi et al., 2024; Aridor et al., 2024a), who are also more likely to rely on the AMPAS award, one of the most famous awards to discover and watch movies.

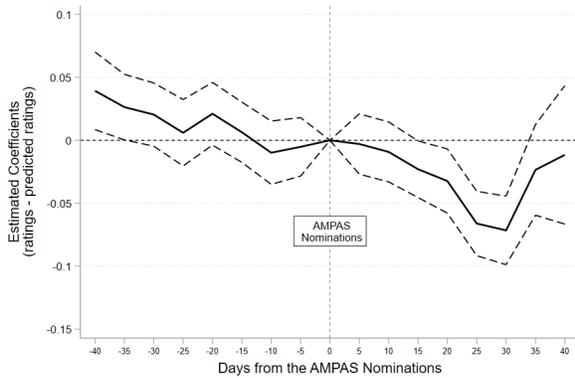
One potential concern is that the stronger effect observed for less experienced users may be partly driven by limitations in the precision of the embedding-based matching procedure, as their embeddings are estimated from fewer observations. If this were the case, we would expect the disappointment effect to be more pronounced among users for whom our predictions are less accurate. However, we find the opposite. In Appendix Table D14, we assess whether the observed heterogeneity in the disappointment effect is related to the accuracy of our user-level predictions. Specifically, we compute, for each user, the average absolute discrepancy between actual and predicted ratings based only on ratings posted before the AMPAS nominations, denoted by  $\bar{e}_i$ , and split users into two groups: those with  $\bar{e}_i \leq 0.5$  and those with  $\bar{e}_i > 0.5$ . The results show that the disappointment effect is significant only for users with lower average discrepancies. Accordingly, the stronger effect observed for less experienced users is not driven by noise in the matching procedure, but reflects differences in how users respond to nomination-induced expectations.

### The Dynamics of Disappointment

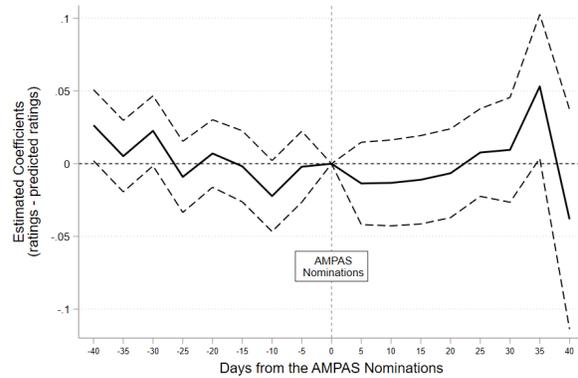
We conclude this Section by offering some insights into the dynamics of disappointment, using the recommendations produced by the recommender system. Specifically, we examine the difference between the actual rating posted by user  $i$  for movie  $j$ ,  $r_{ij}$ , and the predicted rating from the recommender system,  $\hat{r}_{ij}$ . We then explore how this difference evolves over time with a lead-lag model. In this model, the difference  $r_{ij} - \hat{r}_{ij}$  is regressed on movie fixed effects, and a set of dummy variables for each 5-day window around the AMPAS nomination dates:

$$r_{ij} - \hat{r}_{ij} = \gamma_j + \sum_{\tau=-40 \text{ days}}^{40 \text{ days}} \delta_{\tau} 1(t = \tau) + \varepsilon_{ji}. \quad (8)$$

This allows us to observe how the gap between actual and predicted ratings changes relative to the nomination event. Figure 5 presents two event studies, tracking the differences before and after the nominations for nominated and non-nominated movies. Nominated movies show a slight decline in the difference before the nominations, which could be due to minor changes in expectations as users begin to anticipate the potential for nominations. This is followed by a sharp drop starting 10 days after the nominations, which is



(a) Nominated Movies



(b) Non-Nominated Movies

**Figure 5.** The Difference between Actual and Predicted Ratings over time

*Notes:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The difference between actual and predicted ratings is regressed on movie fixed effects and a set of five-day interval dummies centered on the AMPAS nomination date for nominated and non-nominated movies. The graphs display the estimated coefficients of these dummies, with the coefficient for the five-day interval of the nomination date normalized to zero. Standard errors, clustered by movie, are reported at the 10% level.

consistent with the disappointment effect.

The difference begins to increase again around day 30 after the nominations. Part of this uptick may be attributed to the fact that, in many years, the AMPAS awards occur after the 35-day mark. Since we remove ratings posted after the awards, the sample of movies used in the analysis differs for these later time points. To assess the role of this potential composition effect, we re-estimate the model restricting the sample to years in which the AMPAS awards occur at least 40 days after the nominations (1997–2003, 2013, 2014, 2016, and 2018). As shown in Appendix Figure D3, the dynamics remain very similar: the drop in ratings reaches its lowest point around 25–30 days after the nominations, followed by a rebound. This similarity suggests that the uptick is unlikely to be driven solely by changes in movie composition. Specifically, the fading of the initial negative gap suggests that the psychological salience of the nomination event may wane over time. Accordingly, disappointment appears to be strongest when the triggering event (the nomination) is still top-of-mind for users. As attention shifts toward the upcoming award ceremony and users update their beliefs based on subsequent cues, expectations may recalibrate, diminishing the disappointment effect. These dynamics point to the importance of attention and memory in sustaining reference-dependent preferences and are consistent with recent work on disappointment and recency bias in belief formation (Gagnon-Bartsch and Bushong, 2022; Bushong and Gagnon-Bartsch, 2023).

In Appendix Figure D4, we replicate this analysis for nominated movies using alternative predicted ratings generated from three distinct recommendation algorithms provided by the LensKit package: biased

matrix factorization (MF), user-user collaborative filtering (UU), and item-item collaborative filtering (II). Despite differences in algorithmic structure and the way predictions are generated, all three event studies exhibit a strikingly similar pattern: a sharp decline in the difference between actual and predicted ratings beginning roughly 10 days after the nominations, followed by a recovery starting around day 30.

In contrast, non-nominated movies (Figure 5b) exhibit a flatter trend throughout, indicating that the nomination event does not systematically alter the gap between actual and predicted ratings for those movies.

In Appendix Table D16, we extend this analysis by implementing a DiD design, so comparing again nominated and non-nominated movies, leveraging information from predicted ratings and embeddings. In all specifications, we include only users who provided at least one rating 40 days prior to the nominations (a necessary condition for calculating predicted ratings and embeddings). In Column (1), we report the same DiD specification as in Column (4) of Table 3, confirming that, even with a different sample, the results remain consistent and exhibit a very similar magnitude. In Column (2), we replicate the same DiD design but use predicted ratings,  $\hat{r}_{ij}$ , as the dependent variable. The results indicate that nominations have no significant effect on predicted ratings. This is expected, as the recommendation model does not incorporate ratings after the nominations. These findings suggest that, in the absence of a nomination, the ratings of nominated movies would not experience a post-nomination decline. Columns (3) and (4) present variations of the specification in Column (4) of Table 3, but with additional controls: predicted ratings in Column (3) and the 50 embeddings calculated by the recommendation system in Column (4). These serve as alternative methods to address potential selection concerns, and the results consistently show a decline in ratings with a similar magnitude. Finally, in Column (5), we apply the DiD specification to the difference between the actual rating and the predicted rating,  $r_{ij} - \hat{r}_{ij}$ , once again confirming the same negative effect.

## 6 External Validation: IMDb Data

Thus far in the paper, we have leveraged the features of the dataset provided by MovieLens to investigate the role of disappointment in user ratings. The platform’s specific characteristics and the comprehensive dataset allowed us to credibly isolate the disappointment effect. However, in many other contexts, disappointment effects likely exist but are far more difficult to measure reliably. To assess the external validity of our findings, we extend our analysis to IMDb, the largest online repository of user-generated movie content.<sup>24</sup> To replicate the approach used on MovieLens, we scraped IMDb user reviews (including both the numerical ratings and the written content) for the set of nominated movies analyzed before. We also collected the full

---

<sup>24</sup>As of November 2024, IMDb is the third most-visited website in the U.S. in the streaming and online TV category. For more details about IMDb, see: <https://www.similarweb.com/website/imdb.com/>.

**Table 6.** Matching Users Before and After the AMPAS Nominations: IMDb Ratings

	(OLS)	(M1)	(M2)	(M3)	(M4)
Post Nomination	-0.149*** (0.012)	-0.121*** (0.015)	-0.127*** (0.017)	-0.150*** (0.020)	-0.162*** (0.024)
$\bar{r}_i^j$		✓	✓	✓	✓
$N_i^j$			✓	✓	✓
$\bar{s}_i^j$				✓	✓
$Share_i^j$					✓
N	54,838	36,398	32,232	21,767	16,143
Mean Dep. Var.	3.625	3.682	3.682	3.690	3.690

*Note:* The sample includes IMDb ratings posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M4 report ATT estimates matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

set of reviews written by users who rated these nominated titles, allowing us to construct user-level variables similar to those used in the MovieLens analysis. In this Section, we focus exclusively on the numerical ratings; in the next Section, we turn to the textual content of the reviews to explore different sentiments expressed by IMDb users on the platform. IMDb ratings range from 1 to 10, so we rescaled them to a 0.5–5-star scale to ensure comparability with MovieLens.

Despite the absence of predicted ratings and user embeddings from a recommendation system, we employed a matching design based on observable user characteristics, mirroring our approach with MovieLens. Specifically, we included the average rating posted by each user, the number of movies they rated before rating the nominated movies, their stringency levels (calculated as in Section 2.3 but limited to the set of ratings and users scraped from IMDb<sup>25</sup>), and the share of user ratings for movies released in the same year.

Table 6 presents the OLS and matching estimates derived from IMDb data. Column 1 (OLS) includes movie fixed effects, comparing ratings of users who watched the same nominated movies without accounting for user characteristics. Columns 2–5 (M1–M4) provide matching estimates based on exact matching on movie identifiers and quintiles of the aforementioned user attributes, following a methodology similar to Table 4. Our findings confirm that IMDb users also rate nominated movies more negatively after the nominations. Notably, the magnitude of this disappointment effect is approximately three times larger than what we observed on MovieLens. A key reason for this difference may be the composition of the active user base:

<sup>25</sup>On MovieLens, we also have data on users that have never rated nominated movies, allowing us to measure how strict each user was compared to all other users in the sample. On IMDb, this measure is restricted to the subset of users who rated nominated movies.

IMDb attracts a broader and more diverse audience, which includes a higher proportion of inexperienced users who may be more susceptible to heightened expectations set by nominations, whereas MovieLens users tend to be more experienced and engaged. Another possible explanation is the public nature of IMDb reviews, which could amplify societal influences such as groupthink, intensifying disappointment among users (a reverse “reputation effect” as in [Li et al., 2024](#)). However, as we show in the next Section, we do not observe a robust, statistically significant increase in reverse bandwagon sentiments (contrarianism toward mainstream films) on IMDb after nominations. This finding suggests that groupthink is unlikely to be the primary driver of the difference in magnitude between the two platforms.

To further explore the robustness of our findings, we conducted a heterogeneity analysis for IMDb data, similar to the approach used for MovieLens. The results reveal that the patterns of heterogeneity on IMDb closely mimic those observed on MovieLens, albeit with a larger magnitude of the effect. Specifically, in Appendix Table E1, we show that the disappointment effect is more pronounced for movies that receive more than one nomination and for those that eventually win an AMPAS award. These findings suggest that the salience of multiple nominations or an eventual win may elevate expectations further, intensifying the subsequent disappointment among viewers. Interestingly, the results in Appendix Table E2 also highlight that the effect is largely driven by movies produced between 1995 and 2004, consistent with our findings on MovieLens. However, the disappointment effect remains statistically significant even for the most recent years on IMDb. This suggests that while the awards’ influence may have faded, their capacity to shape consumer expectations, and trigger disappointment, has not disappeared entirely. Finally, we explored the role of user experience in driving the differences between IMDb and MovieLens results. Appendix Table E3 shows that IMDb users with fewer than five ratings prior to rating a nominated movie are the most affected by the disappointment effect. As the number of prior reviews increases, the magnitude of the effect decreases, pointing to the critical role of inexperience in amplifying disappointment. This pattern aligns with our findings on MovieLens, where more casual users (posting fewer ratings) also exhibit stronger disappointment. The higher prevalence of inexperienced users on IMDb, compared to the more expert reviewers on MovieLens, likely explains at least part of the larger effect size observed on IMDb.

## **7 Textual Evidences of Disappointment**

In the previous Sections, we presented evidence that online ratings on MovieLens and IMDb decline after the announcement of Academy Award nominations, even after accounting for the selection of users who rate movies before and after the nominations. We interpret this drop as evidence of disappointment: users’

expectations increase following the nominations, but the quality of the movie remains constant, and their heightened expectations are less likely to be met. As a result, users rate nominated movies less positively after the nominations. Our numerical analysis of ratings points strongly in this direction, but it cannot explicitly identify disappointment as the cause. So far, we have only looked at numerical ratings without direct evidence of the motivations or intentions behind users' evaluations.

In this Section, we address this limitation by analyzing the texts that users leave alongside their ratings on the two platforms, aiming to strengthen the interpretation of the observed decline in ratings as a disappointment effect. Specifically, we leverage two types of textual data: user-generated tags on MovieLens and written reviews on IMDb. Tags on MovieLens are typically one- or two-word descriptors that users associate with movies. Not all users add tags, and when they do, tags tend to focus more on objective attributes of the film (e.g., "sci-fi", "surreal", "based on a book") rather than on the user's emotional response (Vig, Sen and Riedl, 2012; Kotkov, Medlar, Satyal, Maslov, Neovius and Glowacka, 2022). On the other hand, the written reviews on IMDb are considerably richer: in addition to the rating, users often provide several sentences, or even paragraphs, expressing their feelings, thoughts, and evaluations after watching the movie. These written reviews are typically more sentiment-driven, offering a valuable window into users' emotional reactions rather than just descriptive content (Topal and Ozsoyoglu, 2016).

By analyzing users' texts, we can also check whether alternative mechanisms could be responsible for the drop in ratings.

One possibility is that the decline could be due to unobservable dimensions of user preferences: for instance, users who watch a movie after its nomination might be less intrinsically interested in it, but feel societal pressure to consume a "prestigious" product. This could lead to lower enjoyment and subsequently lower ratings, independent of disappointment. Such mechanisms have been documented in other contexts, where social dynamics push consumers toward products they might not otherwise choose, potentially resulting in lower satisfaction (Bursztyn, Handel, Jimenez and Roth, 2023). Although the selection corrections in previous sections should already account for such biases, analyzing the sentiments expressed in users' texts offers an additional and complementary validation.

Another potential explanation is a reverse bandwagon or snob effect (Leibenstein, 1950): some viewers may deliberately rate a popular or acclaimed movie more harshly precisely because it has become mainstream and widely praised, as a way to distance themselves from majority opinion (Corneo and Jeanne, 1997). This form of contrarianism, in which individuals derive utility from disliking what others like, could also generate a post-nomination dip in ratings even without any change in the actual experience.

Finally, post-nomination viewers might experience worse viewing conditions, such as more crowded

theaters or limited screening options, which could negatively affect their experience.

We start by describing in more detail the data sources, focusing on the nature of tags on MovieLens and written reviews on IMDb, and the sentiment analysis approach used to detect expressions of disappointment and other emotions. We then present preliminary evidence from MovieLens tags, which provides suggestive support for our disappointment hypothesis. Finally, we turn to a more extensive analysis of IMDb written reviews, where the richness of the text allows us to more convincingly show that disappointment, rather than disinterest, worse viewing conditions or contrarianism, is the main driver of the post-nomination decline in user ratings.

## 7.1 Textual Data and Sentiment Analysis Strategy

On MovieLens, users have the option to contribute tags that describe movies they have rated. Importantly, the process of assigning tags is separate from the act of rating a movie: users rate a movie with a numerical score and may, independently, add descriptive tags. Restricting the observations to the same sample used for the DiD analysis, we observe a total of 41,321 tags, compared to 497,266 ratings posted over the same period. Thus, tagging behavior is much less frequent than rating, and users leaving tags are likely to be drawn from a selected sample of users. Indeed, users who contribute tags are considerably more active: they post on average 328 reviews, compared to an average of 135 reviews for users who never leave any tags. Moreover, tag contributors tend to assign slightly higher ratings, with an average score of 3.77 versus 3.67 for non-tagging users. This suggests that tagging is more common among engaged and more positively inclined users. To quantify disappointment-related sentiment in viewer-contributed tags, we implemented a fuzzy matching approach against a curated list of keywords associated with unmet expectations (e.g., “disappointment”, “overhyped”, or “underwhelming”). This method allows us to detect both exact matches and close variants, accounting for potential misspellings, variations in phrasing, and word combinations. We provide a detailed explanation of the keyword selection, refinement process, and the fuzzy matching algorithm employed in Appendix F.

Unlike MovieLens tags, the textual component in the IMDb dataset is much richer and more consistently available. Specifically, we scraped user reviews from IMDb, each of which includes both a written component and an associated numerical rating. Moreover, IMDb reviews tend to be significantly longer than MovieLens tags and more explicitly reflect users’ opinions. To extract meaningful insights from user-generated movie reviews, we employ a methodology grounded in semantic similarity analysis using transformer-based models. Specifically, we use the `all-MiniLM-L6-v2` model from the Sentence Transformers library available on Hugging Face, which allows us to convert review texts into vector embeddings

that capture their semantic meaning.<sup>26</sup> For each analytical task, we compare these embeddings with a curated set of reference phrases representing the target concept (e.g., disappointment, disinterest). We then calculate cosine similarity scores between each review and the reference set, using those scores to determine whether a review expressed a particular sentiment or contextual reference.

The first analysis focuses on identifying reviews that express disappointment. To capture this, we compare reviews against reference phrases such as “I am disappointed” and “This movie was overrated.” The output is a numeric score indicating how strongly a review aligns with disappointment sentiment. In a second task, we explore a different form of negativity: disinterest. Here, we target reviews where users criticize a movie not necessarily because of poor quality, but because it did not align with their personal taste or preferences. Typical expressions include “Not my style” or “I just do not like this genre.” Again, reviews are scored based on how semantically similar they are to these reference expressions. A third task focuses on reviews mentioning poor viewing conditions. Phrases like “The cinema was crowded” or “There were too many people in the theater” served as indicators. This allowed us to assess whether user dissatisfaction could be linked to viewing circumstances rather than to the movie itself. A fourth task examines the reverse bandwagon, or snob, effect. In this case, we look for reviews where users criticize a movie precisely because it is popular or highly acclaimed, often as a way to distance themselves from mainstream opinion. Examples include “I avoid films that become mainstream successes” or “Too mainstream for me”.

Finally, we also employ a transformer-based classification model specifically trained to detect fine-grained emotions such as disappointment. We use the `SamLowe/roberta-base-go_emotions` model available on Hugging Face, which is fine-tuned on a large dataset of user-generated texts annotated with 28 distinct emotion labels, including “disappointment”.<sup>27</sup> This model allows us to assign a probability score to each review indicating the likelihood that it conveys disappointment. Unlike the semantic similarity method, which relies on custom reference phrases, this approach directly leverages the model’s classification capabilities to infer sentiment from the full context of the review text. We again provide a more detailed explanation of each of these tasks and approaches in Appendix F.

## 7.2 Evidence from MovieLens Tags

Out of the 41,321 tags posted around the nomination periods, only 109 refer to disappointment. This low number is expected, as tags on MovieLens are typically descriptive of the movie’s attributes rather than expressions of the user’s emotional response. Due to the small number of disappointment-related tags, it is

---

<sup>26</sup>For more information, see: <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>.

<sup>27</sup>For more information, see [https://huggingface.co/SamLowe/roberta-base-go\\_emotions](https://huggingface.co/SamLowe/roberta-base-go_emotions).

difficult to draw strong conclusions from this data alone.

Nonetheless, in Appendix Table F1, we show that tags related to disappointment are more likely to be associated with nominated movies compared to non-nominated ones. This pattern also holds when restricting the sample to the post-nomination period. Overall, while limited, this analysis provides suggestive evidence that disappointment may be more likely to arise in the context of Academy nominated movies.

### 7.3 Evidence from IMDb Reviews

Sentiment analysis of the IMDb reviews reveals that for the four dimensions of negativity we focus on (disappointment, disinterest, worse viewing conditions, and contrarianism) the average cosine similarity scores are around 0.35. This indicates medium similarity between the reviews and the target sentiments, which is expected given the typically longer nature of the reviews. To categorize the reviews, we create a dummy variable for each sentiment dimension (disappointment, disinterest, bad viewing conditions, and reverse bandwagon effect), which takes the value of 1 if the cosine similarity score exceeds the median value for that dimension.<sup>28</sup>

In Table 7 and Appendix Tables F2, F3, and F4, we present an analysis similar to the one shown in Table 6, but here the dependent variables are the four dummy variables representing disappointment, disinterest, bad viewing conditions, and reverse bandwagon effect (capturing contrarian reactions to a movie’s popularity). As before, in all four tables, Column 1 (OLS) includes movie fixed effects, comparing users who watched the same nominated movies without accounting for user characteristics. Columns 2–5 (M1–M4) provide matching estimates based on exact matching on movie identifiers and quintiles of the user attributes mentioned earlier.

In Table 7, we observe that users are more likely to write reviews expressing some level of disappointment after the AMPAS nomination. This finding holds consistently across different specifications and after accounting for the selection of users rating nominated movies before or after the nominations, suggesting that disappointment is a prominent sentiment after nominations.

In contrast, we do not find similar patterns for disinterest, worse viewing conditions, or contrarianism, as shown in Appendix Tables F2, F3, and F4. Interestingly, when user characteristics are not controlled for, the effect is significant for all three of these sentiments (though still only about half the size of the disappointment effect). This suggests that some post-nomination raters may indeed find nominated movies outside their preferred genres, experience suboptimal viewing conditions, or rate them more harshly because

---

<sup>28</sup>Specifically, the median scores are 0.3568 for disappointment, 0.3663 for disinterest, 0.2842 for bad viewing conditions, and 0.3545 for the reverse bandwagon effect. This approach naturally introduces some false positives, but being too selective would mean including only extreme negative reviews, which often exhibit high similarity scores across all dimensions.

**Table 7.** Matching Users Before and After the AMPAS Nominations: IMDb Reviews’ Disappointment

	(OLS)	(M1)	(M2)	(M3)	(M4)
Post Nomination	0.024*** (0.004)	0.023*** (0.006)	0.027*** (0.007)	0.030*** (0.008)	0.026*** (0.010)
$\bar{r}_i^j$		✓	✓	✓	✓
$N_i^j$			✓	✓	✓
$\bar{s}_i^j$				✓	✓
$Share_i^j$					✓
N	54,838	36,398	32,232	21,767	16,143
Mean Dep. Var.	0.499	0.453	0.453	0.452	0.452

*Note:* The sample includes IMDb reviews posted within an 80-day window around the AMPAS nomination date, excluding reviews posted after the AMPAS award ceremony. The dependent variable is a dummy variable that takes the value of 1 if the review’s cosine similarity score related to a set of expressions of disappointment exceeds the median value of 0.3568. The table displays estimated treatment effects for nominated movies. The first column presents the OLS estimate from a regression of reviews on movie fixed effects and a dummy variable equal to 1 if the review was posted after the nomination date. Columns M1 to M4 report ATT estimates, matching exactly on the movie identifier and quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies reviewed by the user prior to reviewing the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

of their mainstream status. However, once we account for user characteristics, these effects become statistically insignificant. The absence of a robust reverse bandwagon effect on IMDb reviews further suggests that the larger post-nomination rating decline observed on IMDb compared to MovieLens is more likely driven by differences in the composition of the user base rather than by stronger groupthink dynamics. In particular, MovieLens users tend to be more experienced and engaged, which could make them less susceptible to disappointment effect.

In addition, to check the robustness of our results regarding disappointment, we present additional analyses in Appendix Tables F5 and F6. These tables show that the results are not sensitive to the median threshold used to categorize reviews as expressing disappointment. When we select reviews within the 75th or 90th percentile in terms of cosine similarity to disappointment, we continue to observe a significant and positive effect on the probability of leaving reviews expressing disappointment after the AMPAS nominations. Finally, we replicate the analysis using the alternative measure of disappointment based on the transformer-based classification model `SamLowe/roberta-base-go_emotions`. We create a dummy variable that equals 1 if the predicted probability of the “disappointment” label exceeds the value of 0.5. The results, reported in Appendix Table F7, are largely consistent with those obtained using cosine similarity: across all matching specifications, we find a significant increase in the probability that users express disappointment in their reviews after the nominations, confirming the robustness of our findings to the sentiment detection approach.

## 8 Conclusion

We show that third-party quality disclosures, such as Academy Award nominations, can unintentionally reduce consumer satisfaction by inflating expectations. Even when product quality remains unchanged, consumer evaluations may decline due to a disappointment effect. This finding challenges the conventional view that quality signals always enhance perceived value in markets with imperfect information (Chen and Xie, 2005).

Our analysis reveals a significant short-run drop in user ratings following nomination announcements, suggesting that heightened expectations drive consumer dissatisfaction. This contributes to the growing literature on reference-dependent preferences and expectation-driven utility (Li et al., 2024; Bondi et al., 2024). By isolating the disappointment effect from selection and supply-side responses, we highlight the dynamic relationship between quality signals and consumer experience.

These insights have broader implications for online platforms and rating systems. Because expectations shaped by external signals can skew post-consumption evaluations, ratings may not always reflect intrinsic product quality. Platforms should consider mechanisms to account for such biases, using expectation-adjusted algorithms or dynamically tracking sentiment shifts as proposed by Aridor, Gonçalves, Kong, Kluver and Konstan (2024b). Addressing these biases can improve the reliability of ratings and support better decision-making by both consumers and firms.

Methodologically, our matching approach based on embedding representations offers a flexible framework for causal inference in high-dimensional, sparse data environments. It is particularly well-suited for settings where latent preferences and selection are key concerns. For example, applied to consumer panel data such as Nielsen Homescan, this method can be used to estimate heterogeneous responses to policy interventions like taxes or labeling requirements (Harding, Leibtag and Lovenheim, 2012; Bronnenberg, Dubé, Gentzkow and Shapiro, 2015). The same structure can be leveraged to evaluate platform design choices, including personalization algorithms (Simester, Timoshenko and Zoumpoulis, 2020), misinformation interventions (Fong, Guo and Rao, 2024), and sustainability labels (Proserpio, Goli, Mangini, Lau and Yu, 2025).

Finally, our findings add to the literature on online reviews, which are known to be shaped by manipulation and social dynamics (Luca and Zervas, 2016; Moe and Schweidel, 2012; He et al., 2022). We show that quality disclosures not only influence product choice but also affect how consumers retrospectively evaluate products. This underscores the importance of considering expectation dynamics when interpreting consumer feedback.

While our analysis focuses on user ratings, an important direction for future research is to examine the economic consequences of disappointment. If inflated expectations lead to lower post-consumption evaluations, certified products may suffer losses in sales or long-term demand despite unchanged quality. Quantifying these effects would help illuminate the broader economic relevance of psychological responses to quality signals and further inform the design of certification and marketing strategies.

By understanding and addressing the dynamics of disappointment, stakeholders in certification ecosystems, including firms, certification bodies, and online platforms, can refine their strategies to maximize the effectiveness of quality signals while preserving consumer welfare.

## References

- Anderson, Eric T, and Duncan I Simester (2014) ‘Reviews without a purchase: Low ratings, loyal customers, and deception.’ *Journal of Marketing Research* 51(3), 249–269
- Aridor, Guy, Duarte Gonçalves, Daniel Kluver, Ruoyan Kong, and Joseph Konstan (2024a) ‘The informational role of online recommendations: Evidence from a field experiment.’ *CESifo Working Paper*
- Aridor, Guy, Duarte Gonçalves, Ruoyan Kong, Daniel Kluver, and Joseph Konstan (2024b) ‘The movielens beliefs dataset: Collecting pre-choice data for online recommender systems.’ In ‘Proceedings of the 18th ACM Conference on Recommender Systems’ pp. 1–1
- Athey, Susan, and Guido W Imbens (2019) ‘Machine learning methods that economists should know about.’ *Annual Review of Economics* 11(1), 685–725
- Backus, Matthew, Thomas Blake, Dimitriy Masterov, and Steven Tadelis (2022) ‘Expectation, disappointment, and exit: evidence on reference point formation from an online marketplace.’ *Journal of the European Economic Association* 20(1), 116–149
- Barberis, Nicholas C (2013) ‘Thirty years of prospect theory in economics: A review and assessment.’ *Journal of economic perspectives* 27(1), 173–196
- Billsus, Daniel, Michael J Pazzani et al. (1998) ‘Learning collaborative information filters.’ In ‘Icml,’ vol. 98 pp. 46–54
- Bondi, Tommaso (2023) ‘Alone, together: A model of social (mis) learning from consumer reviews.’ *Available at SSRN 4453685*
- Bondi, Tommaso, Michelangelo Rossi, and Ryan Stevens (2024) ‘The good, the bad and the picky: Consumer heterogeneity and the reversal of product ratings.’ *Management Science*
- Brandes, Leif, and Yaniv Dover (2022) ‘Offline context affects online reviews: The effect of post-consumption weather.’ *Journal of Consumer Research* 49(4), 595–615
- Bronnenberg, Bart J, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M Shapiro (2015) ‘Do pharmacists buy bayer? informed shoppers and the brand premium.’ *The Quarterly Journal of Economics* 130(4), 1669–1726

- Bursztyn, Leonardo, Benjamin R Handel, Rafael Jimenez, and Christopher Roth (2023) ‘When product markets become collective traps: The case of social media.’ Technical Report, National Bureau of Economic Research
- Bushong, Benjamin, and Tristan Gagnon-Bartsch (2023) ‘Reference dependence and attribution bias: evidence from real-effort experiments.’ *American Economic Journal: Microeconomics* 15(2), 271–308
- Chen, Yan, F Maxwell Harper, Joseph Konstan, and Sherry Xin Li (2010) ‘Social comparisons and contributions to online communities: A field experiment on movielens.’ *American Economic Review* 100(4), 1358–1398
- Chen, Yubo, and Jinhong Xie (2005) ‘Third-party product review and firm marketing strategy.’ *Marketing science* 24(2), 218–240
- Corneo, Giacomo, and Olivier Jeanne (1997) ‘Snobs, bandwagons, and the origin of social customs in consumer behavior.’ *Journal of Economic Behavior & Organization* 32(3), 333–347
- De Langhe, Bart, Philip M Fernbach, and Donald R Lichtenstein (2016) ‘Navigating by the stars: Investigating the actual and perceived validity of online user ratings.’ *Journal of Consumer Research* 42(6), 817–833
- Dell, Melissa (2024) ‘Deep learning for economists.’ Technical Report, National Bureau of Economic Research
- 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), 277–285
- Dranove, David, and Ginger Zhe Jin (2010) ‘Quality Disclosure and Certification: Theory and Practice.’ *Journal of Economic Literature* 48(4), 935–63
- Ekstrand, Michael D, John T Riedl, Joseph A Konstan et al. (2011a) ‘Collaborative filtering recommender systems.’ *Foundations and Trends® in Human–Computer Interaction* 4(2), 81–173
- Ekstrand, Michael D, Michael Ludwig, Joseph A Konstan, and John T Riedl (2011b) ‘Rethinking the recommender research ecosystem: reproducibility, openness, and lenskit.’ In ‘Proceedings of the fifth ACM conference on Recommender systems’ pp. 133–140
- Elfenbein, Daniel W, Raymond Fisman, and Brian McManus (2015) ‘Market structure, reputation, and the value of quality certification.’ *American Economic Journal: Microeconomics* 7(4), 83–108

- Farronato, Chiara, and Georgios Zervas (2022) ‘Consumer reviews and regulation: Evidence from nyc restaurants.’ Technical Report, National Bureau of Economic Research
- Farronato, Chiara, Andrey Fradkin, Bradley J Larsen, and Erik Brynjolfsson (2024) ‘Consumer protection in an online world: An analysis of occupational licensing.’ *American Economic Journal: Applied Economics* 16(3), 549–579
- Fong, Jessica, Tong Guo, and Anita Rao (2024) ‘Debunking misinformation about consumer products: Effects on beliefs and purchase behavior.’ *Journal of Marketing Research* 61(4), 659–681
- Fradkin, Andrey, Elena Grewal, and David Holtz (2021) ‘Reciprocity and unveiling in two-sided reputation systems: Evidence from an experiment on airbnb.’ *Marketing Science* 40(6), 1013–1029
- Friberg, Richard, and Erik Grönqvist (2012) ‘Do expert reviews affect the demand for wine?’ *American Economic Journal: Applied Economics* 4(1), 193–211
- Gagnon-Bartsch, Tristan, and Benjamin Bushong (2022) ‘Learning with misattribution of reference dependence.’ *Journal of Economic Theory* 203, 105473
- Gergaud, Olivier, Karl Storchmann, and Vincenzo Verardi (2015) ‘Expert opinion and product quality: Evidence from new york city restaurants.’ *Economic Inquiry* 53(2), 812–835
- Goldberg, David, David Nichols, Brian M. Oki, and Douglas Terry (1992) ‘Using collaborative filtering to weave an information tapestry.’ *Commun. ACM* 35(12), 61–70
- Ham, Sung H, Ingrid Koch, Noah Lim, and Jiabin Wu (2021) ‘Conflict of interest in third-party reviews: an experimental study.’ *Management Science* 67(12), 7535–7559
- Harding, Matthew, Ephraim Leibtag, and Michael F Lovenheim (2012) ‘The heterogeneous geographic and socioeconomic incidence of cigarette taxes: evidence from nielsen homescan data.’ *American Economic Journal: Economic Policy* 4(4), 169–198
- Harper, F Maxwell, and Joseph A Konstan (2015) ‘The movielens datasets: History and context.’ *Acm transactions on interactive intelligent systems (tiis)* 5(4), 1–19
- He, Sherry, Brett Hollenbeck, and Davide Proserpio (2022) ‘The market for fake reviews.’ *Marketing Science* 41(5), 896–921

- Hilger, James, Greg Rafert, and Sofia Villas-Boas (2011) ‘Expert opinion and the demand for experience goods: an experimental approach in the retail wine market.’ *Review of Economics and Statistics* 93(4), 1289–1296
- Hu, Nan, Paul A Pavlou, and Jie Jennifer Zhang (2009) ‘Why do online product reviews have a j-shaped distribution? overcoming biases in online word-of-mouth communication.’ *Communications of the ACM* 52(10), 144–147
- Hu, Nan, Paul A Pavlou, and Jie Zhang (2017) ‘On self-selection biases in online product reviews.’ *MIS quarterly* 41(2), 449–475
- Hui, Xiang, Maryam Saeedi, Zeqian Shen, and Neel Sundaresan (2016) ‘Reputation and regulations: Evidence from ebay.’ *Management Science* 62(12), 3604–3616
- Kastanakis, Minas N, and George Balabanis (2012) ‘Between the mass and the class: Antecedents of the “bandwagon” luxury consumption behavior.’ *Journal of business research* 65(10), 1399–1407
- (2014) ‘Explaining variation in conspicuous luxury consumption: An individual differences’ perspective.’ *Journal of Business Research* 67(10), 2147–2154
- Koren, Yehuda, Robert Bell, and Chris Volinsky (2009) ‘Matrix factorization techniques for recommender systems.’ *Computer* 42(8), 30–37
- Kőszegi, Botond, and Matthew Rabin (2006) ‘A model of reference-dependent preferences.’ *The Quarterly Journal of Economics* 121(4), 1133–1165
- Kotkov, Denis, Alan Medlar, Umesh Raj Satyal, Alexandr Maslov, Mats Neovius, and Dorota Glowacka (2022) ‘Rating consistency is consistently underrated: An exploratory analysis of movie-tag rating inconsistency.’ In ‘Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing’ pp. 1355–1364
- Kovács, Balázs, and Amanda J Sharkey (2014) ‘The paradox of publicity: How awards can negatively affect the evaluation of quality.’ *Administrative science quarterly* 59(1), 1–33
- Lagios, Nicolas, and Pierre-Guillaume Méon (2023) ‘A matter of taste: The negative welfare effect of expert judgments.’ *Working Papers CEB*
- Lange, Richard D, Ankani Chattoraj, Jeffrey M Beck, Jacob L Yates, and Ralf M Haefner (2021) ‘A confirmation bias in perceptual decision-making due to hierarchical approximate inference.’ *PLoS Computational Biology* 17(11), e1009517

- Lee, Young-Jin, Kartik Hosanagar, and Yong Tan (2015) ‘Do i follow my friends or the crowd? information cascades in online movie ratings.’ *Management Science* 61(9), 2241–2258
- Leibenstein, Harvey (1950) ‘Bandwagon, snob, and veblen effects in the theory of consumers’ demand.’ *The quarterly journal of economics* 64(2), 183–207
- Li, Xingyi, Yiting Deng, Puneet Manchanda, and Bert De Reyck (2024) ‘Can lower expert opinions lead to better consumer ratings?: The case of michelin stars.’ *Available at SSRN 4275944*
- Li, Xinxin, and Lorin M Hitt (2008) ‘Self-selection and information role of online product reviews.’ *Information systems research* 19(4), 456–474
- Luca, Michael, and Georgios Zervas (2016) ‘Fake it till you make it: Reputation, competition, and yelp review fraud.’ *Management science* 62(12), 3412–3427
- Mitchell, Terence R, Leigh Thompson, Erika Peterson, and Randy Cronk (1997) ‘Temporal adjustments in the evaluation of events: The “rosy view”.’ *Journal of experimental social psychology* 33(4), 421–448
- Moe, Wendy W, and David A Schweidel (2012) ‘Online product opinions: Incidence, evaluation, and evolution.’ *Marketing Science* 31(3), 372–386
- Moorthy, Sridhar, Brian T Ratchford, and Debabrata Talukdar (1997) ‘Consumer information search revisited: Theory and empirical analysis.’ *Journal of consumer research* 23(4), 263–277
- Nosko, Chris, and Steven Tadelis (2015) ‘The limits of reputation in platform markets: An empirical analysis and field experiment.’ Technical Report, National Bureau of Economic Research
- Parkinson, Brian, Rob B Briner, Shirley Reynolds, and Peter Totterdell (1995) ‘Time frames for mood: Relations between momentary and generalized ratings of affect.’ *Personality and Social Psychology Bulletin* 21(4), 331–339
- Paterek, Arkadiusz (2007) ‘Improving regularized singular value decomposition for collaborative filtering.’ *Proc. KDD Cup and Workshop* ACM Press, 39–42
- Proserpio, Davide, Ali Goli, Tyler Mangini, Ken Lau, and Daniela Yu (2025) ‘The impact of sustainability programs on consumer purchase behavior: Evidence from amazon.’ *International Journal of Research in Marketing*

- Proserpio, Davide, Wendy Xu, and Georgios Zervas (2018) ‘You get what you give: theory and evidence of reciprocity in the sharing economy.’ *Quantitative Marketing and Economics* 16, 371–407
- Rey-Biel, Pedro, Ayelet Gneezy, Uri Gneezy, Dominique Lauga, and Joan Llull (2024) ‘Expectations, satisfaction and utility from experience goods.’ *Manuscript*
- Sarwar, Badrul, George Karypis, Joseph Konstan, and John Riedl (2000) ‘Analysis of recommendation algorithms for e-commerce.’ In ‘Proceedings of the 2nd ACM Conference on Electronic Commerce’ pp. 158–167
- Schoenmueller, Verena, Oded Netzer, and Florian Stahl (2020) ‘The polarity of online reviews: Prevalence, drivers and implications.’ *Journal of Marketing Research* 57(5), 853–877
- Simester, Duncan, Artem Timoshenko, and Spyros I Zoumpoulis (2020) ‘Targeting prospective customers: Robustness of machine-learning methods to typical data challenges.’ *Management Science* 66(6), 2495–2522
- Sunder, Sarang, Kihyun Hannah Kim, and Eric A Yorkston (2019) ‘What drives herding behavior in online ratings? the role of rater experience, product portfolio, and diverging opinions.’ *Journal of Marketing* 83(6), 93–112
- Topal, Kamil, and Gultekin Ozsoyoglu (2016) ‘Movie review analysis: Emotion analysis of imdb movie reviews.’ In ‘2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)’ IEEE pp. 1170–1176
- Vig, Jesse, Shilad Sen, and John Riedl (2012) ‘The tag genome: Encoding community knowledge to support novel interaction.’ *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2(3), 1–44
- Wang, Chong, Xiaoquan Zhang, and Il-Horn Hann (2018) ‘Socially nudged: A quasi-experimental study of friends’ social influence in online product ratings.’ *Information Systems Research* 29(3), 641–655

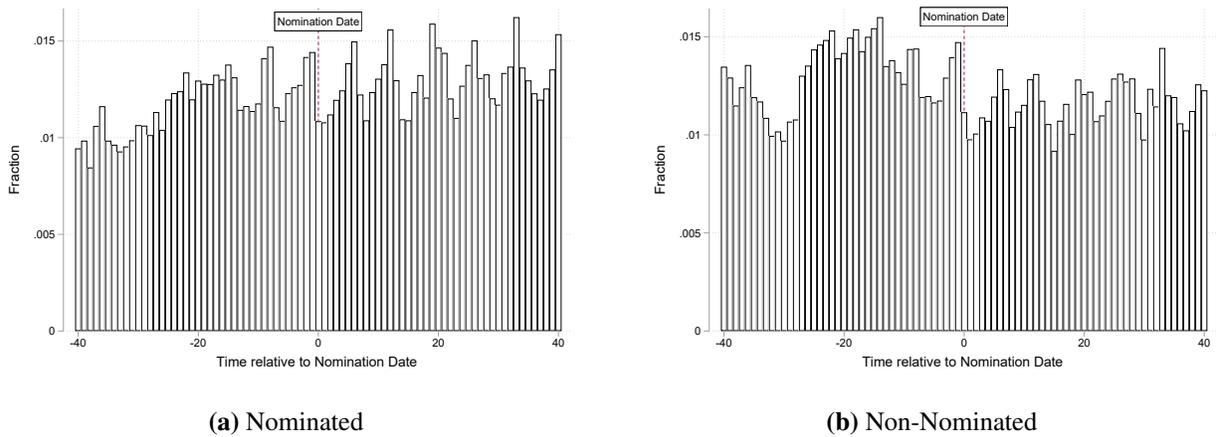
## **APPENDIX - Quality Disclosures and Disappointment: Evidence from the Academy Awards by Michelangelo Rossi and Felix SchleeF**

### **A APPENDIX - Empirical Setting and Dataset**

**Table A1.** Summary Statistics: Non-Nominated Movies

	Mean	SE	Min	Max
<i>Average Ratings</i>				
All	3.212	.0047	.8548	4.483
Before Nom.	3.182	.0063	.5	5
After Nom.	3.213	.0048	.8929	4.48
<i>Number of Ratings</i>				
All	1221	35.5011	30	72674
Before Nom.	127.2	5.23	1	15200
After Nom.	1085	32.0367	4	68582
<i>Average Ratings (80-days window)</i>				
All	3.205	.0071	.5	5
Before Nom.	3.197	.0082	.5	5
After Nom.	3.221	.0083	.5	5
<i>Number of Ratings (80-days window)</i>				
All	43.45	1.206	0	2827
Before Nom.	23.55	.6598	0	1454
After Nom.	20.31	.5767	0	1373

*Note:* The table presents descriptive statistics on all 9,397 movies, non-nominated for AMPAS awards, included in our sample. We report the average ratings and the number of ratings both before and after the nominations. The first two panels include all available ratings without any time restrictions, while the second two panels focus on ratings within a 80-day window surrounding the nominations.



**Figure A1.** Arrival of Ratings for Nominated and Non-Nominated Movies

*Notes:* The two figures show the amount of ratings that are displayed over time for not nominated and nominated movies. On the x-axis, time is measured in terms of days of distance from the nomination dates.

**Table A2.** User Selection Before and After the Nominations: All Ratings

	Mean	SE	Min	Max
<i>Average User Ratings</i>				
All	3.456	.0012	2.956	3.946
Before Nom.	3.411	.0022	1.038	5
After Nom.	3.459	.0013	2.65	3.955
<i>Number of User Ratings</i>				
All	758.5	3.99432	62.09	2927.1
Before Nom.	813.5	6.457	1	17502
After Nom.	766.6	4.35713	62.17	5602
<i>Average User Stringency</i>				
All	.0794	.001	-.5153	.6804
Before Nom.	.1101	.0024	-2.243	2.2
After Nom.	.0768	.0011	-.4975	1.149
<i>Average Share of User Ratings for Current-Year Releases</i>				
All	.041	.000271	.0055	.427
Before Nom.	.1072	.00116	0	.8587
After Nom.	.03	.000134	.0047	.2183

*Note:* The table presents descriptive statistics of users who rated movies nominated for AMPAS Awards included in our sample, with no time restrictions. We report the average user rating, the average number of ratings per user, the average rating stringency (a measure of how strictly users rate, based on how their ratings compare to the average rating of the movie at the time), and the average share of ratings for movies produced in the same year as the rating. The stringency of a user is defined as a moving average of the stringency of their past five ratings. These statistics are presented separately for users who rated the movies before and after the nominations.

## B APPENDIX - Conceptual Framework

### Signals Between Watching and Rating Decisions

In our framework, the decisions to watch and to rate a movie are closely linked: users receive signals about a movie's quality either before or after both actions and form expectations accordingly. In practice, however, it is reasonable to assume that users may often rate a movie days or even weeks after watching it. This temporal gap introduces a crucial nuance, namely, that a new quality signal, such as an AMPAS nomination, may arrive after the user has seen the movie but before they submit a rating.

Our data capture only the timing of the rating, not the moment of consumption. As a result, we cannot directly observe whether a user watched the movie before or after receiving the signal. Nevertheless, we can articulate two competing hypotheses regarding how such signals might affect evaluations when they arrive after viewing but prior to rating.

**Hypothesis 1: Primacy of First Experience.** According to this hypothesis, users who watch a movie before any nomination signal are unaffected by that signal, even if it precedes their rating. This view draws on the notion of primacy or confirmation bias (Lange, Chatteraj, Beck, Yates and Haefner, 2021), where the initial experience (here, the act of watching the film) anchors the user's evaluation. Since expectations were formed without knowledge of the nomination and the film's quality has already been directly experienced, the rating reflects that unmediated encounter. In this case, nominations would not influence ratings, and these users would resemble those who both watched and rated the film before the signal emerged.

**Hypothesis 2: Retrospective Signal Adjustment.** An alternative possibility is that users who watched the movie before the nomination may still be influenced by the nomination if it arrives before they submit their rating. In this view, the evaluation is not solely a function of an objective memory of the film's quality, or the expectations held at the time of viewing. Rather, it is also shaped by the user's psychological and cognitive state at the moment of rating (Parkinson, Briner, Reynolds and Totterdell, 1995), an internal context that may have been altered by the arrival of the signal. A nomination can thus serve as a reframing cue, prompting the user to reassess the film not in isolation but in light of its new categorization as an "Academy-nominated" movie (Mitchell, Thompson, Peterson and Cronk, 1997). The result may be a shift in perceived standards or reference points, leading to a revised rating that reflects a discrepancy between the remembered experience and the elevated status suggested by the nomination. In line with this, Brandes and Dover (2022) demonstrate that a reviewer's offline context at the time of review provision systematically

affects online review content for hotels: they use weather conditions at the time of reviewing and show that 6.5% more reviews are written on rainy days, and that these reviews are 0.1 points lower on average. This mechanism implies that post-viewing signals can retroactively reshape evaluations by influencing how users interpret and contextualize their memories, rather than by modifying expectations beforehand.

These opposing hypotheses illustrate the complexity introduced by the unobserved timing of consumption. Without data on when users watched the movie, we cannot definitively determine which mechanism prevails. However, this distinction becomes especially relevant when we explore heterogeneity in the nomination effect over time, as discussed in Sections 4 and 5. Observing how the size of the effect evolves in the weeks following the nomination may offer indirect evidence on whether such post-viewing, pre-rating influence exists.

Ultimately, resolving this question remains an important avenue for future research. One promising direction would be to leverage belief data as in [Aridor et al. \(2024b\)](#) to study how expectations adjust after a signal changes but before a rating is submitted.

## **C APPENDIX - Comparing Ratings for Nominated and Non-Nominated Movies**

**Table C1.** Difference-in-Differences: Removing Days Right After the Nominations

	(All)	(5 Days After)	(10 Days After)	(15 Days After)
$Nominated_j \times post_t^{Nom}$	-0.018** (0.007)	-0.019** (0.008)	-0.024*** (0.008)	-0.033*** (0.009)
Movie FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
$R^2$	0.21	0.21	0.21	0.21
N	497,266	469,327	437,886	405,986
Mean Dep. Var.	3.503	3.502	3.503	3.502

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. Column 1 (All) reports results based on all ratings in the window. Columns 2 to 4 (5 Days After; 10 Days After; 15 Days After) restrict the sample to ratings posted within 5, 10, and 15 days following the nomination date, respectively. Standard errors, clustered by movie, are shown in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table C2.** Difference-in-Differences: Varying Non-Nominated Movies in the Control Group

	(All)	( $n_i^{critics} > 0$ )	( $\bar{r}_i^{critics} > 50$ )	( $\bar{r}_i^{IMDb} > 5$ )
$Nominated_j \times post_t^{Nom}$	-0.018** (0.007)	-0.015** (0.007)	-0.013* (0.007)	-0.016** (0.007)
Movie FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
$R^2$	0.21	0.20	0.16	0.18
N	497,266	467,818	397,498	484,683
Mean Dep. Var.	3.503	3.511	3.628	3.532

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. Column 1 (All) reports results based on all ratings in the window. Columns 2 to 4 ( $n_i^{critics} > 0$ ,  $\bar{r}_i^{critics} > 50$ , and  $\bar{r}_i^{IMDb} > 5$ ) restrict the sample to movies with at least one professional review on the Metacritic website, those with an average Metacritic score above 50 (out of 100), and those with an average IMDb rating above 5, respectively. Standard errors, clustered by movie, are shown in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table C3.** Difference-in-Differences: Nominated Movie Heterogeneity

	(1)	(2)	(3)	(4)
$Nominated_j^{=1} \times post_t^{Nom}$	-0.012 (0.011)			
$Nominated_j^{>1} \times post_t^{Nom}$		-0.021** (0.008)		
$Nominated_j^{award=0} \times post_t^{Nom}$			-0.018** (0.009)	
$Nominated_j^{award>0} \times post_t^{Nom}$				-0.018* (0.011)
Movie FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
$R^2$	0.19	0.21	0.19	0.20
N	429,955	458,487	455,277	433,165
Mean Dep. Var.	3.421	3.475	3.451	3.446

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. In all specifications, the control group of non-nominated movies remains unchanged, while the treated group varies across columns. In Column (1), the variable  $Nominated_j^{=1}$  identifies movies that received exactly one nomination. In Column (2),  $Nominated_j^{>1}$  captures movies with more than one nomination. In Column (3),  $Nominated_j^{award=0}$  refers to nominated movies that did not win any award. In Column (4),  $Nominated_j^{award>0}$  includes nominated movies that won at least one award. Standard errors, clustered by movie, are shown in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table C4.** Difference-in-Differences: Controlling for User Selection

	(1)	(2)	(3)	(4)	(5)
$Nominated_j \times post_t^{Nom}$	-0.018** (0.007)	-0.018** (0.007)	-0.016** (0.007)	-0.017** (0.007)	-0.027*** (0.007)
$\bar{r}_i^j$	0.689*** (0.008)	0.693*** (0.008)	0.244*** (0.008)	0.244*** (0.008)	-0.156*** (0.015)
$N_i^j$		0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
$\bar{s}_i^j$			-0.520*** (0.005)	-0.521*** (0.005)	-0.201*** (0.006)
$Share_i^j$				-0.189*** (0.022)	-0.258*** (0.049)
Movie FEs	✓	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓	✓
User FEs					✓
$R^2$	0.30	0.30	0.33	0.33	0.43
N	496,044	496,044	493,817	493,817	487,847
Mean Dep. Var.	3.502	3.502	3.503	3.503	3.498

Note: The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. Standard errors, clustered by movie, are shown in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## D APPENDIX - A Recommendation-based Matching Design

### D.1 Computation of User Embeddings for Recommendation System Matching

To implement our recommendation system matching approach, we construct user embeddings that reflect individual preferences over time. We present here in detail the procedure used to compute these embeddings, which are later used in our matching algorithm.

We rely on a standard matrix factorization technique based on singular value decomposition (SVD), as implemented in the `scikit-surprise` python package. This algorithm decomposes the user-item rating matrix into low-dimensional latent factors, which we interpret as embeddings of user tastes and item characteristics. The performance and hyperparameter settings of this algorithm are discussed in Appendix D.2.

The construction of time-varying user embeddings proceeds as follows:

- We loop over all nomination dates between 1995 and 2019, which corresponds to the time span covered by the Movielens 25M dataset. For users who are active in multiple nomination periods, we compute a separate set of embeddings for each period.
- In each iteration of the loop, we restrict the rating data to include only ratings up to 40 days before the respective nomination date. This ensures that user preferences are estimated based only on information available prior to each nomination period.
- We then train the SVD algorithm on the windowed rating data and recover the user embeddings.
- Lastly, we merge the resulting user embeddings with the rating data. Each film in the dataset is assigned a nomination reference date (i.e., the date at which it would have been eligible for nomination, regardless of whether it was actually nominated). For each rating, we assign the embedding corresponding to the nomination period in which the film falls.

This procedure allows user embeddings to evolve across time, reflecting changing preferences or reactions to past outcomes. In particular, we allow ratings from the previous nomination period to enter the training data of the subsequent period. This design choice ensures that prior disappointment from award outcomes can influence the embedding used for future matching, capturing potential dynamics in user tastes.

**Table D1.** Recommender Algorithm Performance with Different Specifications

<b>No. of Epochs = 10</b>				
<b>No. of Factors</b>	(LR=0.002, RR=0.02)	(LR=0.002, RR=0.01)	(LR=0.005, RR=0.02)	(LR=0.005, RR=0.01)
10	.8586	.8689	.8185	.864
20	.8561	.8688	.8128	.8623
50	.8513	.8688	.8074	.8598
100	.8487	.8686	.8035	.8582
200	.8457	.8684	.801	.8566
<b>No. of Epochs = 20</b>				
<b>No. of Factors</b>	(LR=0.002, RR=0.02)	(LR=0.002, RR=0.01)	(LR=0.005, RR=0.02)	(LR=0.005, RR=0.01)
10	.8268	.8644	.7903	.847
20	.8226	.8635	.7836	.8461
50	.8162	.8617	.7789	.8449
100	.8118	.86	.7779	.8439
200	.8088	.8578	.7781	.8427

*Note:* The sample includes all MovieLens ratings from 1995 to 2019, posted up to 40 days before the nomination date of each year. The table displays the average root-mean-squared error (RMSE) obtained via 5-fold cross-validation across a range of hyperparameter values: the number of factors (embedding dimensions), the number of training epochs, the learning rate (LR), and the regularization rate (RR).

## D.2 The Hyperparameter Choice for the Recommender Algorithm and its Efficiency

The calculation of embedding values for the matching procedure involves several degrees of freedom, governed by four key hyperparameters: the number of embedding dimensions, the number of training epochs, the learning rate (LR), and the regularization rate (RR). The number of embedding dimensions determines the size of the vector that encodes a user’s preferences. The number of epochs indicates how many iterations the optimization routine performs to update and refine these embeddings. The learning rate controls the step size of these updates, affecting the speed and stability of convergence. Finally, the regularization rate penalizes large parameter values, helping to prevent overfitting.

Appendix Table D1 reports the average root-mean-squared error (RMSE) obtained via 5-fold cross-validation across a range of hyperparameter values. The results indicate that a learning rate of 0.005 and a regularization rate of 0.02 yield the best predictive performance. While increasing the number of epochs from 10 to 20 does lead to some improvement, the RMSE is less sensitive to this change than to variations in the learning and regularization rates. Similarly, expanding the number of embedding dimensions from 50 to 100 or 200 yields only marginal gains. Based on these findings, we adopt the following hyperparameter configuration for all embedding calculations used in the matching estimations: LR = 0.005, RR = 0.02, number of epochs = 20, and number of embedding = 50. The chosen values for learning rate, regularization rate, and number of epochs coincide with the default settings of the `scikit-surprise` library. The only deviation from the default is the number of embeddings, which we set to 50 instead of the default of 100.

### D.3 Embedding Distances and Rating Similarity

A key assumption underlying our identification strategy is that the user embeddings produced by the recommendation model meaningfully reflect users’ tastes and preferences. While the model’s predictive accuracy (e.g., RMSE) validates the overall quality of the rating predictions, it does not guarantee that the learned embeddings capture behaviorally-relevant preference structures. Here, we present a validation exercise designed to test whether users who express more similar opinions on specific movies also tend to have more similar embeddings.

We focus on the 50 most-rated movies in our dataset, as these ensure a large and diverse set of users for comparison. For each movie  $j$  and user  $i$ , we compute the absolute difference between user  $i$ ’s rating and the rating of every other user  $k$  who also rated movie  $j$ :  $\delta_{ik} = |r_{ij} - r_{kj}|$ . This measure,  $\delta_{ik}$ , captures the degree of rating similarity between users  $i$  and  $k$  for movie  $j$ . By construction,  $\delta_{ik}$  takes values ranging from a minimum of 0, when users give the same rating, to a maximum of 4, corresponding to the largest possible difference in ratings (e.g., one user gives a 5 and the other a 1).<sup>29</sup>

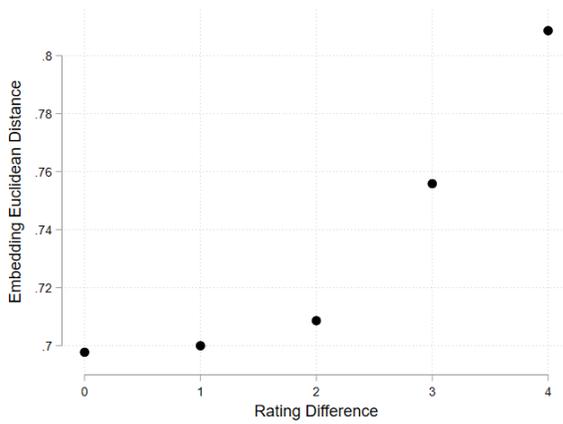
We then compute the Euclidean and Mahalanobis distances between the embedding of user  $i$  and that of user  $k$ , for all pairs of users who rated the same movie. If embeddings reflect preferences, we should observe that smaller values of  $\delta_{ik}$  are associated with smaller distances in embedding space.

Appendix Figures D1 and D2 show the average Euclidean and Mahalanobis distances, respectively, between users’ embeddings as a function of  $\delta_{ik}$ . Each point in the plot represents the average embedding distance for all user pairs  $(i, k)$  with a given  $\delta_{ik}$  value.

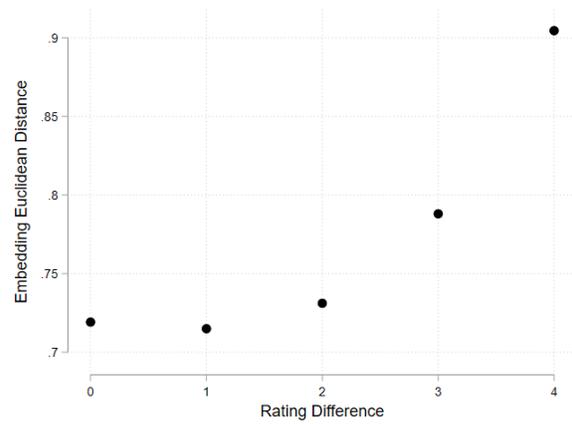
To provide granularity, Appendix Figures D1a—D1e and D2a—D2e show the same analysis separately for each of the five most-rated movies in the dataset (listed in chronological order): *Crouching Tiger, Hidden Dragon* (2000); *Lord of the Rings: The Fellowship of the Ring* (2001); *Lord of the Rings: The Return of the King* (2003); *Avatar* (2009); and *The Martian* (2015). Finally, Appendix Figures D1f and D2f show the results for all top-rated 50 movies. Across all specifications, we observe a consistent monotonic relationship: as  $\delta_{ik}$  increases, so does the average embedding distance. Specifically, we find that users with a rating difference of 0 or 1 tend to have relatively similar embeddings, indicating aligned preferences. In contrast, user pairs with the largest possible differences in ratings (such as one user loving and the other disliking the same movie) tend to have much larger distances in embedding space, suggesting that their overall movie preferences diverge substantially. This pattern holds for both Euclidean and Mahalanobis distances, both in individual movies and in the aggregated sample.

---

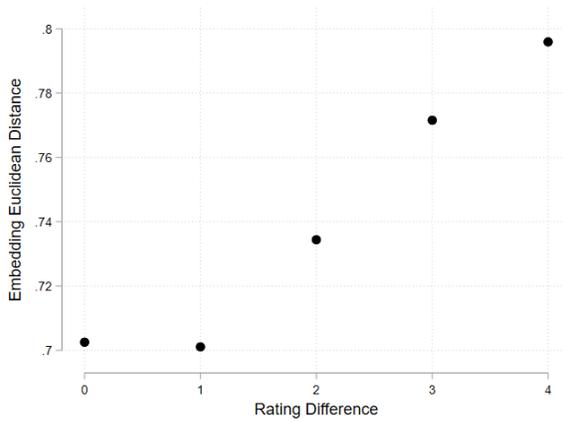
<sup>29</sup>To ensure consistency before and after the 2003 change in the rating system (Harper and Konstan, 2015), we restrict our analysis to ratings with integer values (i.e., full stars), excluding half-star ratings.



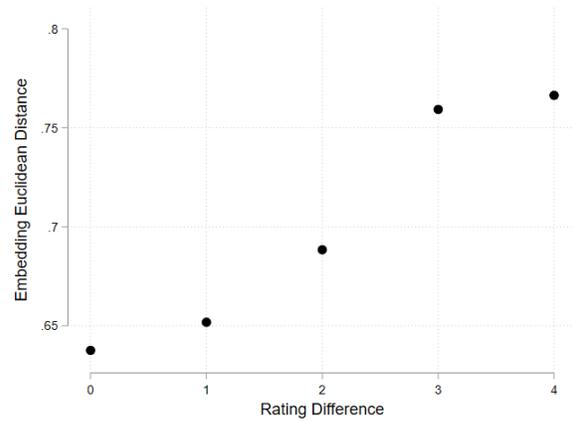
(a) *Crouching Tiger, Hidden Dragon (2000)*



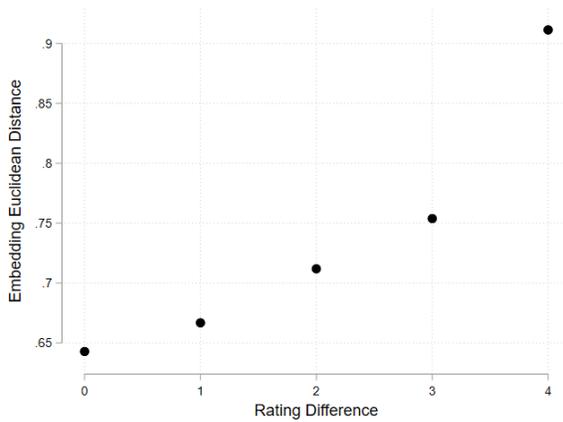
(b) *Lord of the Rings: The Fellowship of the Ring (2001)*



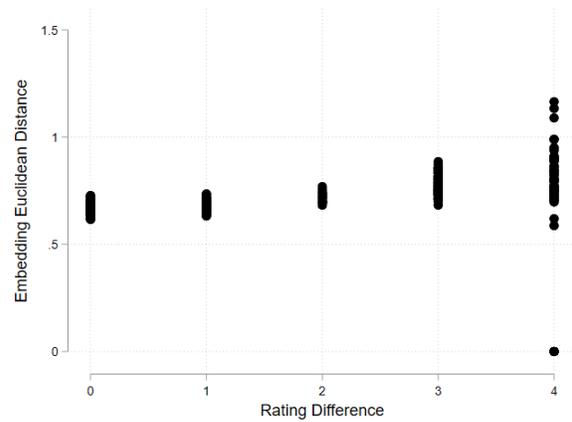
(c) *Lord of the Rings: The Return of the King (2003)*



(d) *Avatar (2009)*



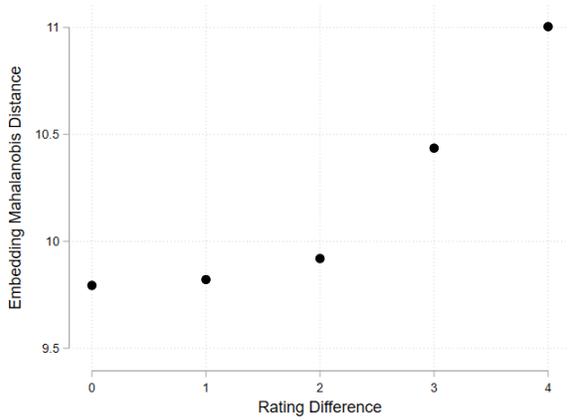
(e) *The Martian (2015)*



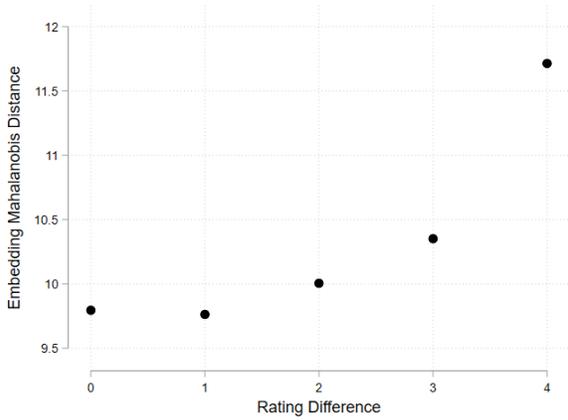
(f) 50 Most-Rated Movies

**Figure D1.** Euclidean Embedding Distances by Rating Difference  $\delta_{ik}$

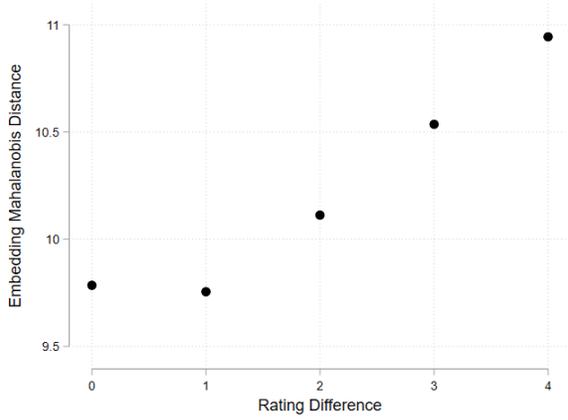
Notes: Notes: The graphs compare the average Euclidean distance between user embeddings for pairs of users who rated the same movie, grouped by their absolute rating difference. The six sub-figures correspond to: the five most-rated movies (*Crouching Tiger, Hidden Dragon*; *Lord of the Rings: The Fellowship of the Ring*; *Lord of the Rings: The Return of the King*; *Avatar*; and *The Martian*) and the results across all 50 most-rated movies.



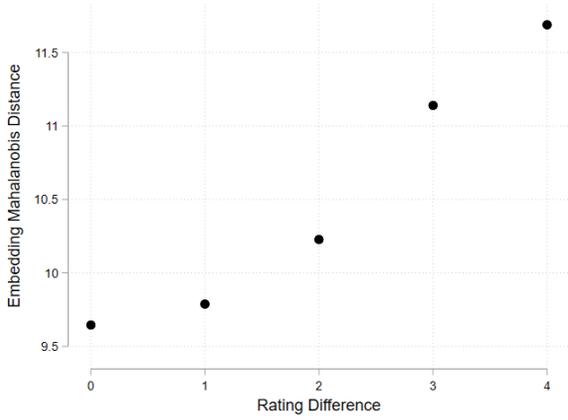
(a) *Crouching Tiger, Hidden Dragon* (2000)



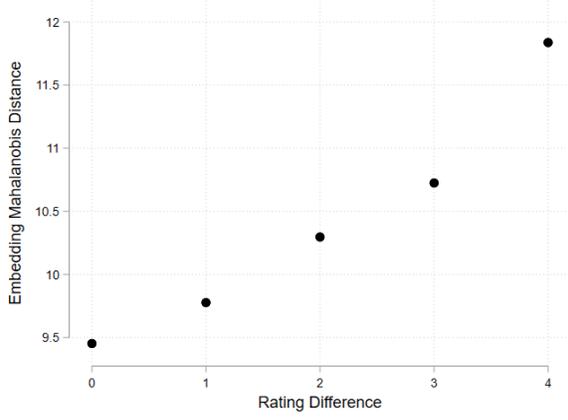
(b) *Lord of the Rings: The Fellowship of the Ring* (2001)



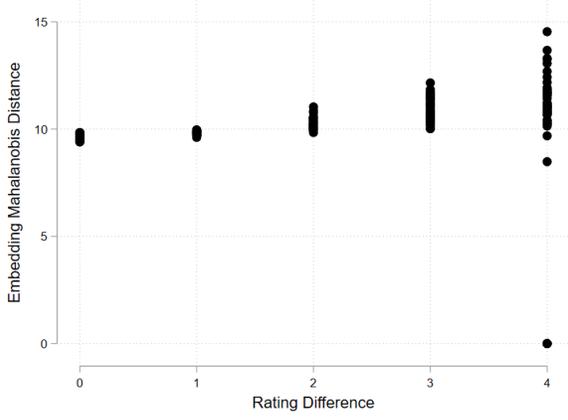
(c) *Lord of the Rings: The Return of the King* (2003)



(d) *Avatar* (2009)



(e) *The Martian* (2015)



(f) 50 Most-Rated Movies

**Figure D2.** Mahalanobis Embedding Distances by Rating Difference  $\delta_{ik}$

Notes: The graphs compare the average Mahalanobis distance between user embeddings for pairs of users who rated the same movie, grouped by their absolute rating difference. The six sub-figures correspond to: the five most-rated movies (*Crouching Tiger, Hidden Dragon*; *Lord of the Rings: The Fellowship of the Ring*; *Lord of the Rings: The Return of the King*; *Avatar*; and *The Martian*) and the results across all 50 most-rated movies.

## **D.4 Recommendation-based Matching - Robustness**

**Table D2.** Matching Users Before and After the AMPAS Nominations: embeddings with 10-dimensions

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	-0.039*** (0.007)	-0.036*** (0.007)	-0.052*** (0.007)	-0.044*** (0.007)	-0.040*** (0.010)	-0.046*** (0.016)
User Embeddings		✓	✓	✓	✓	✓
$\bar{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\bar{s}_i^j$					✓	✓
$Share_i^j$						✓
N	63849	59937	59975	55965	40354	18237
Mean Dep. Var.	3.919	3.919	3.919	3.919	3.921	3.921

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M5 report ATT estimates from a multivariate-distance matching with user embeddings (10 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D3.** Matching Users Before and After the AMPAS Nominations: embeddings with 20-dimensions

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	-0.039*** (0.007)	-0.036*** (0.009)	-0.042*** (0.007)	-0.037*** (0.008)	-0.032*** (0.010)	-0.042*** (0.016)
User Embeddings		✓	✓	✓	✓	✓
$\bar{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\bar{s}_i^j$					✓	✓
$Share_i^j$						✓
N	63849	60043	59974	56068	40266	18165
Mean Dep. Var.	3.919	3.919	3.919	3.919	3.921	3.921

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M5 report ATT estimates from a multivariate-distance matching with user embeddings (20 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D4.** Matching Users Before and After the AMPAS Nominations: embeddings with 100-dimensions

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	-0.039*** (0.007)	-0.040*** (0.008)	-0.042*** (0.007)	-0.046*** (0.007)	-0.040*** (0.010)	-0.048*** (0.017)
User Embeddings		✓	✓	✓	✓	✓
$\bar{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\bar{s}_i^j$					✓	✓
$Share_i^j$						✓
N	63849	60015	59979	56210	40384	18266
Mean Dep. Var.	3.919	3.919	3.919	3.919	3.921	3.921

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M5 report ATT estimates from a multivariate-distance matching with user embeddings (100 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D5.** Matching Users Before and After the AMPAS Nominations: embeddings with 200-dimensions

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	-0.039*** (0.007)	-0.038*** (0.008)	-0.039*** (0.008)	-0.043*** (0.008)	-0.039*** (0.011)	-0.048*** (0.015)
User Embeddings		✓	✓	✓	✓	✓
$\bar{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\bar{s}_i^j$					✓	✓
$Share_i^j$						✓
N	63849	59980	59859	56322	40338	18261
Mean Dep. Var.	3.919	3.919	3.919	3.919	3.921	3.921

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M5 report ATT estimates from a multivariate-distance matching with user embeddings (200 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D6.** Matching Users Before and After the AMPAS Nominations: LensKit MF Recommender Algorithm and Embeddings

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	-0.039*** (0.007)	-0.036*** (0.008)	-0.049*** (0.007)	-0.043*** (0.008)	-0.037*** (0.010)	-0.043*** (0.016)
User Embeddings		✓	✓	✓	✓	✓
$\vec{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\vec{s}_i^j$					✓	✓
$Share_i^j$						✓
N	63849	59869	59815	55914	39993	18093
Mean Dep. Var.	3.919	3.919	3.919	3.919	3.921	3.921

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M5 report ATT estimates from a propensity score matching with user embeddings (50 dimensions - Lenskit matrix factorization algorithm), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D7.** Matching Users Before and After the AMPAS Nominations: Removing Days Right After the Nominations

	(All)	(5 Days After)	(10 Days After)	(15 Days After)	(No First Session After Nom.)
Post Nomination	-0.047*** (0.015)	-0.040** (0.016)	-0.042** (0.020)	-0.042** (0.020)	-0.034** (0.017)
User Embeddings	✓	✓	✓	✓	✓
$\bar{r}_i^j$	✓	✓	✓	✓	✓
$N_i^j$	✓	✓	✓	✓	✓
$\bar{s}_i^j$	✓	✓	✓	✓	✓
$Share_i^j$	✓	✓	✓	✓	✓
N	18218	16038	13733	10964	14900
Mean Dep. Var.	3.921	3.922	3.921	3.921	3.923

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. In all columns we report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics (as in Column (M5) in Table 4). Column 1 (All) reports results based on all ratings in the window. Columns 2 to 4 (5 Days After; 10 Days After; 15 Days After) restrict the sample to ratings posted within 5, 10, and 15 days following the nomination date, respectively. Column 5 (No First Session After Nom.) restricts the sample excluding ratings that are written during the user’s first active session on MovieLens after the nomination. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D8.** Matching Users Before and After the AMPAS Nominations: Users who Posted More than 50 Ratings Before Ratings the Nominated Movie

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	-0.038*** (0.007)	-0.044*** (0.008)	-0.044*** (0.008)	-0.048*** (0.009)	-0.040*** (0.010)	-0.049*** (0.017)
User Embeddings		✓	✓	✓	✓	✓
$\bar{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\bar{s}_i^j$					✓	✓
$Share_i^j$						✓
N	61529	57829	57716	53872	38308	16815
Mean Dep. Var.	3.914	3.914	3.914	3.914	3.915	3.915

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. We exclude all users that posted less than 50 ratings on MovieLens before posting the rating for the nominated movie. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M5 report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D9.** Matching Users Before and After the AMPAS Nominations: Users who Posted More than 50 Ratings in Total

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	-0.039*** (0.007)	-0.041*** (0.008)	-0.043*** (0.008)	-0.046*** (0.009)	-0.041*** (0.010)	-0.052*** (0.017)
User Embeddings		✓	✓	✓	✓	✓
$\bar{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\bar{s}_i^j$					✓	✓
$Share_i^j$						✓
N	63054	59154	59130	55411	39616	17647
Mean Dep. Var.	3.917	3.917	3.917	3.917	3.919	3.919

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. We exclude all users that posted less than 50 ratings on MovieLens. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M5 report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D10.** Matching Users Before and After the AMPAS Nominations: Predicted Ratings (LensKit MF, UU, and II Recommender Algorithms)

	(MF)	(UU)	(II)
Post Nomination	-0.002 (0.005)	0.000 (0.005)	-0.003 (0.006)
User Embeddings	✓	✓	✓
$\bar{r}_i^j$	✓	✓	✓
$N_i^j$	✓	✓	✓
$\bar{s}_i^j$	✓	✓	✓
$Share_i^j$	✓	✓	✓
N	15472	15261	15353
Mean Dep. Var.	3.910	3.910	3.912

*Note:* The sample includes predicted ratings using matrix-factorization, user-user, and item-item LensKit recommender algorithms for movies rated from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. The first column is the OLS estimate from a regression of predicted ratings on movie fixed effects and a dummy equal to one if a predicted rating is related to a rating posted after the nomination date. Columns M1 to M5 report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D11.** Matching Users Before and After the AMPAS Nominations: Non-Nominated Movies

	(OLS)	(M1)	(M2)	(M3)	(M4)	(M5)
Post Nomination	0.001 (0.007)	-0.006 (0.009)	0.001 (0.008)	-0.015 (0.011)	-0.031* (0.017)	-0.023 (0.030)
User Embeddings		✓	✓	✓	✓	✓
$\bar{r}_i^j$			✓	✓	✓	✓
$N_i^j$				✓	✓	✓
$\bar{s}_i^j$					✓	✓
$Share_i^j$						✓
N	99608	83266	68369	42141	19578	6331
Mean Dep. Var.	3.397	3.397	3.397	3.397	3.399	3.399

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for non-nominated movies, excluding all users who have rated nominated movies. The first column is the OLS estimate from a regression of ratings on movie fixed effects and a dummy equal to one if a rating has been given after the nomination date. Columns M1 to M5 report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies rated by the user prior to the rating of the non-nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D12.** Matching Users Before and After the AMPAS Nominations: Number of Nominations and Awards

	( $Nom = 1$ )	( $Nom > 1$ )	( $Award = 0$ )	( $Award > 0$ )
Post Nomination	-0.049 (0.038)	-0.043** (0.020)	-0.037* (0.021)	-0.061*** (0.021)
User Embeddings	✓	✓	✓	✓
$\bar{r}_i^j$	✓	✓	✓	✓
$N_i^j$	✓	✓	✓	✓
$\bar{s}_i^j$	✓	✓	✓	✓
$Share_i^j$	✓	✓	✓	✓
N	4964	13277	10049	8200
Mean Dep. Var.	3.795	3.993	3.847	4.037

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. In all columns we report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics (as in Column (M5) in Table 4). Column 1 to 4 ( $Nom = 1$ ,  $Nom > 1$ ,  $Award = 0$ , and  $Award > 0$ ) restrict the sample to movies that received exactly one nomination, with more than one nomination, nominated movies that did not win any award, and nominated movies that won at least one award, respectively. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D13.** Matching Users Before and After the AMPAS Nominations: Movie Production Year

	(1995 – 2004)	(2004 – 2011)	(2011 – 2019)
Post Nomination	-0.080*** (0.028)	-0.059** (0.023)	0.004 (0.025)
User Embeddings	✓	✓	✓
$\bar{r}_i^j$	✓	✓	✓
$N_i^j$	✓	✓	✓
$\bar{s}_i^j$	✓	✓	✓
$Share_i^j$	✓	✓	✓
N	6863	5574	5478
Mean Dep. Var.	4.024	3.913	3.822

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. In all columns we report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics (as in Column (M5) in Table 4). Column 1 to 3 (1995 – 2004, 2004 – 2011, and 2011 – 2019) restrict the sample to movies that received at least one AMPAS nomination between 1995 and 2004, between 2005 and 2011, and between 2012 and 2019, respectively. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table D14.** Matching Users Before and After the AMPAS Nominations: User Characteristics ( $N_i^j$ ,  $\bar{s}_i^j$  and  $\bar{e}_i$ )

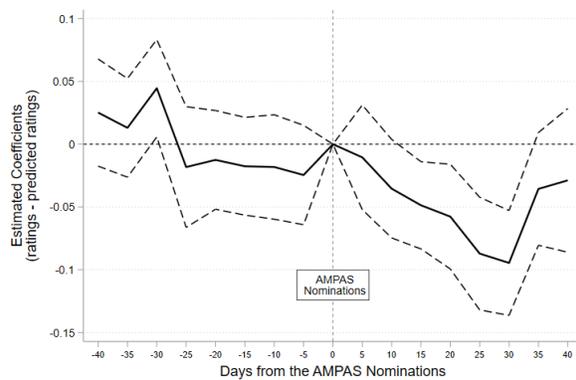
	( $N_i^j < 400$ )	( $N_i^j > 400$ )	( $\bar{s}_i^j < 0$ )	( $\bar{s}_i^j > 0$ )	( $\bar{e}_i < 0.5$ )	( $\bar{e}_i > 0.5$ )
Post Nomination	-0.047** (0.022)	-0.037 (0.025)	-0.063*** (0.020)	-0.035 (0.026)	-0.094*** (0.023)	0.032 (0.036)
User Embeddings	✓	✓	✓	✓	✓	✓
$\bar{r}_i^j$	✓	✓	✓	✓	✓	✓
$N_i^j$	✓	✓	✓	✓	✓	✓
$\bar{s}_i^j$	✓	✓	✓	✓	✓	✓
$Share_i^j$	✓	✓	✓	✓	✓	✓
N	9225	8402	8827	8185	4533	6337
Mean Dep. Var.	3.996	3.859	4.129	3.729	3.973	3.875

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. In all columns we report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics (as in Column (M5) in Table 4). Columns 1 to 6 ( $N_i^j < 400$ ,  $N_i^j > 400$ ,  $\bar{s}_i^j < 0$ ,  $\bar{s}_i^j > 0$ ,  $\bar{e}_i < 0.5$ , and  $\bar{e}_i > 0.5$ ) restrict the sample based on user characteristics: the number of prior ratings posted before rating the nominated movie ( $N_i^j$ ), the stringency level ( $\bar{s}_i^j$ ), and the average discrepancy between predicted and actual ratings before the nominations ( $\bar{e}_i$ ). Specifically, the columns correspond to users with fewer or more than 400 prior ratings; with stringency levels below or above 0; and with average discrepancies below or above 0.5, respectively. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

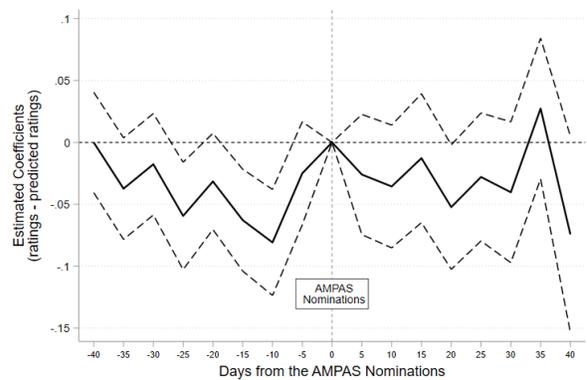
**Table D15.** Matching Users Before and After the AMPAS Nominations: User Characteristics ( $\bar{r}_i^j$  and  $Share_i^j$ )

	( $\bar{r}_i^j < 3.5$ )	( $\bar{r}_i^j > 3.5$ )	( $Share_i^j < 0.05$ )	( $Share_i^j > 0.05$ )
Post Nomination	-0.051* (0.028)	-0.049** (0.021)	-0.046** (0.023)	-0.047** (0.022)
User Embeddings	✓	✓	✓	✓
$\bar{r}_i^j$	✓	✓	✓	✓
$N_i^j$	✓	✓	✓	✓
$\bar{s}_i^j$	✓	✓	✓	✓
$Share_i^j$	✓	✓	✓	✓
N	6679	10353	8099	8807
Mean Dep. Var.	3.698	4.101	3.918	3.923

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. In all columns we report ATT estimates from a propensity score matching with user embeddings (50 dimensions), matching exactly on the movie identifier as well as quintiles of different user characteristics (as in Column (M5) in Table 4). Columns 1 to 4 ( $\bar{r}_i^j < 3.5$ ,  $\bar{r}_i^j > 3.5$ ,  $Share_i^j < 0.05$ , and  $Share_i^j > 0.05$ ) restrict the sample based on user characteristics: the average user rating posted before rating the nominated movie ( $\bar{r}_i^j$ ), and the share of movies rated in the same year ( $Share_i^j$ ). Specifically, the columns correspond to users with an average rating below or above 3.5; and with a share of movies rated in the same year below or above 0.05, respectively. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .



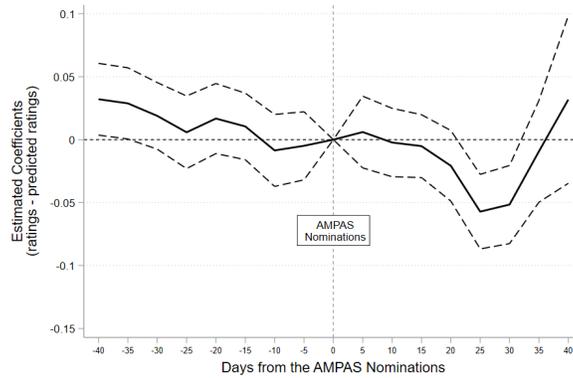
(a) Nominated Movies



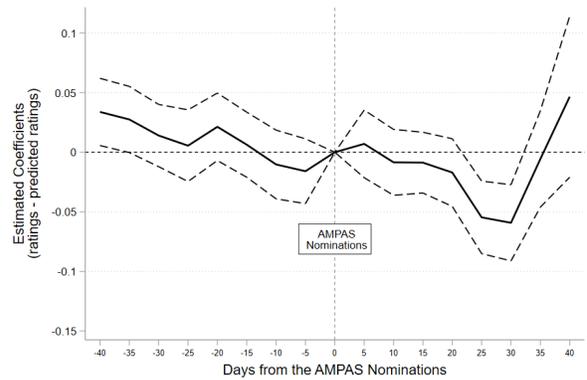
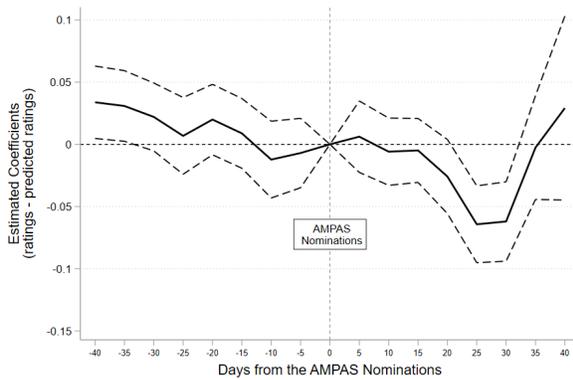
(b) Non-Nominated Movies

**Figure D3.** The Difference between Actual and Predicted Ratings over time: AMPAS Awards 40 Days After Nominations

*Notes:* The sample includes MovieLens ratings posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. Additionally, we restrict the sample to years in which the AMPAS awards take place at least 40 days after the nominations (these years are 1997–2003, 2013, 2014, 2016, and 2018). The difference between actual and predicted ratings is regressed on movie fixed effects and a set of five-day interval dummies centered on the AMPAS nomination date for nominated and non-nominated movies. The graphs display the estimated coefficients of these dummies, with the coefficient for the five-day interval of the nomination date normalized to zero. Standard errors, clustered by movie, are reported at the 10% level.



(a) Nominated Movies: LensKit Predicted Ratings (MF)



(b) Nominated Movies: LensKit Predicted Ratings (UU)

(c) Nominated Movies: LensKit Predicted Ratings (II)

**Figure D4.** The Difference between Actual and Predicted Ratings over time

*Notes:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The difference between actual and predicted ratings (using LensKit predicted ratings based on matrix-factorization, user-user, and item-item recommender algorithms) are regressed on movie fixed effects and a set of five-day interval dummies centered on the AMPAS nomination date for nominated and non-nominated movies. The graphs display the estimated coefficients of these dummies, with the coefficient for the five-day interval of the nomination date normalized to zero. Standard errors, clustered by movie, are reported at the 10% level.

**Table D16.** Difference-in-Differences: Exploiting Recommended Systems Predictions and Embeddings

	$(r_{ij})$	$(\hat{r}_{ij})$	$(r_{ij})$	$(r_{ij})$	$(r_{ij} - \hat{r}_{ij})$
$Nominated_j \times post_t^{Nom}$	-0.017*	-0.003	-0.014*	-0.017**	-0.014*
	(0.009)	(0.004)	(0.008)	(0.009)	(0.008)
Movie FEs	✓	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓	✓
Predicted Ratings			✓		
Embeddings				✓	
$R^2$	0.25	0.50	0.39	0.25	0.12
N	254,482	254,482	254,482	254,482	254,482
Mean Dep. Var.	3.503	3.539	3.503	3.503	-0.036

*Note:* The sample includes MovieLens ratings from 1995 to 2019, posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. In all specifications, we only include users with at least one rating 40 days before the nominations (a necessary condition to calculate predicted ratings and embeddings). Columns 1 to 5 ( $r_{ij}$ ,  $\hat{r}_{ij}$ ,  $r_{ij}$ ,  $r_{ij}$ , and  $r_{ij} - \hat{r}_{ij}$ ) report results from different Difference-in-Differences specifications, using distinct outcome or control variables: actual ratings (Column 1), predicted ratings (Column 2), actual ratings controlling for either predicted ratings (Column 3) or user embeddings (Column 4), and the difference between actual and predicted ratings (Column 5). Standard errors, clustered by movie, are shown in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## **E APPENDIX - External Validity: IMDb Data**

**Table E1.** Matching Users Before and After the AMPAS Nominations (IMDb Ratings): Number of Nominations and Awards

	( <i>Nom</i> = 1 )	( <i>Nom</i> > 1 )	( <i>Award</i> = 0 )	( <i>Award</i> > 0 )
Post Nomination	-0.136** (0.059)	-0.168*** (0.035)	-0.092** (0.039)	-0.224*** (0.028)
$\bar{r}_i^j$	✓	✓	✓	✓
$N_i^j$	✓	✓	✓	✓
$\bar{s}_i^j$	✓	✓	✓	✓
$Share_i^j$	✓	✓	✓	✓
N	3570	12573	7909	8234
Mean Dep. Var.	7.389	7.375	7.206	7.604

*Note:* The sample includes IMDb ratings posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. In all columns we report ATT estimates matching exactly on the movie identifier as well as quintiles of different user characteristics (as in Column (M5) in Table 4). Column 1 to 4 ( *Nom* = 1, *Nom* > 1, *Award* = 0, and *Award* > 0 restrict the sample to movies that received exactly one nomination, with more than one nomination, nominated movies that did not win any award, and nominated movies that won at least one award, respectively. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table E2.** Matching Users Before and After the AMPAS Nominations (IMDb Ratings): Movie Production Year

	( 1995 – 2004 )	( 2004 – 2011 )	( 2011 – 2019 )
Post Nomination	-0.260*** (0.074)	-0.150*** (0.056)	-0.134*** (0.033)
$\bar{r}_i^j$	✓	✓	✓
$N_i^j$	✓	✓	✓
$\bar{s}_i^j$	✓	✓	✓
$Share_i^j$	✓	✓	✓
N	2540	4337	9036
Mean Dep. Var.	7.646	7.499	7.225

*Note:* The sample includes IMDb ratings posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. In all columns we report ATT estimates matching exactly on the movie identifier as well as quintiles of different user characteristics (as in Column (M5) in Table 4). Column 1 to 3 (1995 – 2004, 2004 – 2011, and 2011 – 2019) restrict the sample to movies that received at least one AMPAS nomination between 1995 and 2004, between 2005 and 2011, and between 2012 and 2019, respectively. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table E3.** Matching Users Before and After the AMPAS Nominations (IMDb Ratings): User Number of Ratings

	$(N_i^j < 5)$	$(5 > N_i^j > 20)$	$(N_i^j > 20)$
Post Nomination	-0.190*** (0.042)	-0.159*** (0.052)	-0.107*** (0.038)
$\bar{r}_i^j$	✓	✓	✓
$N_i^j$	✓	✓	✓
$\bar{s}_i^j$	✓	✓	✓
$Share_i^j$	✓	✓	✓
N	8070	3559	3771
Mean Dep. Var.	7.167	7.283	7.598

*Note:* The sample includes IMDb ratings posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table displays estimated treatment effects for nominated movies. In all columns we report ATT estimates matching exactly on the movie identifier as well as quintiles of different user characteristics (as in Column (M5) in Table 4). Columns 1 to 3 ( $N_i^j < 5$ ,  $5 > N_i^j > 20$ , and  $N_i^j > 20$ ) restrict the sample based on the number of prior ratings posted before rating the nominated movie ( $N_i^j$ ). Specifically, the columns correspond to users with fewer than 5 prior ratings; with a number of prior ratings between 5 and 20, and with more than 20 prior ratings, respectively. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## **F APPENDIX - Textual Evidences of Disappointment**

### **Textual Analysis of Movielens Tags**

To identify disappointment-related sentiment in MovieLens user-contributed tags, we employed a fuzzy string matching approach built on a curated dictionary of disappointment-related expressions.

Tags are user-supplied and attached to movies independently of ratings. The dataset includes 41,321 tags associated with the same sample of movies and time window used in the difference-in-differences analysis. After deduplication and normalization (lowercasing and trimming whitespace), we applied a basic cleaning pipeline to remove stopwords and punctuation, though in practice, tags tend to be short and rarely contain such elements.

We curated a list of disappointment-related keywords through an iterative process:

- Manual review of a random subset of tags;
- Lexical expansion using thesaurus tools (e.g., WordNet) and prior domain-specific sentiment lexicons;
- Inclusion of common variations, abbreviations, and informal spellings.

The final keyword set included terms such as: “disappointing”, “overhyped”, “let down”, “underwhelming”, “expected more”.

To account for variability, we used fuzzy string matching via the `fuzzywuzzy` Python package (now maintained as `thefuzz`), implementing token sort ratio and partial ratio scoring. Matches were accepted based on a similarity threshold of 85%.

The following list contains the set of keywords and short phrases used to identify disappointment-related sentiment in MovieLens tags: “Disappointing”, “Let down”, “Over-rated”, “Overrated”, “disappointed”, “disappointing”, “disappointing ending”, “disappointments”, “expected more”, “let down”, “over-hyped”, “overrated”, “somewhat”, “overrated”, “way overrated.”

### **Textual Analysis of IMDb Reviews**

Unlike MovieLens tags, IMDb reviews are longer, richer, and more sentiment-focused. The dataset includes written reviews for every rating used in the matching analysis. Each review consists of free-form user text, often a paragraph or more.

Before embedding, reviews were preprocessed as follows: lowercasing; removal of punctuation and HTML formatting; Tokenization and lemmatization; removal of non-alphanumeric characters and isolated stopwords (e.g., “the”, “was”, etc.)

We employed sentence-transformer models to perform semantic similarity analysis. Specifically, we used the all-MiniLM-L6-v2 model from the Sentence Transformers library, which provides dense vector representations of text. This model was chosen for its balance between speed and performance and is widely used in short-to-medium length semantic matching tasks.

Procedure:

- **Encoding Reviews:** Each review was transformed into a 384-dimensional vector using the pretrained MiniLM model via sentence-transformers (version 2.2.2);
- **Reference Phrase Encoding:** For each target sentiment (e.g., disappointment, disinterest, viewing conditions), we defined a small set of seed phrases. Examples for Disappointment: “I am disappointed”, “This movie was overrated”, “It was overhyped”. Examples for Disinterest: “Not my style”, “I don’t enjoy these kinds of stories”, “I’m not a fan of the director”. Example for Viewing Conditions: “The cinema was crowded”, “The screen was too small”, “Streaming quality was not great”; Example for Reverse Bandwagon: “Mainstream movies rarely appeal to me”, “I avoid films that become mainstream successes”, “I don’t watch mainstream movies”;
- **Cosine Similarity Scoring:** We computed cosine similarity between each review vector and the set of reference vectors. For each sentiment type, the final score was defined as the maximum similarity to any reference phrase.

Below are sample reviews classified as highly likely to contain the target sentiment, based on being in the top 1% of the cosine similarity distribution with reference phrases for each sentiment category.

### **Disappointment**

- “One of the most overrated movies of all time. I personally do not get the hype with this movie. Most of it just did not interest me. at all.”
- “Overrated Overrated Overrated.”
- “I was expecting this movie to be a huge bore. Well, it wasn’t really, it was dull at times but not a bore or anything. Is it Overrated? Yes. I wasn’t WOWed by it like the critics were but i thought it was a good movie. Does it deserve a Best Picture nomination? NO. Did it deserve 9 nominations? NO.The acting was pretty good. Nicole Kidman gave a wonderful performance but she didn’t give the Best Actress performance of the year. Meryl Streep was very good but not great. She was far

better in Adaptation. Julianne was good but not worth a nomination. Her story was the weakest, what i mean is it was weakly written one. The dialogue wasn't as good and I never expected Julianne's character to leave in the end, that was the script's fault. Ed Harris was so-so, definitely didn't deserve a nomination."

### **Disinterest**

- "This movie was boring and absolutely uninteresting for me, if you wanna waste couple of hours, this movie is for you."
- "The film is bad, boring and no sense. One of those ends that just leaves you wondering why have you wasted two hours of your life."
- "I found this movie to be too slow moving and boring. At two hours and fifteen minutes, it is also far too long for the meager content. It just wasn't interesting, and tedious."

### **Viewing Conditions**

- "I watched it on Netflix. Wow this was hard to watch without falling asleep. Didn't bother finishing it. I really wanted to like this movie."
- "This movie was horrible.End of story.It lacked emotion, feeling, and any sense of mood.It was like watching the History Channel in vivid color without any narration. Oh, and Surround Sound. Sorry, I forgot to mention the fact that it was on a rather large screen."
- "May have looked great on an IMAX screen. Good sound effects, however, the rest is disappointing and boring."

### **Reverse Bandwagon** (some parts of the review text have been truncated due to excessive length)

- "One of Hollywood's worst clichés. It is, in my opinion, a film without content made just to create box office receipts. The music is painful to hear, the performance of the actors leaves much to be desired and is full of errors not acceptable for a project of this magnitude, the film is based on inspirations from other films, which gives the sensation of already having seen it."

- “A movie that mainstream Hollywood isn’t talking about. A wonderful story. Phenomenaly acted. Definite feel good movie.”
- ‘Firstly I would like to say that sometimes I genuinely worry about the world today. I admit that I went to see this film with very low expectations, I thought that this film was going to be, quite simply, a typical "American" comedy, and yet, I still went to see it. This is because I am not so full of myself that I can’t enjoy a film for what it is. A bloody good laugh! Not everything has to be deep and meaningful. As I’ve said above get some perspective, ITS JUST A FILM!’”

As a complementary method to semantic similarity scoring, we also employed a pre-trained transformer model fine-tuned specifically for emotion classification. The `SamLowe/roberta-base-go_emotions` model, available on Hugging Face, was trained on the GoEmotions dataset, a large corpus of Reddit comments labeled with 28 emotion categories including “disappointment”. This model allows us to generate probability scores for each emotion class directly from the full review text.

Before inputting reviews into the classifier, we applied the same text preprocessing steps as in the semantic similarity approach: lowercasing, punctuation and HTML removal, tokenization and lemmatization, removal of non-alphanumeric characters, and exclusion of isolated stopwords.

We used the Hugging Face Transformers library (version 4.39.3) and the `pipeline` utility for text classification. This model outputs a set of emotion-probability pairs for each review. We retained the predicted probability score for the “disappointment” label and used it as a continuous measure of disappointment sentiment.

Procedure:

- Preprocessing: Text normalization as detailed above;
- Classification: Reviews were passed to the `text-classification` pipeline using the `SamLowe/roberta-base-go_emotions` model;
- Score Extraction: For each review, we extracted the predicted probability assigned to the “disappointment” label as a measure of sentiment strength.

Below are sample reviews classified as highly likely to contain disappointment sentiment, based on being in the top 1% of the predicted probability distribution for the “disappointment” label.

## **Disappointment**

- “I was looking forward to this coming out on ppv. How disappointing !”

- “This disappointing film is remarkably patchy. Periods of real excitement and some terrific acting by Day-Lewis are interspersed with melodrama, some grotty special effects, and a story that includes coincidences and developments that are unbelievable. Altogether a film that does not enhance Scorsese’s reputation.”
- “I was appalled that they changed history. Go watch the real Queen on YouTube instead of this campy failure to recreate. Very disappointing.”

**Table F1.** Movielens Tags with Disappointment Sentiment

	Not-Nominated	Nominated	Difference	p-value	Obs. Not-Nominated	Obs. Nominated
Full sample	0.003	0.004	-0.002	0.018	33559	7762
Post nomination only	0.002	0.004	-0.002	0.012	11568	4176

*Note:* The table compares the rates of “disappointment sentiment” for Movielens tags posted by users in two distinct groups: those associated with movies that were nominated for the AMPAS awards and those that were not. The sample includes Movielens tags posted within an 80-day window around the AMPAS nomination date, excluding ratings posted after the AMPAS award ceremony. The table report the means for each group, the difference between them, and the associated p-values.

**Table F2.** Matching Users Before and After the AMPAS Nominations: IMDb Reviews’ Disinterest Dimension

	(OLS)	(M1)	(M2)	(M3)	(M4)
Post Nomination	0.010** (0.004)	0.008 (0.007)	0.011 (0.007)	0.012 (0.008)	0.013 (0.009)
$\bar{r}_i^j$		✓	✓	✓	✓
$N_i^j$			✓	✓	✓
$\bar{s}_i^j$				✓	✓
$Share_i^j$					✓
N	54838	36398	32232	21767	16143
Mean Dep. Var.	0.499	0.467	0.467	0.465	0.465

*Note:* The sample includes IMDb reviews posted within an 80-day window around the AMPAS nomination date, excluding reviews posted after the AMPAS award ceremony. The dependent variable is a dummy variable that takes the value of 1 if the review’s cosine similarity score related to a set of expressions of disinterest exceeds the median value of 0.3663. The table displays estimated treatment effects for nominated movies. The first column presents the OLS estimate from a regression of reviews on movie fixed effects and a dummy variable equal to 1 if the review was posted after the nomination date. Columns M1 to M4 report ATT estimates, matching exactly on the movie identifier and quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies reviewed by the user prior to reviewing the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table F3.** Matching Users Before and After the AMPAS Nominations: IMDb Reviews' Viewing Conditions Dimension

	(OLS)	(M1)	(M2)	(M3)	(M4)
Post Nomination	0.008* (0.004)	0.005 (0.006)	0.008 (0.006)	0.002 (0.008)	0.002 (0.010)
$\bar{r}_i^j$		✓	✓	✓	✓
$N_i^j$			✓	✓	✓
$\bar{s}_i^j$				✓	✓
$Share_i^j$					✓
N	54838	36398	32232	21767	16143
Mean Dep. Var.	0.499	0.462	0.462	0.460	0.460

*Note:* The sample includes IMDb reviews posted within an 80-day window around the AMPAS nomination date, excluding reviews posted after the AMPAS award ceremony. The dependent variable is a dummy variable that takes the value of 1 if the review's cosine similarity score related to a set of expressions related to bad viewing conditions exceeds the median value of 0.2842. The table displays estimated treatment effects for nominated movies. The first column presents the OLS estimate from a regression of reviews on movie fixed effects and a dummy variable equal to 1 if the review was posted after the nomination date. Columns M1 to M4 report ATT estimates, matching exactly on the movie identifier and quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies reviewed by the user prior to reviewing the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table F4.** Matching Users Before and After the AMPAS Nominations: IMDb Reviews' Reverse Bandwagon Dimension

	(OLS)	(M1)	(M2)	(M3)	(M4)
Post Nomination	0.012*** (0.004)	0.010 (0.006)	0.009 (0.007)	0.006 (0.008)	0.009 (0.010)
$\bar{r}_i^j$		✓	✓	✓	✓
$N_i^j$			✓	✓	✓
$\bar{s}_i^j$				✓	✓
$Share_i^j$					✓
N	54838	36398	32232	21767	16143
Mean Dep. Var.	0.499	0.462	0.462	0.460	0.460

*Note:* The sample includes IMDb reviews posted within an 80-day window around the AMPAS nomination date, excluding reviews posted after the AMPAS award ceremony. The dependent variable is a dummy variable that takes the value of 1 if the review's cosine similarity score related to a set of expressions related to the reverse bandwagon effect exceeds the median value of 0.3545. The table displays estimated treatment effects for nominated movies. The first column presents the OLS estimate from a regression of reviews on movie fixed effects and a dummy variable equal to 1 if the review was posted after the nomination date. Columns M1 to M4 report ATT estimates, matching exactly on the movie identifier and quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies reviewed by the user prior to reviewing the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table F5.** Matching Users Before and After the AMPAS Nominations: IMDb Reviews' Disappointment (25% High Similarity Scores)

	(OLS)	(M1)	(M2)	(M3)	(M4)
Post Nomination	0.034*** (0.004)	0.036*** (0.005)	0.037*** (0.006)	0.037*** (0.007)	0.035*** (0.008)
$\bar{r}_i^j$		✓	✓	✓	✓
$N_i^j$			✓	✓	✓
$\bar{s}_i^j$				✓	✓
$Share_i^j$					✓
N	54838	36398	32232	21767	16143
Mean Dep. Var.	0.250	0.214	0.214	0.213	0.213

*Note:* The sample includes IMDb reviews posted within an 80-day window around the AMPAS nomination date, excluding reviews posted after the AMPAS award ceremony. The dependent variable is a dummy variable that takes the value of 1 if the review's cosine similarity score related to a set of expressions of disappointment exceeds the value of 0.4120 (75th percentile). The table displays estimated treatment effects for nominated movies. The first column presents the OLS estimate from a regression of reviews on movie fixed effects and a dummy variable equal to 1 if the review was posted after the nomination date. Columns M1 to M4 report ATT estimates, matching exactly on the movie identifier and quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies reviewed by the user prior to reviewing the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table F6.** Matching Users Before and After the AMPAS Nominations: IMDb Reviews' Disappointment (10% High Similarity Scores)

	(OLS)	(M1)	(M2)	(M3)	(M4)
Post Nomination	0.026*** (0.003)	0.026*** (0.003)	0.027*** (0.003)	0.029*** (0.005)	0.031*** (0.006)
$\bar{r}_i^j$		✓	✓	✓	✓
$N_i^j$			✓	✓	✓
$\bar{s}_i^j$				✓	✓
$Share_i^j$					✓
N	54838	36398	32232	21767	16143
Mean Dep. Var.	0.100	0.083	0.083	0.083	0.083

*Note:* The sample includes IMDb reviews posted within an 80-day window around the AMPAS nomination date, excluding reviews posted after the AMPAS award ceremony. The dependent variable is a dummy variable that takes the value of 1 if the review's cosine similarity score related to a set of expressions of disappointment exceeds the value of 0.4569 (90th percentile). The table displays estimated treatment effects for nominated movies. The first column presents the OLS estimate from a regression of reviews on movie fixed effects and a dummy variable equal to 1 if the review was posted after the nomination date. Columns M1 to M4 report ATT estimates, matching exactly on the movie identifier and quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies reviewed by the user prior to reviewing the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table F7.** Matching Users Before and After the AMPAS Nominations: IMDb Reviews’ Disappointment (transformer-based classification model)

	(OLS)	(M1)	(M2)	(M3)	(M4)
Post Nomination	0.010*** (0.002)	0.012*** (0.003)	0.011*** (0.004)	0.014*** (0.005)	0.016*** (0.006)
$\bar{r}_i^j$		✓	✓	✓	✓
$N_i^j$			✓	✓	✓
$\bar{s}_i^j$				✓	✓
$Share_i^j$					✓
N	54838	36398	32232	21767	16143
Mean Dep. Var.	0.080	0.071	0.071	0.071	0.071

*Note:* The sample includes IMDb reviews posted within an 80-day window around the AMPAS nomination date, excluding reviews posted after the AMPAS award ceremony. The dependent variable is a dummy variable that takes the value of 1 if the predicted probability of the “disappointment” label exceeds the value of 0.5. The table displays estimated treatment effects for nominated movies. The first column presents the OLS estimate from a regression of reviews on movie fixed effects and a dummy variable equal to 1 if the review was posted after the nomination date. Columns M1 to M4 report ATT estimates, matching exactly on the movie identifier and quintiles of different user characteristics. All characteristics are time-varying, based on the set of movies reviewed by the user prior to reviewing the nominated movie. Standard errors are obtained via bootstrap (100 replications). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .