CAGE

Social Influence in Online Reviews: Evidence from the Steam Store

CAGE working paper no. 714

July 2024

Adam Di Lizia

WARWICK
THE UNIVERSITY OF WARWICK

Economic and Social Research Council

Social Influence in Online Reviews: Evidence from the Steam Store*

Adam Di Lizia †

July 16, 2024

Abstract

How does social influence affect consumer ratings? Using a dataset from the popular *Steam* gaming platform I investigate how quality judgements depend on pre-existing consumer assessments. In 2019, *Steam* introduced a new review system which decreased the exposure of users to previous ratings. Firstly, I find that user ratings are dependent on average ratings. A 10 percent increase in average rating increases the probability a review is positive by 5.4% before the policy change, but only by 2.8% after. The result is not due to selection, and is robust to a wide range of alternative specifications. Secondly, the effect is heavily asymmetric: individual reviewers are more negative when exposed to a lower average rating, but do not respond to a higher one. This negativity compounds and inflates the gap between lower rated and higher rated games. Overall, these social influence effects are driven by less experienced users on the platform. Finally, using estimates of owner data, I run a structural model of game choice. A 1% increase in rating is equivalent to a 2.5 dollar price reduction. This suggests social influence has large implications for buyers and sellers.

*I would like to thank Mirko Draca, Ao Wang, Lily Shevchenko, Manuel Bagues, Mingli Chen, Carlo Schwarz, Dan Bernhardt and my MRes cohort for their advice and feedback. I am grateful to the ESRC for helping fund this research agenda. All mistakes are my own.

†Department of Economics, University of Warwick

1 Introduction

User reviews are an essential part of the online experience, and vital in understanding how purchasing decisions are made. A platform's reviews are crucial to users choosing between products, and are arguably one of the biggest benefits to shopping online. It has been shown that these reviews have a large, causal impact on purchase decisions (Anderson and Magruder 2012; Reinstein and Snyder 2005), and that firms care about these reviews enough to attempt to manipulate them (Mayzlin, Dover, and Chevalier 2014). The online market is both large and growing[1] and therefore understanding how reviews are generated is important for understanding the decisions made by the modern consumer.

In theory, these reviews should be representative and informative about a product's quality. This does not often hold in practice. On many platforms, users are exposed to the ratings of others when leaving their review. The issue explored in this paper is whether those opinions influence reviewers. This "social influence" would persist across many platforms and harm the ability of a consumer to accurately gauge product quality.[2]

Examining these issues in practice is difficult. This paper's key contribution is leveraging an exogenous shock in the information available to users to causally estimate social influence effects. I also analyse who is most susceptible to social influence, and how social influence affects sales.

I use data from *Steam*, a popular gaming platform and the main retailer for PC games, to identify how the average rating affects user reviews. *Steam* is a large platform that distributes video games over the internet. It is both a substantial market in its own right and is typical of the review systems across the internet. Further description of *Steam*, its structure and content can be found in the next section and Appendix A. Importantly, only those who own the game can leave a review.

An update to the platform in 2019 resulted in reviewers no longer seeing the average rating by default when leaving their own review. Before users had to go to the store page to leave a review and thus were exposed to the average rating. Since the update, users are able to review directly, without visiting the store, avoiding the average rating.[3]

This work sits as part of the larger peer effects literature. Manski (1993) shows the challenges of identification in this setting, which this paper will deal with through quasi-experimental variation. More specifically, this work belongs to the reduced-form style as described by Sacerdote (2014). This type of work often looks at peer effects and student performance, such as Angrist and Lang (2004) or Imberman, Kugler, and Sacerdote (2012), using natural experiments for identification. A non-educational example of this type of peer effects design is Aral and Nicolaides (2017), who use data from the *Strava* running app and exogenous variation in weather to show that users exercise more when their peers do.

I apply this framework to online reviews with two key benefits. Firstly, the common in-means specification has a concrete interpretation. Users actually observe the average rating directly, rather than it being a proxy for more general inter-

[1] The fraction of all purchases made that are online is around 16%

[2] Selection bias is also a key issue. This is the topic of Di Lizia (2024) looking at the same Steam data.

[3] Care will be taken to ensure selection of different types of reviews do not cause the results in section 5

dependencies. Furthermore, due to the nature of the update I can separately identify the social influence effect from the more general effect of peer group characteristics.

I show that this exogenous change in the information set has substantial impacts on reviewing behaviour. A 10 percent[4] increase in average rating increases the probability a review is positive by 5.4% before the policy change, but only by 2.8% after. This suggests user ratings do depend on the average rating. Overall, games benefit from the policy, mostly due to more positive reviews across worse rated games. This suggests compounding negativity is the primary cause of social influence bias in this context. In a simple structural model I show that this change to rating affects sales. Consumers value a 1% increase in rating equivalently to a 2.5 dollar decrease in price, suggesting social influence does affect product choice.

This paper adds to the literature examining the importance of social influence across online platforms. Muchnik, Aral, and Taylor (2013) experimentally up-vote or down-vote comments on a news aggregator site and find an initial positive vote leads to a higher probability the comment is voted positively by others. Bursztyn and Jensen (2015) leverage a digital policy change across online courses where students were suddenly exposed to a leaderboard, and they show this information worsens student performance. More general herding effects across online reviews have been examined in the quantitative marketing literature. Lee, Hosanagar, and Tan (2015) and Sunder, Kim, and Yorkston (2019) both show that user rating and the rating of friends consistently move together across a movie website and board game website respectively. However, this could be driven by the similarity between users[5]. In my setting this is controlled for allowing the precise social influence effect to be identified.

Structure. The paper is structured as follows. Section 2 describes the background of Steam as a platform and the policy change. Section 3 onwards describes the data, framework and results, and Section 7 concludes.

2 Background

2.1 Broad Overview

Steam was founded by *Valve*, a gaming company, in order to automatically provide updates for their products over the internet. Eventually this evolved into a service where entire games could be sold without the need for a physical copy. Initially, only Valve-made games were offered on the platform, but by 2005 other game companies started distributing their products on *Steam*.[6] This made it easier to buy and sell games, and has become the main way to distribute PC games.

While competitors do exist, such as Epic Games, Steam enjoys an almost monopolistic status with around 75% of the market for PC game distribution. The value of Steam has expanded rapidly in the last decade as video games have become more digital. The service generates around $6 billion in revenue and has 132 million active users circa 2020. Thus while this research aims to provide a more general understanding of online reviews, *Steam* is an incredibly valuable platform in its own right.

[4] In absolute terms as average rating is measured as a percent in this setting

[5] Even when exposure to the average rating is removed in my setting, there is still a large positive correlation between average rating and user rating suggesting that similarity between users is a key reason for this correlation. See Table 1.

[6] In return they give Valve a 30% cut of all sales. This is 20% for larger games.

Grand Theft Auto 5, a popular game sold on Steam, is the highest grossing media product of all time[7]. For a further discussion on the types of games on Steam and Valve's goals as a platform, see Appendix A.

2.2 Steam Reviews

User reviews were introduced to the platform in November 2013.[8] Reviews are binary: users can either recommend or not recommend a game. Users can also write text explaining their rating. This information is aggregated to the consumer as an average rating (% who recommended) across all reviews and reviews left in the last 30 days (since 2016). Consumers can also search through the history of all posted reviews. These reviews display the reviewer's username, the number of reviews they have left, number of games they own, playtime at review and the text of the review. Users must own the game in order to leave a review.[9] A valuable feature of the platform is that games are tied to an individual's account, so the average user is unlikely to switch between accounts.

2.3 Policy Change

The biggest change to this review system occured as part of the "Library Update" on October 31, 2019. Before the update, the reviewer had to navigate back to the store page of a game to leave a review. This resulted in the user seeing the average rating.[10] After the update, when exiting a game played for more than a few hours users are given the option on the page from which the game was launched to leave a review. Users can post a review entirely within a pop-up box without being exposed to the average rating.[11] The old system is still present, but is more difficult to use. Even if some reviewers still use the old system, or seek out the average rating regardless, it unlikely to be common behaviour and at worst would attenuate the results.

It appears the aim of the update was to increase the sheer number of reviews. *Steam* notes in a blog post how the daily number of reviews has tripled after the update.[12] However, this also provides an excellent setting for exploring how reviews are generated and how platforms may influence this process.

[7]https://www.marketwatch.com/story/this-violent-videogame-has-made-more-money-than-any-movie-ever-2018-04-06.

[8]Originally, only the metacritic score - calculated by taking the average across all verified game critic reviews for a game- was displayed. It is unchanging once all critics have reviewed a given game, usually in the first week or two after release.

[9]This can occur by purchasing the product directly from Steam, or purchasing a Steam Key - a code giving them access. Only verified Steam purchases are aggregated into the average rating. In addition Steam for some games chooses not to incorporate reviews as part of review bombs (off topic review activity rating a game as negative for actions of the publisher or developer rather than game quality) into the average rating. These reviews can be identified while scraping to best recreate the closest rating measure. Nevertheless, as shown in Figure 3 and Table C.2, the effect is not sensitive to small changes in how rating is measured.

[10]See Figure A.2

[11]See Figure A.3

[12]See https://store.steampowered.com/news/group/4145017/view/1697229969004830849

3 Data

3.1 Review Data

Review data is scraped directly from the public *Steam* store pages of individual games. It includes user rating, text of review, time of review, username and other supplementary information. I collect all English language reviews for the top 250 games by their number of reviews in May 2019.[13] I obtain the history of reviews from a game's release to November 2023. The top 250 games represent around 20 % of all reviews left on the platform as a whole.[14]

3.2 Average Rating

The rating at time of review is not observed, but can be calculated as Steam would by averaging over all previous reviews at the time of review. This is an imperfect measure due to some reviews being updated[15] and a restriction to English language reviews. However, both are a small fraction of total reviews. In addition, for games released after 2016, I have daily ratings from *SteamSpy*. These ratings include non-English and unedited reviews and using them does not substantially change the results.

3.3 Text Analysis

Additionally, I calculate review length and sentiment of a text. Review length is calculated as the number of characters in a review. The sentiment measure, *Spacy text blob*, is taken directly from the popular *Spacy NLP library*, and is a measure of how positive the text is. It can differentiate between negation i.e "not good" is correctly recognised as a negative phrase. Sentiment supplements the binary rating with a continuous measure of like and dislike.

3.4 Price, Player Number and Sales Data

Separately, the full history of a game's price, sales, player numbers and discounts at the time of review since 2015 is scraped from *SteamSpy.com*[16] These will be useful controls; price and sale data will also be useful for running a simple structural model in section 6. See Appendix B for further comments on this data.

3.5 Descriptive Statistics

Descriptive statistics and simple correlations for the variables are shown in Table B.1. As shown in Figure 1, the update resulted in a large change to the average positivity of a review. Figure 1 includes only reviewers present before and after the policy change to avoid impact from user selection.

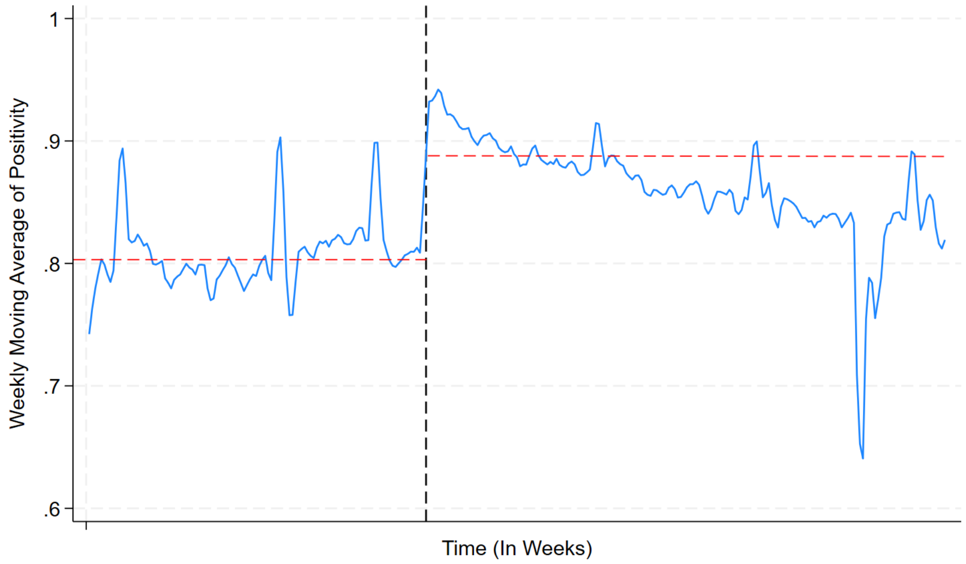
[13]Due to one of these games being re-branded as a 'game of the year edition" the final number of games in my sample is 249

[14]Calculated as a percentage of all reviews in November 2023

[15]Steam users can update their review after leaving it. When scraping reviews, only the most recent versions of reviews are seen. Since the observed characteristics of the review occured at the update, I treat updated reviews as being created at their date of update, but results are robust to alternate codes or dropping all updated reviews from the data set.

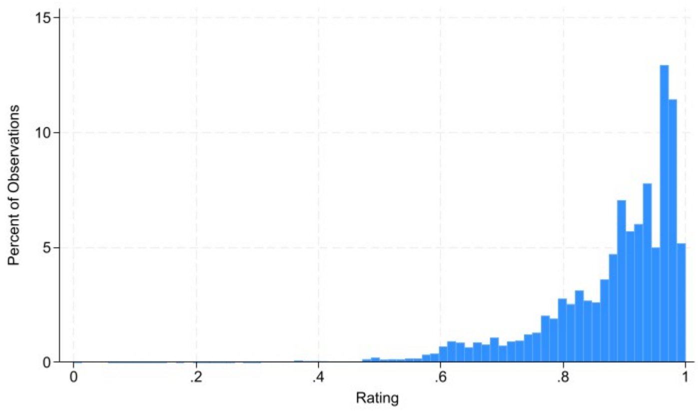
[16]See https://steamspy.com/ for the data and a wider discussion on owner data estimates.

Figure 1: Change in Positivity



Change in weekly average positivity across all games in the sample around the policy change, denoted by the dashed black line. This is the effect only on reviewers present before the policy change and after so is not driven by selection or attrition. This is a 3 period weekly moving average of the positivity of reviews across all games in a given week. The large spikes are related to sales or new game releases. In the full specification, these will all be controlled for.

Figure 2: Ratings Exposed to Users at Time of Review



Histogram of the average rating at time of given review. This is the distribution of calculated ratings displayed to users, across all games and time. The y-axis represents the percentage of all posted reviews that at time of writing were exposed to a rating within that bin.

Figure 2 shows how a game's average rating varies at time of review for each user review in the dataset. Most reviews are left when the average game rating is upwards of 60%. Similar to platforms like Uber, where 5 stars is the modal rating, games can reasonably be 100% positively rated. Therefore, games rated at 70% are not necessarily well rated, as this reflects that 30% of all users would not recommend the experience.

Basic correlations of characteristics within reviews can also be seen in Table B.2. Negative reviews tend to be longer, and be left by users who own more games.[17] The average rating correlates positively with user rating.

4 Empirical Analysis

To identify the effect of the average rating on user behaviour, I run the following regression:

$$y_{ijt} = \alpha_i + \mu_j + \tau_t + \beta \bar{y}_{jt} \times D_{jt} + \kappa D_{jt} + \lambda \bar{y}_{jt} + X'\gamma + \epsilon_{ijt} \tag{1}$$

In equation (1) y_{ijt} is the binary rating left by individual i for game j at review cycle time t. Review cycle time is calculated as the number of weeks since a game's release. $\bar{y}_{jt}$ is the average of all previous user ratings for game j at review cycle time t. This is the information displayed to the reviewer before the policy change but not afterwards. $\bar{y}_{jt}$ will be different for each new reviewer leading to variation within game. D_{jt} is a dummy for treatment which is equal to 1 if game j at review cycle time t is reviewed using the new system and 0 otherwise. X' is a vector of controls containing price, player numbers, playtime at review, and whether the game has had a recent update. α_i are the individual fixed effects of a given user. μ_j are game fixed effects. τ_t are the time fixed effects for each week since a game's release. ϵ_{ijt} is a stochastic error term. The key parameters of interest are β and κ.

The main issue with attempting to identify the effect of the average rating on individual user ratings is the co-movement between three different types of effects, as described by Manski (1993). This co-movement is known as the *reflection problem*. The first type, *endogenous effects*, refer to whether seeing a different average rating changes an individual reviewer's own behaviour. These are the main effects of interest. The second type, *exogenous effects*, refer to whether the average characteristics of a reviewer's reference group change their behaviour. Finally, *correlated effects* refer to the effects driven by the similarity between group members. Similar people generate similar responses but this is not due to interaction between them.

Manski (1993) shows that the presence of exogenous and correlated effects leads to difficulty in identifying the endogenous effect. In the case of online reviews, the key issue is the presence of correlated effects. Similar users review similar games at similar times. This leads to spurious correlation between the average rating and a user's rating. In this setting, however, the endogenous effect channel is shut down after the policy change. Reviewers no longer see the average rating by default. This change now allows separate identification of the endogenous effect. Intuitively, any shift in the correlation of the rating post-policy will capture the endogenous effect while keeping the exogenous and correlated channels constant, since the policy implementation was

[17] For further empirical observations, see Lin et al. (2019) which looks at these correlations within review.

unrelated to characteristics of games and users.[18]

The regression is run only on users present both before and after the policy change. Identification comes from the within-reviewer, within-game analysis. Equation (1) is a difference-in-differences design comparing the rating dependency for a given user, rating, game and week since release both before and after the policy change. Standard errors are clustered at the game level. I control for price and player numbers, as well as include time fixed effects for each week since a game's release to ensure trends within a game's life cycle do not drive the results.

Crucially, the policy change is an exogenous treatment that reduces the information seen by a user at time of review. If this policy changes the correlation between user and average rating, then the average rating causally impacts how an individual user reviews.

In the next section, I discuss the results as well as alternative mechanisms and potential concerns, such as selection. In short, the evidence strongly suggests a genuine change in the opinions formed by reviewers.

5 Results

5.1 Baseline Results

Table 1 shows the change in the rating dependency after the policy change. As expected, the probability of a positive review is highly correlated with average previous rating due to the reflection problem. Column 3 presents the main results. A 10 percent[19] increase in average rating increases the probability a review is positive by 5.4% before the policy change, but only by 2.8% after. This suggests that reducing the exposure of reviewers to the average rating reduces the correlation of their reviews with this rating. The total correlation between individual rating and average rating is around 0.9. Therefore, around 30% of the total correlation is driven by the endogenous effect.[20]

In addition, the probability to leave a positive review increases by 26.6% after the policy. This suggests heterogeneous marginal effects. Removing exposure to the average rating increases the overall positivity in the review system. However, this is distributed unevenly. Worse rated games do much better after the policy change relative to better rated games. To illustrate, a 50% rated game has a 13.5% higher probability to be rated positively after the policy change. In contrast, a 100% rated game would see no change to the probability it is rated positively. This suggests that the average rating reinforces negative opinions across games.

[18] The regression is run with the average rating rather than the average rating for the last 30 days for a few reasons. Mostly, it will be co-linear with average rating, especially for games with little variation month by month. While average rating was always displayed, average rating for the past 30 days was only displayed from 2016 onwards. Furthermore, the 30 day average rating will likely be very co-linear with both game and time fixed effects and user ratings due to it varying at a similar frequency.

[19] In absolute terms.

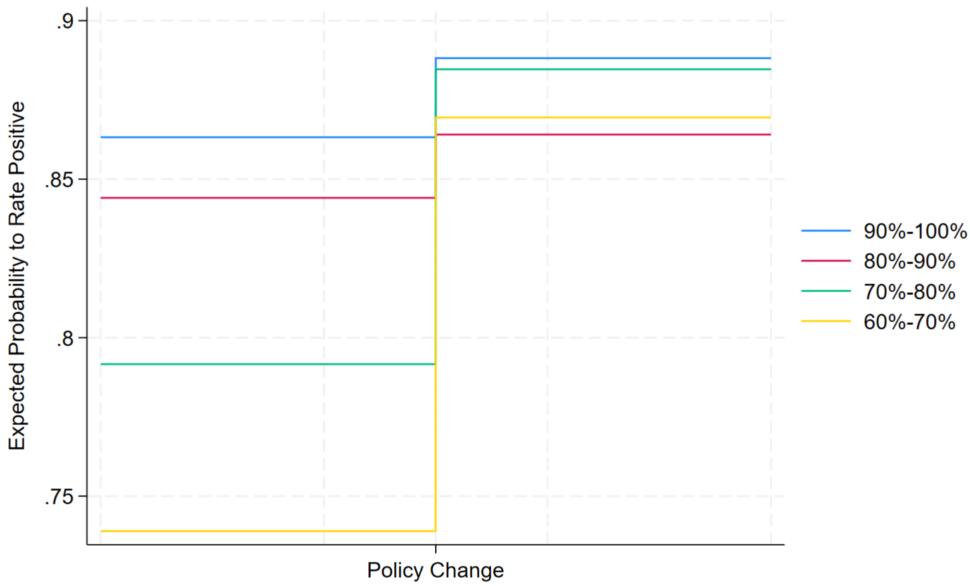
[20] There could be concern that updates might be driving the results if games change more after the policy. However there are on average fewer updates to the games in my sample after the policy change compared to an equivalent window before. Therefore at worst the coefficients would be underestimated. In addition Table C.5 runs the baseline including a unique Game $\times$ Time fixed effect for each game-week pair to attempt to fully capture updates, and the results are unchanged.

Table 1: Baseline Results

	(1) Positivity	(2) Positivity	(3) Positivity	(4) Sentiment
Policy × Rating	-0.179**	-0.156**	-0.269***	-0.060***
	(0.072)	(0.061)	(0.063)	(0.019)
Rating	0.900***	0.843***	0.544***	0.096***
	(0.047)	(0.046)	(0.080)	(0.033)
Policy	0.153**	0.120**	0.266***	0.063***
	(0.067)	(0.057)	(0.056)	(0.016)
Observations	6,510,809	6,403,981	6,403,981	6,403,981
Games (Clusters)	249	249	249	249
Controls	Yes	Yes	Yes	Yes
Individual FE	No	Yes	Yes	Yes
Game FE	No	No	Yes	Yes
Time FE	No	No	Yes	Yes

Results from the Baseline regression. Controls include price, player numbers, a dummy equal to 1 if any update has occured, review length and playtime at review. The sample is restricted only to users who left reviews before and after the policy change, which is 1,467,981 unique users leaving on average 5 reviews (on average 2.7 before and 2.3 after). See Figure C.1 for a frequency chart of total reviews left by those in the sample. See Table C.1 to see results are robust to excluding users above or below a certain number of reviews. There is a slight drop of observations when including fixed effects as while all users have left more than one review, some might have missing values for price if left before 2015. The outcome positivity refers to a binary variable which is 1 if the review was positive and 0 otherwise. Column 4 runs the same regression as column 3 but with sentiment as the dependent variable. This is an absolute measure between 0 and 1. Standard errors are clustered by game and in parentheses.

Figure 3: Group-Level Results



A graph of the change in expected probability to rate positive by rating group. The rating at time of review is split into 4 dummies, from 60% and increasing by 10 to 100%. These are the coefficients from a regression identical to equation (1) but using four rating bins rather than a continuous rating. The dependent variable is again the binary rating of a user review. This compares the effect on a review for a rating in that bin before and after the policy. To see the change in probability a review is positive for a given group, see the jump after the policy. Standard errors are clustered at the game level. See Table C.3 for the companion regression, and see Figure C.3 for the effect in the raw data.

Group Level Responses A straightforward way to see this asymmetry is to generate an indicator for each of the main rating groups in the sample: $60\% - 70\%$, $70\% - 80\%$, $80\% - 90\%$ and $90\% - 100\%$. These variables substitute $\bar{y_{jt}}$ in equation (1). The results are shown in Figure 3, which displays the probability that each group is rated positively. The lower the rating group, the greater the effect, with the $60\% - 70\%$ getting a 13% increase. $70\% - 80\%$ rated games receive a 10% increase, and $80\% - 90\%$ and $90\% - 100\%$ receive a 2% increase each. The result being so similar to the main specification even with such coarse aggregation suggests that small errors in the way average rating is calculated do not affect the results. An alternate specification allowing the effect to be non-linear in average rating is shown in Figure C.2, and supports this mechanism. Furthermore, this effect is even visible in the raw data, shown in Figure C.3. This strongly suggests that the average rating reinforces negativity in online review systems.

Alternate Rating Measures Results are robust to changing how the average rating is measured, as shown in Table C.2. Column 1 displays the results when considering reviews to be left when created rather than updated. Column 2 shows the results after dropping all updated reviews. In both cases the results are similar. In addition, results are robust to measuring average rating using *SteamSpy* data. Since *SteamSpy* only has data from mid 2015, the sample is restricted to games released from 2016. The effect is still present though smaller due to fewer games in this sample.

Sentiment Column 4 in Table 1 presents the results of estimating equation (1) using sentiment as the dependent variable. The coefficient on Policy reflects a 12% increase in positivity relative to the average before the policy change. The effect is again asymmetric: a game rated at 50% gets 6% more positive reviews relative to the mean before the update, with no change for a 100% rated game.

The effect is smaller, as the measure of sentiment is noisy. Additionally, we would expect the average rating to affect those most unsure about their review rather than those with strong opinions. Therefore we do not expect the absolute change in sentiment to be as high as the change in probability a review is positive. This exercise suggests social influence affects not just binary ratings but also the positivity of written text.

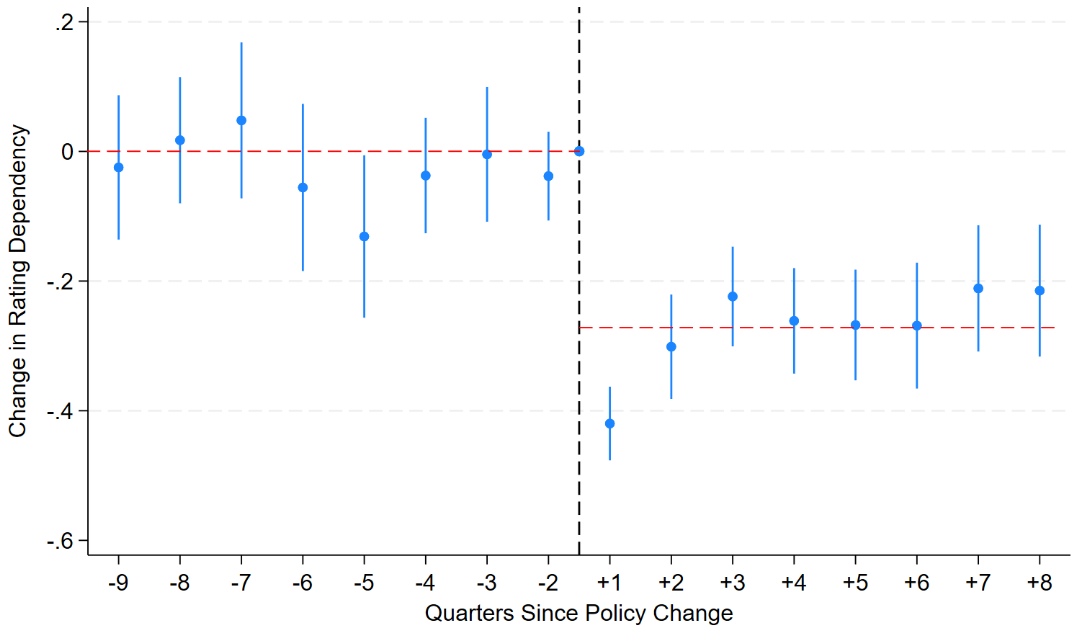
5.2 Selection Effects

Selection due to Change in Review Cost One potential concern is that the results are driven by selection bias. Under the new system, the individual cost of leaving a review may change. This could in turn affect which types of reviews a user chooses to leave, making them more positive purely through selection. For this selection to be consistent with my results, reviewers would have to choose to post different types of reviews only for games rated poorly, and gradually change this selection rule in response to the average rating. This seems unlikely.

Even if this were the case, the effect would not persist. Average ratings would adjust to become more positive over time and eventually average and user rating would correlate as they did before the update. To address this, I run the main specification allowing the coefficients to vary quarterly.

These results are shown graphically in Figure 4. Each point represents the change in rating dependency for a given quarter relative to the last quarter before the policy change. The coefficients only become significant after the policy change. The rating

Figure 4: Variation Around the Policy Change



The quarterly change in rating dependency is plotted above, using the quarter before the policy change as a base month. The specification includes all users to allow for comparisons between quarters before and after the policy change. Sales, player numbers and their interactions are also controlled for in this specification. The Steam Summer Sale for the last 2 weeks of June 2019 has been specifically omitted due to the magnitude of its flash sales. All rating coefficients before the policy change are insignificant bar 5 quarters previous which is barely significant. There is a clear sustained drop before and after the policy. Confidence Intervals are at the 95% level and are displayed as error bars, clustered at the game level.

stays less predictive of review positivity after the policy change, suggesting the results are not driven by user selection.

Selection due to Rating Exposure Alternatively, the results could be driven by users choosing not to post their review when exposed to a lower average rating. If this is the case, then after the policy change we would observe a larger number of reviews specifically on lower rated games. To test this, equation (1) is estimated with daily number of game reviews as the dependent variable and all other variables aggregated by day. Results are shown in Table C.4. The coefficient on Rating $\times$ Policy is positive, and all other coefficients are insignificant. Thus we do not observe an increase in the number of reviews for lower rated games after the update. Exposure to the average rating changes user opinion rather than how users post.

Selection Tests In general, if reviewers change the type of reviews they select to post we would expect the average review sentiment to change. Since the policy made it strictly easier to leave a review, we would expect that users post reviews for games they care less strongly about. This would be reflected in average sentiment moving closer to zero, that is a fall in the absolute value of sentiment. Column 1 of Table C.4 estimates equation (1) with the absolute value of sentiment as the dependent variable. There is very little effect on absolute sentiment after the policy change. Sentiment

Table 2: Heterogeneous Effects

	(1) Positivity	(2) Positivity
Policy × Rating	-0.156**	-0.372***
	(0.061)	(0.072)
Policy × Rating × Log Playtime	0.040***	0.050***
	(0.014)	(0.014)
Policy × Rating × Log Reviews		0.087***
		(0.012)
Policy	0.153***	0.358***
	(0.054)	(0.063)
Policy × Log Playtime	-0.049***	-0.059***
	(0.012)	(0.012)
Policy × Log Reviews		-0.084***
		(0.011)
Observations	6,403,981	6,403,981
Games (Clusters)	249	249
Controls	Yes	Yes
Individual FE	Yes	Yes
Game FE	Yes	Yes
Time FE	Yes	Yes

The regression is re-run including a full set of interaction terms. The additional interactions with ratings are included in the controls but not displayed for ease, as the interest is on the total effect of the policy change. Column 1 is run only on playtime at review, though the results are similar including lifetime number of reviews as shown in column 2. To see the effect of interest, note how the Policy × Rating coefficient is negative, but the further interactions are positive. The same holds for the Policy coefficient but in reverse. As experience increases, this total effect of negativity compounding falls. Note that since playtime is normalised, log of playtime will be negative for anyone below the mean so the effect of the policy for someone with 1 game owned and a small amount of playtime will be much larger than the effect in Table 1. Standard errors are clustered by game and in parentheses.

is not a perfect measure of opinion strength, but helps corroborate the presence of social influence.

5.3 Heterogeneity by Experience

To look at heterogeneity of results by user type, the regression is re-run interacting rating × policy with user's playtime at review[21] and lifetime number of reviews left. These are proxies of user experience within and across games. Of most interest is the user's playtime at review, as availability of the exact time spent with a product at time of review is a unique feature of the *Steam* platform.

[21]Playtime at review is normalised by dividing by the mean playtime at review for a given game. This ensures that a low playtime at review reflects an objective lack of experience rather than a short game.

As we can see from Table 2, experience reduces the susceptibility to social influence. An increase in the playtime at review for a given game or the lifetime number of reviews left both decrease the impact of the policy change. It makes sense that inexperienced users might be more likely to be swayed by the opinions of others compared to users who have left more reviews or played the game for longer. Less experienced users might supplement their lack of experience with the ratings of others. Reviewers try to be helpful to potential consumers, as it is purely a cost to the user to leave a review. Less experienced users who are unsure about the quality of a product may attempt to learn from the ratings of others when posting a review.

6 Consequences of Social Influence Bias

6.1 Outcome on Market Shares

Another question is the extent to which this social influence might affect sales. In general, work from Anderson and Magruder (2012) suggests that ratings do have causal impacts on purchase decisions. However, with *SteamSpy* approximations for owners of games, I can directly analyse the monetary impact of the policy change.

I run a simple multinomial logit BLP model in the style of the structural literature.[22] This allows aggregate market data to be used to make claims about an individual's product choice. *SteamSpy* data is available only from mid 2015, so only games released in 2016 or later are included. This then becomes a discrete choice problem across the 76 games in my sample and the outside option of some other game not included in my sample. Descriptive statistics for this data set are shown in Table B.3 and additional comments are made in appendix B. Following the literature[23] we assume the utility will take the form:

$$u_{ijt} = -\alpha p_{jt} + \beta x_j + \gamma r_{jt} + \xi_{jt} + \epsilon_{ijt} \tag{2}$$

Where p_{jt} are prices for product j at time t, x_j is the vector of observed product characteristics, r_{jt} is the rating of product j at time t, ξ_j is the unobserved product characteristics on utility, and ϵ_{ijt} is a Type-1 extreme value mean 0 error term. Product characteristics are taken from user submitted tags on games and then normalised. These tags describe the game, and are visible on the store page e.g single-player, adventure, action.

Care must be taken to ensure both rating and price are not endogenous. Price is surprisingly straightforward in this context. Games sold on *Steam* are digital copies and so have 0 marginal cost, and are in infinite supply. Prices do not have to adjust in response to demand to account for supply effects. This may not be true in the long run. For example, cheaper games might get updated more frequently[24] or games being kept at a low price might encourage the entry of other products. However, in the short run this is purely a demand shock, and common on *Steam* are many flash sales where products have their prices reduced for a short time. Often these sales are during certain times of the year such as Lunar New Year or Christmas, but games

[22]The framework will follow Nevo (2000). This model will be estimated by multinomial logit for simplicity as the goal is to identify the broad magnitude of the effect.

[23]See McFadden (1981)

[24]This is an important supply side consideration in Teng (2022) who looks at a structural model of the app store

Table 3: Multinomial Logit Results

	(1) Mkt. Share	(2) Mkt. Share	(3) Mkt. Share	(4) Mkt. Share
Rating	10.62***	10.88***	7.57***	11.19***
	(0.562)	(0.630)	(0.547)	(0.600)
Price	-0.041***	-0.041***	-0.058***	-0.021***
	(0.003)	(0.002)	(0.004)	(0.001)
Product Fixed Effects	Yes	Yes	No	Yes
BLP Instruments	Yes	No	Yes	No
Price Instruments	Yes	Yes	Yes	No

The results from a simple multinomial logit model on the aggregate market shares. The dependant variable is the total market share for product j in week w. The outside option is defined with an approximation on weekly aggregate Steam sales (around 40 million). Estimation occurs via linear IV-GMM, where price and rating are both endogenous except for column 4. Instruments for rating include Game $\times$ Policy, for price Game $\times$ Discount and for both the sum of characteristics of other products when BLP instruments are used. To ensure measurement issues are kept to a minimum, the product fixed effects or vector of characteristics are always interacted with a dummy which is 1 post April 2018. As discussed in appendix B, SteamSpy's owner estimation algorithm changed substantially at this time. When product fixed effects are not included, the instruments are instead Characteristics $\times$ Policy and Characteristics $\times$ Discount. Standard errors are in parenthesis

often go on sales semi-regularly at all times of the year. These short flash sales are a good instrument for game price.

Rating is also endogenous as choosing a given product means you value it, and are more likely to rate it highly. However, the policy change by game is a good instrument as it triggered an exogenous shift in rating. Furthermore, due to the nature of the update, the policy should not have affected sales in any other way. In addition, the classic BLP instruments of sums of product characteristics in the market will be used for both price and rating, though this will not be crucial for the results.

Results Table 3 shows the results of the simple multinomial logit estimation on the aggregate market data. Game specific dummies are used as suggested by Nevo (2000) and so product characteristics are absorbed.[25] We can see that rating does indeed have a positive effect on the market share for a given product. This is to be expected, but since the exogenous variation here is coming from the policy change itself this shows that policies reducing social influence do affect sales.

In terms of magnitudes, a 1% increase in rating[26] increases the market share relative to the outside option by 0.1%. A better understanding of the magnitude is to see that a 1% increase in rating has the same effect on market shares as a 2.5 dollar decrease in price. This reflects the importance to both sellers and consumers in reducing bias in online reviews. Consumers prefer to choose a product with a higher rating, and so a product where negativity has compounded due to social influence bias will receive fewer sales than in the counterfactual without this bias. This simple model supports the previous literature that suggests ratings affect purchase decisions.[27] This

[25] They can be recovered if we assume $E[\xi_{jt}|x] = 0$ but we are not particularly interested in these specific coefficients here.

[26] In absolute terms e.g 60% to 61% or 78% to 79%

[27] Simulating the policy is not necessarily revealing here. Ex-ante consumers would prefer the highest rating possible, even if this resulted in a worse matching probability. It is clear that varying

suggests that platforms would benefit from aiming to reduce social influence bias, as if negativity compounds sales will be artificially depressed.

7 Conclusion

Social influence is present in online review systems. I show that when a reviewer is exposed to the average rating of a product, their review is biased in line with that average rating. More importantly, this is driven by an asymmetrical effect. Negativity compounds in online reviews where the average rating is shown. Removing exposure to the average rating closes the gap between worse-rated and well-rated games. This effect is visible linearly, non-linearly and by rating group, and is robust to a battery of alternative measures. The effect is even visible in the raw data.

Many sites across the web show users the average rating before they leave a review. This paper suggests that this has serious outcomes for both product ratings and sales. An artificial gap between worse rated and better rated products is formed which not only makes the average rating a worse measure of quality, but likely reduces the sales of these products too.

I also show that this effect varies by user experience. Steam provides a unique opportunity to see how long a user spends with a product before they review. Social influence is driven by less experienced users on the platform. This has implications for products which might select for these types of users and could be more at risk to social influence.

Finally, using a simple structural model I show that the variation of the average rating due to the policy change has substantial impact on market share and consumer choice. Crucially, this bias affects sales. Games rated better after the policy sold better after the policy. The magnitudes suggest there is value in reducing this bias across platforms, especially if the asymmetrical compounding of negativity is common across review sites in general. Ultimately, platforms should take great care in how they display their average rating to incoming reviewers. It results in a social influence bias that affects both user rating and purchase decisions.

ratings has monetary value, but a structural model in this context cannot make claims about whether rating increases are good for matching users to games.

References

[1] Michael Anderson and Jeremy Magruder. "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database". In: *The Economic Journal* 122.563 (2012), pp. 957–989.

[2] Joshua D Angrist and Kevin Lang. "Does school integration generate peer effects? Evidence from Boston's Metco Program". In: *American Economic Review* 94.5 (2004), pp. 1613–1634.

[3] Sinan Aral and Christos Nicolaides. "Exercise contagion in a global social network". In: *Nature Communications* 8 (Apr. 2017), p. 14753. DOI: 10.1038/ncomms14753.

[4] Leonardo Bursztyn and Robert Jensen. " How Does Peer Pressure Affect Educational Investments? *". In: *The Quarterly Journal of Economics* 130.3 (May 2015), pp. 1329–1367. ISSN: 0033-5533. DOI: 10.1093/qje/qjv021. eprint: https://academic.oup.com/qje/article-pdf/130/3/1329/17098054/qjv021.pdf. URL: https://doi.org/10.1093/qje/qjv021.

[5] Jiafeng Chen and Jonathan Roth. "Logs with zeros? Some problems and solutions". In: *The Quarterly Journal of Economics* (2023), qjad054.

[6] Scott A Imberman, Adriana D Kugler, and Bruce Sacerdote. "Katrina's children: Evidence on the structure of peer effects from hurricane evacuees". In: *American Economic Review* 102.5 (2012), pp. 2048–2082.

[7] Young-Jin Lee, Kartik Hosanagar, and Yong Tan. "Do I follow my friends or the crowd? Information cascades in online movie ratings". In: *Management Science* 61.9 (2015), pp. 2241–2258.

[8] Dayi Lin et al. "An empirical study of game reviews on the Steam platform". In: *Empirical Software Engineering* 24 (2019), pp. 170–207.

[9] Charles F Manski. "Identification of endogenous social effects: The reflection problem". In: *The review of economic studies* 60.3 (1993), pp. 531–542.

[10] Dina Mayzlin, Yaniv Dover, and Judith Chevalier. "Promotional reviews: An empirical investigation of online review manipulation". In: *American Economic Review* 104.8 (2014), pp. 2421–2455.

[11] Daniel McFadden. "Econometric models of probabilistic choice". In: *Structural analysis of discrete data with econometric applications* 198272 (1981).

[12] Lev Muchnik, Sinan Aral, and Sean J Taylor. "Social influence bias: A randomized experiment". In: *Science* 341.6146 (2013), pp. 647–651.

[13] Aviv Nevo. "A practitioner's guide to estimation of random-coefficients logit models of demand". In: *Journal of economics & management strategy* 9.4 (2000), pp. 513–548.

[14] David A Reinstein and Christopher M Snyder. "The influence of expert reviews on consumer demand for experience goods: A case study of movie critics". In: *The journal of industrial economics* 53.1 (2005), pp. 27–51.

[15] Bruce Sacerdote. "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?" In: *Annual Review of Economics* 6.Volume 6, 2014 (2014), pp. 253–272. ISSN: 1941-1391. DOI: https://doi.org/10.1146/annurev-economics-071813-104217. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-economics-071813-104217.

[16] Sarang Sunder, Kihyun Hannah Kim, and Eric A Yorkston. "What drives herding behavior in online ratings? The role of rater experience, product portfolio, and diverging opinions". In: *Journal of Marketing* 83.6 (2019), pp. 93–112.

[17] Xuan Teng. "Self-Preferencing, Quality Provision, and Welfare in Mobile Application Markets". In: (2022).

Appendix A: Steam Background

Figure A.1: Example Review

An example of a simple review of the game 'Grand Theft Auto 5."

Steam Games

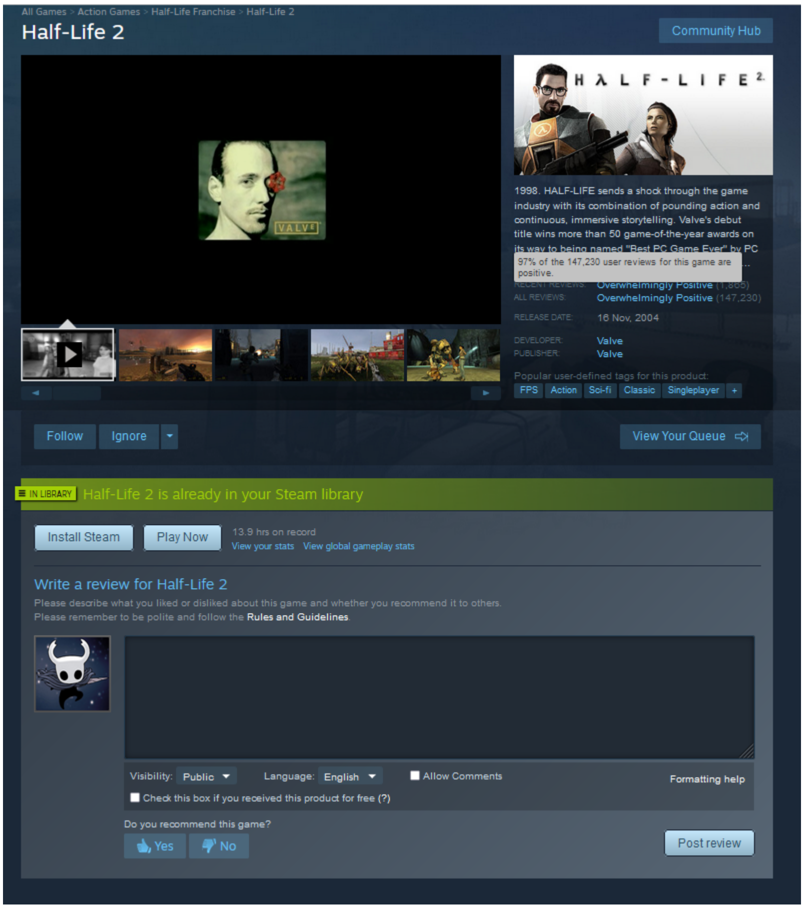
Games on Steam are often substantial experiences in line with games sold for consoles such as Xbox or Playstation. Most games available on those consoles since 2005 have been made available on the Steam store, except for some that are exclusive to a specific console. Some games are small experiences, cheap and short to finish, while some cost hundreds of millions to make and take many hours to complete. Some games have a multiplayer component and so are theoretically limitless in how much they can be played (some reviewers of online games have thousands of hours of playtime). One important aspect to keep in mind is that games, unlike movies or TV shows, can be updated or changed after the release. As a result, ratings can reasonably change over time and there can be large variation even within games. For example, often games can be released with glitches or 'bugs" that result in a poor user experience, but if the developers fix these issues then the game could be a much better experience. Steam has the largest selection of games across all platforms, including consoles, with around 30,000 products to choose from and so certainly represents well a wide variety of game types. Within my sample I have multi-player games, single-player games both short and long and even games made by a single person.

Valve's Goals

Valve has detailed in many blog posts [28] that their goal is to make a good review system that matches consumers to games well. Of course no company would admit to attempting to manipulate reviews to encourage purchase, but it does not seem that Valve is engaging in this type of behaviour. Valve has a voluntary refund policy whereby consumers who have played the game for less than 2 hours can get a guaranteed refund, so do not benefit selling games individuals do not enjoy. In addition, while Valve does not compete that closely with other online game distributors, they

[28] See https://steamcommunity.com/groups/steamworks/announcements/detail/1697229969000435735 for the policy change, user stats and https://store.steampowered.com/oldnews/24155 for a general discussion of making an accurate rating.

Figure A.2: Review System Before the Policy Change



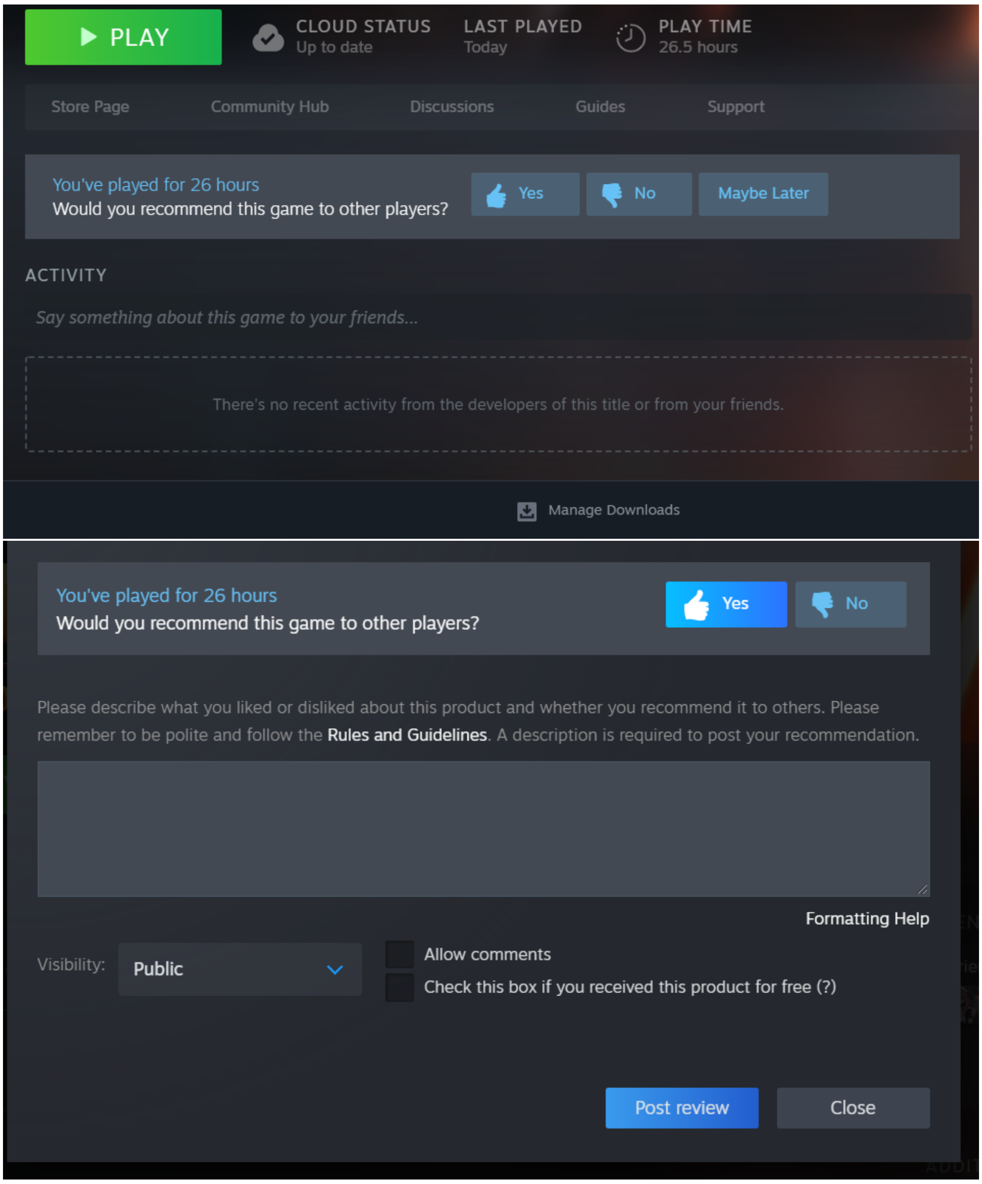
An example of the initial review posting system. The rating of others is immediately obvious when going to leave a review, and the exact rating is clearly visible above the review box. This channel was the only way of reviewing before the Library Update in 2019. This is still a usable channel for reviewers, but directly more obtuse than the new method. Even in the worst case where large groups of users still use the old method which seems unlikely, this would depress the coefficients in magnitude.

do compete with the outside option of illegally downloading games, or "pirating" games. Gabe Newell, the founder of Valve has gone on record detailing this:

'The easiest way to stop piracy is not by putting antipiracy technology to work. It's by giving those people a service that's better than what they're receiving from the pirates."

Broadly speaking, Steam as a platform has a goal that aligns with consumers in building a review system that increases the probability a game purchased is actually enjoyed by the consumer. This ensures users trust the review system and continue to purchase games on Steam. In addition fake reviews are unlikely to be an issue here. This paper focuses on the top 250 reviewed games at the time of the policy, and since an account must have purchased and played the game to be able to leave a review, it is not particularly feasible when the average game has 100,000s of reviews.

Figure A.3: Review System After the Policy Change



An example of the review posting system after the policy. The user is prompted upon going to play the game or exiting the game to leave a review directly. When they click on the yes or no buttons, the pop up box below appears and allows them to directly leave a review within the window. Note that the box is very similar to before, the only difference being the average rating not shown to the reviewer.

Appendix B: Data

Table B.1: Descriptive Statistics for Main Data

	Mean	Standard Deviation	Min	Max
Positive	0.88	0.32	0	1
Number of Games Owned	95.70	301.00	0	31,071
Number of Reviews Left	14.67	53.53	1	10,606
Total Playtime	31,601.10	72,630.80	0	6,049,533
Playtime in Last Two Weeks	107.16	695.32	0	127,474
Playtime at Review (Minutes)	15,740.75	42,609.80	1	4,776,595
Review Length (In Characters)	169.05	2,400.33	1	9,638,999
Policy (At Created)	0.55	0.50	0	1
Policy (At Updated)	0.58	0.49	0	1
Rating (At Created)	0.88	0.10	0	1
Rating Last 30 Days (At Created)	0.88	0.12	0	1
Rating (At Updated)	0.89	0.10	0	1
Rating Last 30 Days (At Updated)	0.89	0.12	0	1
Price (USD)	18.13	15.52	0	59.99
Percent of Total Owners Playing	12.97	16.46	0	100
Player Numbers	151,603.10	324,669.90	0	3,080,310
Updated	0.06	0.24	0	1
Rating (From SteamSpy)	0.87	0.11	0	1
Game Released after 2015	0.24	0.42	0	1
Log Playtime at Review	7.92	2.01	0	15.38
Log Number of Reviews	1.73	1.29	0	9.27
N	20,485,646			

Table B.2: Simple Correlations

	Positive	Length	Games	Playtime	Price	Play Num	Rating
Positive	1.00						
Length	-0.12	1.00					
Games	-0.04	0.13	1.00				
Playtime	0.00	0.00	-0.04	1.00			
Price	-0.12	0.12	0.05	-0.00	1.00		
Play Num	-0.13	-0.03	-0.05	0.06	0.01	1.00	
Rating	0.32	-0.07	-0.00	-0.04	-0.36	-0.28	1.00

Variables used as stored in the data set with associated descriptive statistics.

Table B.3: Descriptive Statistics for Weekly Market Data

	Mean	Standard Deviation	Min	Max
Price	27.60	18.60	0.00	59.99
Discount	8.03	18.82	0.00	92.00
Rating	0.84	0.13	0.31	0.99
SteamSpy Rating	0.84	0.13	0.29	1.00
Sales	26,459.66	114,338.61	0.00	3,937,275.00
Owners	3,930,472.45	7,064,129.74	1,000.00	85,161,578.75
Action	0.13	0.09	0.00	0.43
Adventure	0.09	0.08	0.00	0.29
Atmospheric	0.08	0.08	0.00	0.27
Classic	0.01	0.02	0.00	0.13
Co-op	0.06	0.08	0.00	0.30
Competitive	0.00	0.02	0.00	0.17
Dystopian	0.00	0.01	0.00	0.07
Fps	0.04	0.06	0.00	0.32
First Person	0.05	0.07	0.00	0.29
Free to Play	0.03	0.07	0.00	0.41
MMORPG	0.00	0.02	0.00	0.15
MOBA	0.00	0.02	0.00	0.16
Massively Multiplayer	0.02	0.05	0.00	0.27
Moddable	0.02	0.04	0.00	0.16
Multiplayer	0.14	0.11	0.00	0.41
Puzzle	0.01	0.04	0.00	0.22
Shooter	0.04	0.06	0.00	0.23
Singleplayer	0.14	0.10	0.00	0.40
Story Rich	0.06	0.09	0.00	0.34
Strategy	0.09	0.15	0.00	0.56
Team Based	0.01	0.03	0.00	0.11
E-Sports	0.00	0.01	0.00	0.07
N	16,098			

Descriptive Statistics for the data used in Section 6. Variables are a weekly average, and all taken from *SteamSpy* except for Rating which is calculated as described in the main text by aggregating reviews, then further aggregated by week. Using *SteamSpy* rating instead has virtually no difference on the regression. The horizontal line break shows the tags used in generating the vector of characteristics for each product j. Their sums are used to generate BLP instruments. The results are robust to excluding these tags altogether and using product fixed effects.

Review Data

Direct characteristics of the review are displayed both to users and when scraping, such as whether it was voted up, the number of games owned, number of reviews and playtime at review. Steam users can also rate reviews as funny or helpful, and this data is also collected though has little bearing on the specific channel of interest. The number of games owned and the number of reviews left is a measure that is not fixed at time of review, so it reflects the number of games owned or reviews left at the time of scraping. This is oddly how the data is displayed for Steam users too: when reading a review from someone with 100 reviews left, it could be that this was their first review and they have left 99 others since then. Some users can private the number of games they own if their profile is set to private. They are displayed as 0s in the data-set, but everything else is forced to be shown if a review is left.

SteamSpy

SteamSpy is a website that has been scraping the *Steam* Store and user profiles daily since mid 2015. The website as a result has the largest archive of price data, player data and additional data available for analysis. Price and player number data is fairly straightforward as described in the main text: *SteamSpy* is needed to save the daily history but price and player numbers are directly shown on *Steam* and so these are completely accurate for the given day.

Sales Data

Unfortunately, *Steam* does not reveal sales data for games. However, *SteamSpy* uses an algorithm that trawls through user profiles, sees which games users own, and then estimates the total number of owners based off this information. It is very accurate anecdotally according to the site creator and some developers who can compare estimates to their actual sales.

There are however issues. If games have free weekends, where a game is free to download for a weekend so users can try it out, then estimates for owners will be skewed. After the weekend users must then buy the game in order to keep playing. However, *Steam* incorrectly considers them owners for much longer than the weekend lasts. This can result in a reduction in owners spread out over oddly long periods of time, and therefore daily sales calculated as the difference in owners can be negative. If this occurs, the negative number is replaced with a missing value [29] This is far from ideal, but this is the most accurate data available to anyone outside of *Steam*. In addition, rather than look at daily markets, the structural model in Section 6 aggregates weekly to smooth over potential measurement issues.

Another issue discussed on *SteamSpy* is that in 2018 user profiles became private rather than public by default. The algorithm is still anecdotally accurate but the owner numbers jump for all games around that time. All regressions in Section 6 add a unique game dummy at this time to control for the mean shift after this date.

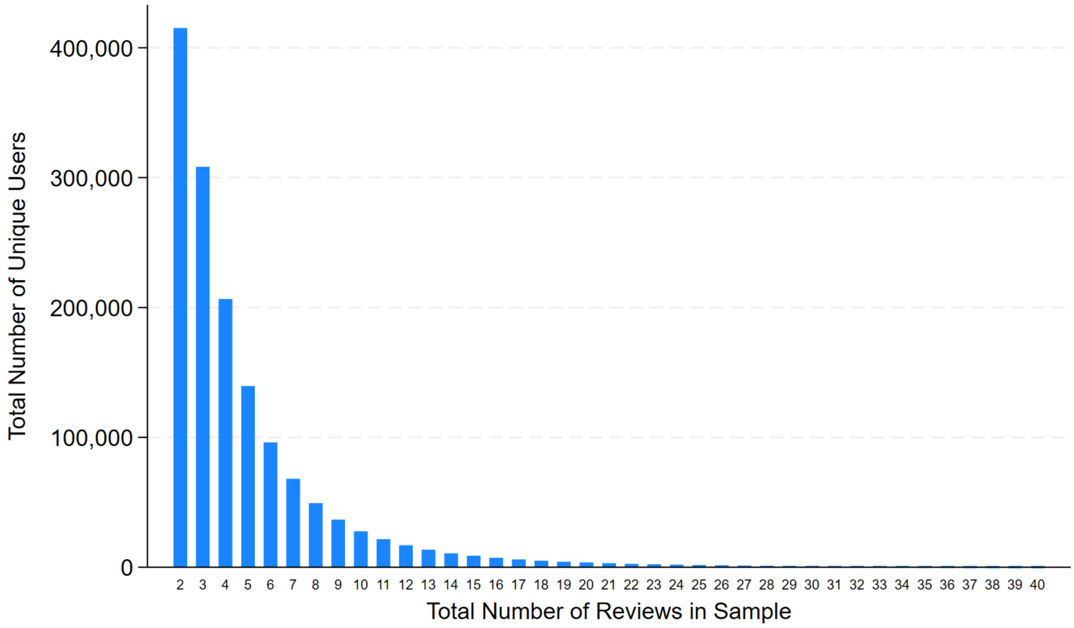
Game Tags

While not necessary for the results, the usual set of BLP instruments and game characteristics are constructed using user submitted tags for games. These include things like a game being single player, multi-player, sci-fi, survival, horror etc. This data is also available on SteamSpy. The tags change over time, but the proportions stay broadly similar and the tags for a game at May 2019 are used as measures of a game's characteristics. To account for bigger games being tagged more, the tags are normalised by the total number of tags left to get a proportion between 0 and 1. For example, game j tagged 50 times as single-player and 50 times as survival would have characteristics of 0.5 each. There are a huge list of possible tags, so the tags are pruned into more main categories and are shown in table B.3.

[29] see Chen and Roth (2023) for a discussion as to why replacing with a small value is not a good idea.

Appendix C: Additional Checks

Figure C.1: Frequency of Lifetime Reviews



This frequency chart displays the number of unique users with a given lifetime number of reviews left. This is exclusively taken from those with at least one review before the policy and one after, hence the minimum value at 2. The graph is cut off at 40 for ease of display but there are users with more than 100 reviews in the sample. While the modal number is 2, there are many in the sample leaving more than 2 reviews. We can see the average user leaves 5 reviews, and there are many 100,000s of users leaving more than 4 reviews. Note that the identification strategy demeans individuals from their fixed effects, and so provided demeaning can occur (i.e there is more than 1 observation) there is no advantage for an individual having 6 reviews rather than 5 except in how well the user can be demeaned. To confirm, the results displayed in Table C.1, are from re-running the baseline regression dropping users with a certain number of reviews and are very similar.

Table C.1: Robustness to Number of Reviews

	(1) Positivity	(2) Positivity	(3) Positivity
Policy	0.251^{***}	0.250^{***}	0.317^{***}
	(0.054)	(0.048)	(0.054)
Rating	0.547^{***}	0.538^{***}	0.539^{***}
	(0.078)	(0.076)	(0.078)
Policy $\times$ Rating	-0.252^{***}	-0.239^{***}	-0.331^{***}
	(0.061)	(0.053)	(0.061)
Reviews in Sample	> 2	> 4	< 4
Observations	4,781,368	2,355,491	3,882,213
Games (Clusters)	249	249	249
Controls	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Game FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes

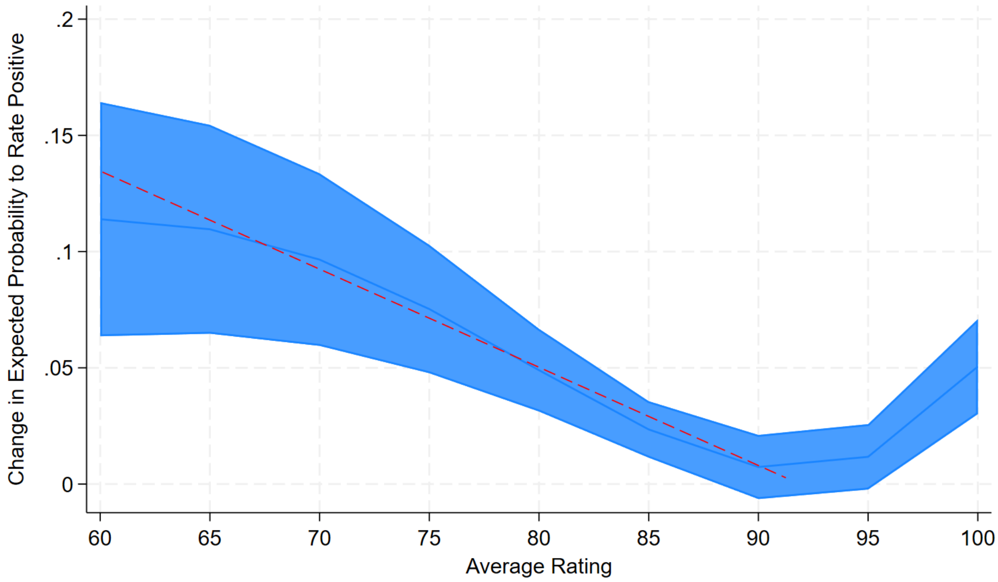
The baseline regression is re-run dropping users based on how many reviews they have left in the sample. This is to ensure results are not being driven by users with few reviews or many reviews. Column 1 drops all users with only 2 reviews. Column 2 drops all users with 4 or fewer reviews. Column 3 drops all users with more than 4 reviews. In all cases, the results stay very similar. The higher coefficients for column 3 compared to column 2 reflects the heterogeneity by experience. Reviews in sample correlate with lifetime number of reviews and hence we would expect the effect to be larger for these users. Standard errors are clustered at the game level and in parentheses.

Table C.2: Alternate Rating Measures

	(1) Positivity	(2) Positivity	(3) Positivity
Policy	0.220^{***}	0.318^{***}	0.141^{*}
	(0.060)	(0.054)	(0.074)
Rating	0.360^{***}	0.550^{***}	0.421^{***}
	(0.071)	(0.074)	(0.101)
Policy $\times$ Rating	-0.261^{***}	-0.328^{***}	-0.178^{**}
	(0.067)	(0.062)	(0.084)
Measure	Created	No Updated	SteamSpy
Observations	6,434,055	5,185,856	1,478,341
Games (Clusters)	249	249	76
Controls	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Game FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes

The baseline regression is re-run with alternate measures for rating. "Created" reflects that the date a review is considered to be left is when it was created rather than updated. "No updated" simply drops all updated reviews from the data set (around 2 million total, and 1 million from those leaving multiple reviews). "SteamSpy" uses the rating calculated from SteamSpy for games released in 2016 and after. This is available for 76 games in my sample, and is only daily. This lower variation results in the greater attenuation. Standard errors are clustered at the game level and in parentheses.

Figure C.2: Non Linear Responses



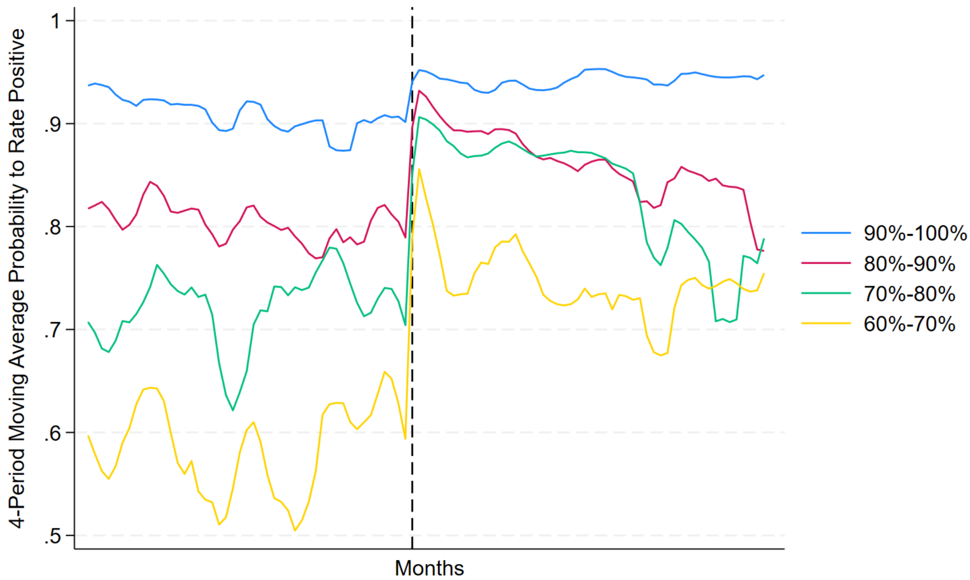
The main specification is re-run with both rating and rating interaction terms up to and including the fourth power. The above graph shows the marginal effect of the policy change at different levels of average rating. For example, for a game with an average rating of 60%, there would be an approximate 12% increase in the probability said game was rated positively. There would be no significant change to games with an average rating of 90%. The effect is calculated for each 5% jump in rating. Even allowing the dependency to take a more general functional form, the change is roughly linear with respect to rating up to about 90%, after which there is no clear effect for the highest rated games and the uptrend is likely an artefact of the functional form. Standard errors are clustered at the game level, and confidence intervals are at 95% coverage and calculated via the *margins* command in *Stata*.

Table C.3: Group Level Results

	(1)
	Positivity
60% - 70%	-0.124***
	(0.021)
70% - 80%	-0.072***
	(0.012)
80% - 90%	-0.019**
	(0.008)
Constant (90% - 100%)	0.884***
	(0.006)
Policy × 60% - 70%	0.106***
	(0.037)
Policy × 70% - 80%	0.068***
	(0.017)
Policy × 80% - 90%	-0.005
	(0.009)
Policy (90% - 100%)	0.025***
	(0.007)
Observations	6,203,303
Games (Clusters)	249
Controls	Yes
Individual FE	Yes
Game FE	Yes
Time FE	Yes

The companion regression to Figure 3. The main specification is repeated with rating bins rather than a continuous rating. The base category is the 90%-100% group before the policy change. All other aspects are identical to the main specification, other than reviews left at an average rating less than 60% are dropped due to too few observations. To see the total effect as displayed on the graph, add coefficients if the dummy variable is 1 to the constant. For example, the expected probability to rate positive for the 70-80 group after the policy is 0.884 - 0.072 + 0.068 + 0.025. Standard errors are clustered at the game level and in parentheses.

Figure C.3: Group Level Response in Raw Data



Similar to Figure 1, the raw average positivity is displayed except across different rating bins. This aims to as best as possible recreate Figure 3 in the raw data. Due to the large spikes and variability, data is aggregated monthly and a 4-period moving average is taken. This is to prevent the graph being so noisy that the lines constantly cover each other. The y-axis is the average probability to rate positive which simply takes the mean positivity of all user ratings in the bin for a given month. The x-axis is time in months. The dotted line marks the policy change. The graph is surprisingly similar to Figure 3 in just the raw data, supporting this being a genuine outcome of the policy that is clearly visible rather than an artefact of the regression specification.

Table C.4: Mechanism Checks

	(1) Absolute Value of Sentiment	(2) Daily Total of Reviews
Policy	0.008	-1.763
	(0.005)	(3.733)
Rating	0.026***	-10.358
	(0.008)	(9.818)
Policy × Rating	-0.009*	11.103**
	(0.006)	(4.444)
Observations	6,403,981	629,643
Games (Clusters)	249	249
Controls	Yes	Yes
Individual FE	Yes	-
Game FE	Yes	Yes
Time FE	Yes	Yes

Column 1 re-runs the baseline regression with the absolute value of sentiment as the dependent variable. If large individual selection occurs, we would expect the absolute value of sentiment to fall after the policy as reviews that users are more indifferent about are selected for. Results show that there is no significant reduction to the absolute value of sentiment post policy suggesting that results are not driven by selection of less extreme reviews. Indeed there is a minuscule increase for lower rated games which is likely just picking up the fact sentiment increases for lower rated games. This is an inaccurate measure but suggests review selection is not so significant. Column 2 re-runs the baseline regression with daily total of reviews left by those in the sample. All variables are aggregated by day hence the smaller number of observations and no individual fixed effects. This is done to see whether the effect is driven by users posting reviews on lower rated games they would not have when exposed to the average rating. The coefficient of interest is slightly significantly positive and therefore it seems the social influence bias is driven by an actual change in opinion. We would expect to see a negative Policy × Rating coefficient if selection was purely driving results, either generic or otherwise.

Table C.5: Game × Time Fixed Effects

	(1)
	Positivity
Policy	0.243***
	(0.058)
Rating	0.095***
	(0.105)
Policy × Rating	-0.240***
	(0.052)
Observations	6,402,502
Games (Clusters)	249
Controls	Yes
Individual FE	Yes
Game × Time FE	Yes

The regression is re-run including a unique fixed effect for each week-game pair. The variation here comes from looking at the rating correlation within a given game at a given week before the policy compared to after the policy. If there is a social influence effect, we would expect that even controlling for the average positivity in a given week we would still see more correlation when users are exposed to the average rating. Results are robust to these fixed effects. Even if games get updated or have good or bad weeks, it seems that this is not driving the results. Standard errors are in parentheses and are clustered at the game level.