Learning from Users: Impact of Online Reviews and Differences across Producer Types

The online user community serves as an important external source of information for both consumers and producers: it not only informs consumers' purchasing decisions but also influences producers' future research and development (R&D). However, the latter has received comparatively less attention. In this paper, I examine the role of user reviews in product development within the video game industry. Using textual data from a two-sided platform, I assess the impact of user reviews on video game updates. The empirical results show that games receiving more design-related reviews are more likely to be updated in the following month when users express dissatisfaction. Furthermore, I provide evidence that incumbent firms leverage reliability-related reviews for more complex product development, whereas small studios fail to do so. This divergence is driven by differences in both the capability of producers and stakeholder assessment standards.

Key words: Platform, Online user reviews, R&D

Acknowledgments

The author gratefully acknowledges the financial support provided by the PhD student stipend at Georgia Institute of Technology, which enabled the completion of this research. Additionally, the author sincerely appreciates the valuable feedback and insights provided by faculty members at both Copenhagen Business School and Georgia Institute of Technology.

1. Introduction

Online feedback mechanisms have become the most common way for consumers to express their attitudes toward products, which also provides producers with a potentially valuable external information source. Especially on two-sided platforms, online reviews facilitate information exchange within user communities, and they replace the consumer survey as the most common way to understand the consumer population for firms. The formulation of user communities in the platform-hosted environment has important implications for many management activities, including customer retention and product development (Dellarocas 2003). As a result, a considerable amount of academic research has examined how firms respond to online user communities and user reviews (Gans et al. 2021, Zhu and Liu 2018).

While most prior literature focuses on the advertising and pricing behaviors of producers (Archak et al. 2011, Ba and Pavlou 2002, Hollenbeck et al. 2019), the impact of user reviews on firms goes beyond the demand side of the market. From the producers' perspective, user reviews can thus

be viewed as a valuable information source that is at least partially unknown to the producers. User reviews as an external information source provide knowledge about consumer experiences, replacing conventional consumer questionnaires and surveys. Producers are thus motivated to delve into the information pool to learn about buyers' preferences to compete in the market. For example, Bertschek and Kesler (2022) provide indirect evidence that the probability of introducing a product innovation is positively correlated with firms' adoption of online communication via a Facebook page. Karanam et al. (2024) find that apps that echo novel ideas from users enjoy an increase in downloads.

In light of the importance of online user communities and reviews, it is important to understand the impact that reviews would have for product development. Thus far, however, there has been little empirical research in a large scale on this question. The broader literature on product development and user reviews has investigated the immediate proximity between user reviews and product features (Zhang et al. 2018) yet lacks large scale empirical evidence of the extent to which kinds of reviews on producers are affected.

To investigate the impact of reviews in user communities on product development and subsequent performance, I examine the relationship between online user reviews and product development in the video game industry. Platforms in the video game industry is not an advertising channel, and there is ample evidence that game producers value consumer opinions due to its informative characteristics (Xu and Ni 2022). Using a unique dataset from the largest online video game platform, *Steam*, I leverage natural language processing techinique to understand the heterogenous impacts of reviews on product development. I show that the effect of user reviews on product development depends on the types of reviews: design-related ones are more likely to trigger further development when the producers are faced with negative ratings. At the same time, producers also tend to strategically integrate the ideas from the user reviews to meet their user's expectation, which drives large firms but not small studios to learn from reliability-related reviews.

This paper speaks to multiple streams of research. First, it contributes to the literature of online reviews. Miller et al. (2009) emphasize the importance of firms' online strategy as the diffusion of opinions becomes rapid. As a result of its popularity, scholars have extensively examined the impact of online user reviews and their characteristics on consumers and producers. I join the discussion through the lens of the resource-based view. I show that the effect of online reviews is not exogenous but dependent on firm-specific characteristics in the first place, including size, experience, and age. These firm-specific resources and capabilities are correlated with how firms strategically learn from the reviews. Second, it contributes to the discussion of user communities and ecosystem research. User communities have become a common search source for external knowledge, and scholars are asking how the environment shapes the openness of the communities (Garriga et al. 2013). The

ecosystem creates value by bringing in indirect network effect and knowledge exchange, I provide new insights for platform owners to motivate complementors' innovative efforts and hence align with the platform's value creation and evolution strategy (Panico and Cennamo 2022). Last but not least, this paper contributes to the ongoing discussion on the asymmetries between startups and established firms in terms of innovation activities (e.g. Criscuolo et al. 2012, Gimenez-Fernandez et al. 2020). I extend this literature by demonstrating that large firms and small startups diverge in their innovation strategies due to differing stakeholder expectations.

2. Background: User communities and asymmetries between established firms and studios

2.1. User Communities

Shah and Nagle (2019) define user communities as "organizations composed primarily of users working collaboratively, voluntarily, and with minimal oversight to freely and openly develop and exchange knowledge in an area of common interest around an artifact." The value of user communities has been widely discussed, and how firms interact with the communities varies. While it is common for firms to organize communities to support knowledge exchange among users, they can also simply participate in a user community that users or third parties govern. Von Hippel (1976) points out that innovation activity does not necessarily originate from inside the firm, and many successful innovations of scientific instruments were created and tested by users. While most studies focus on the impact of lead users on firm-hosted platforms, two critical differences from firm-hosted platforms are that ordinary users are less technical, and firms do not have control over the discussions.

First, in a corporate-hosted user community designed for innovation, the participants differ significantly from the normal users in the discussion groups on a third-party platform with respect to motivations and capabilities. Firms host communities to invite users who draw from a set of knowledge that differs from firms (Von Hippel 2005, Lettl et al. 2009), and firms use this information to complement their own knowledge and boost corporate innovation (Baldwin and Von Hippel 2011, Bayus 2013, Chatterji and Fabrizio 2014). They are willing to share their innovations and respond to firm recognition (Jeppesen and Frederiksen 2006, Jeppesen and Laursen 2009, Schweisfurth and Raasch 2015). However, a large proportion of users do not share these characteristics: they narrowly focus on their own needs but not the firms (Chatterji and Fabrizio 2012) and mostly speak to impact other customers (Chevalier et al. 2018).

Second, firms have less control over the platform-hosted user communities. In a firm-hosted user community, firms have governance strategies to maintain the integrity of the master rules (Nickerson et al. 2017), which helps them diagnose the potential of the ideas (Sivam et al. 2019). However, in a user community hosted by third-party platforms, firms do not have control over the

knowledge exchange, indicating that discussions might be distant from the knowledge firms seek. The lower control level in platform-hosted user communities suggests that the knowledge is less organized and cannot be as effective as those in firm-hosted user communities where firms can moderate the discussion to help corporate innovation.

Differences in user knowledge present both opportunities and challenges for producers. Common benefits include learning from discussions to enhance branding and product support and even facilitate talent recruitment (Shah and Nagle 2019). Future customers can leverage the wealth of available information when evaluating a product and making purchasing decisions, while producers can, in principle, listen to consumers and respond strategically (Chen and Xie 2005). For example, hotels may reduce advertising expenditures when they receive highly positive evaluations on an online platform, recognizing that satisfied users already generate sufficient organic promotion through their reviews (Hollenbeck et al. 2019). User knowledge exchange is also valuable for innovation and product development. Within user communities, discussions often center around unmet needs (Shah and Tripsas 2007), providing firms with critical learning opportunities to refine and advance their products (Antorini et al. 2012, Ogawa and Piller 2006). As a result, a larger volume of reviews can better assist producers in understanding overall user preferences when making product design improvements.

However, limited control over user communities means that the alignment between user discussions and producers' strategic focus may be weak. While firms may strategically engage with dissatisfied consumers on review platforms (Cheng et al. 2024), motivating further user contributions, this does not guarantee that the additional information will be relevant to the firm's strategic concerns. For example, firms pursuing a differentiation strategy may largely ignore reviews focused on pricing concerns, as they assume their target consumers are less price-sensitive, making such feedback less critical to their core business strategy. Also, innovative ideas generated by users do not necessarily lead to high performance and may underperform compared to imitative ideas when implemented inappropriately (Karanam et al. 2024).Especially for resource-constrained producers, the resources available for implementation are limited, making it impossible to act on everything mentioned in user reviews. Therefore, producers must screen user reviews carefully and evaluate them based on their strategic alignment, market conditions, and capabilities before implementation.

2.2. Dimensional Informativeness

Platforms that manage online reviews offer various approaches to sorting reviews. In addition to sorting by recency and helpfulness votes—long-standing methods—a multidimensional rating system makes specific product attributes more accessible (Schneider et al. 2021). This system explicitly scores different facets of a review, based on the premise that various review dimensions

may have distinct implications for both users and producers (Chen et al. 2018). Moreover, specific attribute recalls may be more accessible within a multidimensional system (Schneider et al. 2021).

Dimensional information can have varied subsequent impacts. For instance, in the airline industry, service-specific reviews—rather than value-specific sentiment—drive and predict user recommendations (Siering et al. 2018). Similarly, in the context of interior design for automobile models, reviews associated with hedonic attributes primarily influence sales (Kong et al. 2020). The underlying rationale is that dimensional information plays a crucial role in a differentiated competitive market: producers catering to spendthrift consumers do not prioritize dimensions beyond price, whereas producers aiming to deliver high-quality products must focus on the corresponding dimension. Consequently, dimensional information is considered highly important, as it may have varied subsequent impacts. Even when a multidimensional rating system is unavailable, producers on third-party platforms may still be motivated to sort reviews in a similar manner and analyze them independently. Producers focus on the dimensions most relevant to their business, scrutinizing feasible and potential ideas from reviews to align user preferences with their development strategies—whether by offering niche content or competing through superior quality.

2.3. Producer capabilities and orientations

The classic resource-based view and dynamic capabilities perspective argue that firms possess resources and capabilities that enable them to achieve competitive advantage and continuously renew them to remain competitive in a changing business environment (Teece et al. 1997). One such resource is IT capability, which allows firms to detect potential environmental changes and varies across organizations in its effectiveness (Aral and Weill 2007, Zhu and Kraemer 2002). For example, firms with strong data analysis capabilities can outperform competitors by expanding the search space of existing knowledge during the incremental R&D process (Wu et al. 2020). This advantage arises because IT competency enables firms to learn from external actors, such as partners and consumers (Wu et al. 2020, Kleis et al. 2012).

In the digital product industry, there are at least two types of organizations: large firms such as Electronic Arts and small studios. They significantly differ in the efficiency of using user reviews. Firm-specific capability is accumulated through experience and stored in human capital. Large firms have many more resources and capable employees with the ability to respond to user demands, and they are experienced in translating user needs into products. The resources and process when acquiring knowledge from user communities are specific to firms, which impacts the success of organizational learning (Sivam et al. 2019). On the other hand, small studios that develop apps are usually composed of around 20 employees with limited resources in R&D. A significant proportion of the small studios funded through Kickstarter lack the capability to hire talent, indicating less

stored human capital in absorbing information. Compared to incumbents, they are less experienced in multidimensional requirements from users, and they narrowly focus on their products.

Despite the importance of IT capability heterogeneity in industries such as biotech and banking, firms in IT industries generally excel at leveraging IT capabilities to learn from external resources. Therefore, beyond firm capabilities, I argue that another factor shapes the asymmetries between incumbents and small startups: the evaluation standards used by consumers. Customers, as important stakeholders, develop expectations based on the organization's history and positioning. For example, consumers tend to be more critical when shopping for utilitarian products than for hedonic products (Lee and Hosanagar 2021), as utilitarian products must precisely meet user demands. These asymmetries between producers suggest that users evaluate products relative to their expectations. When comparing startups and incumbents, startups generally benefit from a legitimacy buffer from stakeholders, allowing for greater performance volatility (Fisher et al. 2016). In contrast, incumbents face stronger conformity pressures to maintain a stable product offering (Taeuscher and Rothe 2021). Structural differences in auction success further illustrate this asymmetry; for instance, offering shipping insurance is critical for experienced sellers but not for new sellers (Goes et al. 2013), as consumers expect experienced sellers to provide this service and penalize them when it is missing. Given the differences in the evaluation criteria used by consumers, firms will learn differently from their knowledge base to meet consumer demands. Therefore, heterogeneous producers must understand their users' pre-existing evaluation standards and align their offerings accordingly.

3. Data and Methods

3.1. Empirical Setting and Data

The video game industry has been used for studies in online user reviews and product development, given its unique features for two reasons. First, similar to the software industry, the published game can be further modified and developed as a strategic move. The update history of the games enables one to examine the trajectory of the product development. Second, users of video games are encouraged to leave reviews on the games, and the cost of leaving reviews is much lower, indicating the reviews are more representative of the whole population compared to the other industries. A project manager from *Blizzard*, a famous game developer and publisher that has released Diablo and the Warcraft series, claims in my interview that the video game industry is heavily community-based, and feedback from existing users is vital in the way that it helps attract potential players. He says:

> We care a lot about ratings and reviews. We listen to our existing users' feedback because we believe that the existing users are good representatives of the whole market that we are serving,

> and basically, we are creating value by meeting their demands. And as a consequence, we tend to focus more on negative feedback since we try our best to eliminate them. As we do try to integrate the users' ideas into our products, it is hard to claim to what extent our development can be attributed to user's feedback, since the development process is complicated, and it is challenging for us to identify valuable feedback. We receive thousands of reviews from different sources, and unless we can find a pattern, most of them cannot be embraced directly since the reviews per se do not contain valuable information

Steam, a digital distribution platform for video games, was launched in September 2003. It is now the most prominent digital distribution platform for PC gaming. By 2019, it had over one billion registered accounts, with 90 million active users per month. It had over 30,000 games by 2019, with an additional 20,000 downloadable content extensions to games (DLCs). In this study, I collect review information from a sample of games officially published in 2016 and 2017. I select all the games that exceed 1 million sales until the end of 2021 and remove the games that are not listed on the platform for various reasons. The sample criteria guarantee that the products are successful, indicating that they are not terminated quickly due to reasons such as lack of revenue. The sample includes 263 unique games, and I collect user review text from three years prior to the official publishment to four years post publishment. I also match the reviews with the sales and active user data from *Steam Spy*, which monitors the data using *Steam* API (Rietveld and Ploog 2022).

3.2. Dependent Variable

The main dependent variable is $Update_{it}$. The dummy variable equals 1 if game i publishes an Update News in month t.

3.3. Independent Variables

The main challenge is to differentiate the dimensions embedded in the review text since *Steam* does not adopt the multidimensional review system. Unsupervised machine learning algorithms are commonly used in review analysis for extracting topics from text automatically (Timoshenko and Hauser 2019). I make use of a supervised machine-learning algorithm to recognize the dimensions of the reviews.

BERT is a technique for Natural Language Processing (NLP) pre-training developed by computer scientists. To tailor the algorithm to my data, I manually label a subsample of reviews in multiple dimensions and then train the model in order to identify the information embedded in review texts in terms of these dimensions. As Steam does not provide a multidimensional rating system, I classify the reviews into four dimensions: reliability (bugs, glitches, and optimizations), aesthetics (graphics and sounds of the game), story (storyline and characters), and gameplay (game features

and game mode). I randomly select 800 positive reviews and 800 negative reviews, and manually label them as 1 or 0 in terms of these four dimensions. Table 1 illustrates the examples of the concordance between review text and classifications. I manually recognized reviews for both positive and negative reviews regarding these dimensions.

Table 1 Examples of Mannually Labeled Reviews

Review Text	Reliability	Aesthetics	Story	Gameplay
...was a game that I played during my mid-childhood. I was hoping that they would fix the game issues (Physics, Camera, and Mouth syncing) before releasing it to steam, but I guess it's free money...it's still a fun game. Still a fun storyline.	1	0	1	0
It's incredibly satisfying killing the orcs, making small mazes, messing around with the various traps and weapons...I love how each play-through can be different and custom to how YOU want to play...Highly recommend. :)	0	0	0	1
A surprisingly entertaining game with clever character design, Funny voice acting, great art and amazing music at all points. Although it does come with some minor problems, none of them are bad enough to hinder the great experience the game provides	0	1	1	0
...having to play the same 3 to 5 minute but what was terrible was the number of times the game just crashed in a row (up to 8 tonight in the last half hour), which is why I can't recommend the mac version of this game.	1	0	0	0

After training the NLP model with the manually labeled reviews, I test its prediction accuracy on another 200 positive reviews and 200 negative reviews. These 400 reviews are also manually labeled in the four dimensions based on the contents to verify the reliability of the machine prediction. The accuracies of the tuned model for these dimensions are shown below in Table 2. This result is comparable to previous research where the accuracy of sentiment analysis reaches 94% by deep learning, 74.2% by Naïve Bayes, and 76.46% by Support Vector Machine (Day and Lin 2017), and 70% to 74% where the deep learning model is used to identify the informative of the consumers' review (Timoshenko and Hauser 2019). As the story, aesthetics, and gameplay cannot be modified easily through updates, I combine them together as a *Design* dimension.

Table 2 NLP Model Performance

	Reliability	Design		
		Aesthetics	Story	Gameplay
Accuracy	0.865	0.932	0.910	0.816

I apply this fine-tuned machine learning model to the reviews of the game in the sample. For example, the game *Stardew Valley* received 172 reviews in January 2017, with 10 in reliability, 23 in aesthetics, 7 in story, and 102 in gameplay. For each dimension, I calculate the score. Among the 10 reliability-oriented reviews, 2 are positive and 8 are negative, which yields the $ReiliabilityScore_{it}$ = 0.2 and $Reiliability_{it} = 10$. Similarly, I calculate all the *Design Score* and *Design Count* for the selected game samples.

The detailed description of variables is presented in Table 3. The descriptive statistics are in Table 4, and the correlations are in Table 5. In Figure 1, I provide a locally-weighted scatterplot of data to show the dynamic patterns of games. Panel A shows that the likelihood of product development increases first and then decreases with time. Panel B shows that the relationship between the number of users and time is an inverse-U shape, and it is not perfectly correlated with the number of reviews. Panel C shows that reliability is less commonly mentioned than game design in the reviews, while Panel D shows that the scores of the two dimensions are not parallel: the reliability score is much lower; the game design score corresponds with the overall score, and both of them increase first and stay stable after official publish.

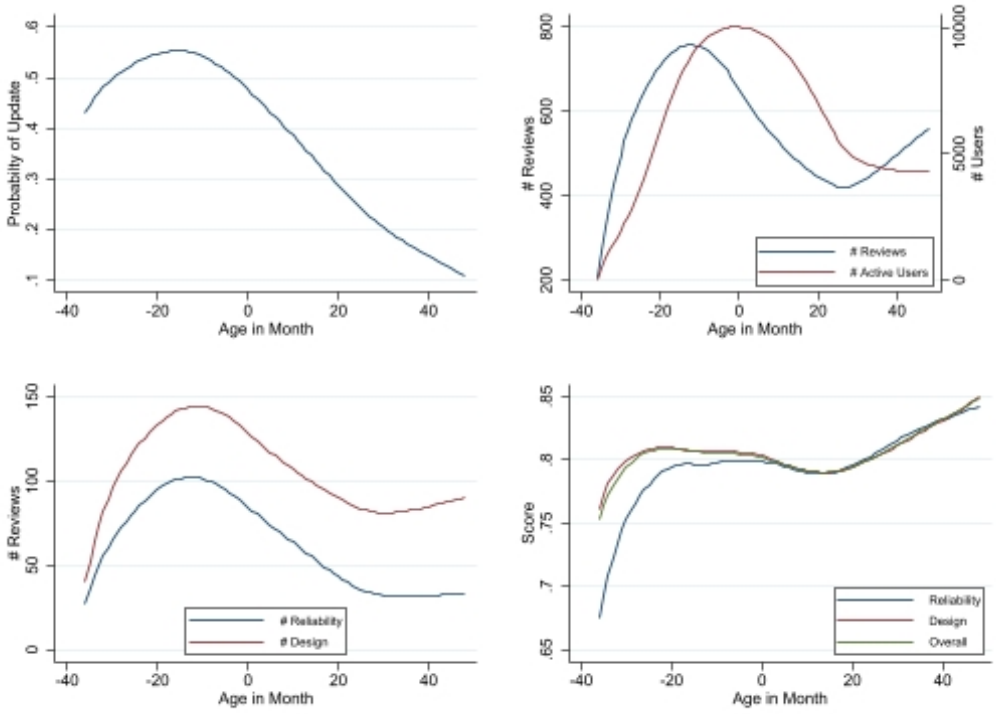


Figure 1 Game and Review Trend

Table 3 Construction of Variables

Construct	Variable	Description
Dependent Variable	$Update_{it}$	1 if game i published an update on month t, 0 otherwise.
Independent Variables	$Reliablity_{it}$	Count of reviews mentioning reliability posted for game i on month t.
	$Design_{it}$	Count of reviews mentioning design posted for game i on month t.
	$ReliablityScore_{it}$	The percentage of positive out of all the reviews mentioning reliability posted for game i on month t.
	$DesignScore_{it}$	The percentage of positive out of all the reviews mentioning design posted for game i on month t.
	$Firm_i$	1 if game i is published by a firm with more than 50 employees, 0 otherwise.
	$ProducerAge_{it}$	The age of the developer of game i on month t, in year.
	$Experience_{it}$	Number of games that the developer of game i published before the focal game.
	$Collaboration_i$	1 if game i is collaborated by multiple game developers, 0 otherwise.
Control Variables	$Log(Count)_{it}$	*Log* of 1 plus all review count posted for game i on month t.
	$Log(Users)_{it}$	*Log* of 1 plus average daily active users for game i on month t.
	$Log(Owners)_{it}$	*Log* of 1 plus accumulating owners (sales) for game i on month t.
	Age_{it}	The age of the game i on month t, in months.
	$Positive_{it}$	The average attitude for reviews posted for game i on month t.
	$Promotion_{it}$	1 if game$i$ has a non-updated related promotion on month t, 0 otherwise.

3.4. Specifications

The baseline specification is

$$\begin{aligned} Y_{it+1} =& f(\alpha_0 + \beta_0 \times X_{it} + \beta_1 \times log(Reliability_{it} + 1) + \beta_2 \times ReliabilityScore_{it} \\ &+ \beta_3 \times log(Design_{it} + 1) + \beta_4 \times DesignScore_{it} + \delta_t + \gamma_i + \epsilon_{it}) \end{aligned}$$

where Y_{ist} is the dummy variable for updates for game i on month t. X_{it} are the time-variant market controls including user count, overall score, age, and promotion information. I include game-fixed effects and time-fixed effects, which account for many omitted variables and sources of endogeneity.

4. Results

4.1. Baseline Results

The baseline results of the linear probability models are presented in Table 6. Column (1) only includes the controls, and it is evident that active users have a positive impact on the probability of

Table 4 Descriptive Statistics

	Mean	SD	Min	Max	N
Update	0.241	0.428	0	1	13,305
Log(Reliablity)	2.559	1.469	0	9.558	13,305
Log(Design)	3.507	1.554	0	9.187	13,305
ReliablityScore	0.818	0.213	0	1	13,305
DesignScore	0.817	0.19	0	1	13,305
Firm	0.387	0.487	0	1	13,305
Producer Age	12.959	10.181	0	53	13,305
Experience	6.957	17.283	0	90	13,305
Collaboration	0.228	0.419	0	1	13,305
Log(Count)	5.062	1.598	0.693	10.542	13,305
Log(Users)	6.531	2.002	0.304	14.859	13,305
Log(Owners)	13.897	1.319	6.909	18.311	13,305
Age	27.986	19.024	-24	60	13,305
Positive	0.817	0.158	0	1	13,305
Promotion	0.179	0.383	0	1	13,305

Table 5 Correlation Matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*Update*	1														
2	*Log(Reliability)*	0.203	1													
3	*Log(Design)*	0.137	0.793	1												
4	*ReliablityScore*	-0.11	0.062	0.181	1											
5	*DesignScore*	-0.129	0.076	0.212	0.601	1										
6	*Firm*	-0.031	0.241	0.236	-0.102	-0.106	1									
7	*Producer Age*	-0.122	0.134	0.181	-0.021	-0.014	0.678	1								
8	*Experience*	-0.099	0.057	0.036	-0.049	-0.043	0.435	0.45	1							
9	*Collaboration*	-0.038	0.014	-0.034	-0.078	-0.086	0.227	0.319	0.303	1						
10	*Log(Count)*	0.16	0.82	0.909	0.18	0.21	0.188	0.137	0.031	-0.05	1					
11	*Log(Users)*	0.259	0.719	0.713	0.02	0.029	0.332	0.254	0.048	0.029	0.762	1				
12	*Log(Owners)*	-0.029	0.377	0.346	-0.02	-0.037	0.122	0.138	-0.051	-0.003	0.383	0.479	1			
13	*Age*	-0.336	-0.11	-0.057	0.117	0.117	0.021	0.179	0.031	0.027	-0.065	-0.079	0.412	1		
14	*Positive*	-0.157	0.089	0.251	0.733	0.84	-0.13	-0.019	-0.058	-0.098	0.249	0.032	-0.037	0.141	1	
15	*Promotion*	-0.263	-0.017	0.025	0.069	0.084	-0.031	-0.009	-0.009	-0.023	0.033	-0.035	-0.005	0.075	0.098	1

updates but the accumulated sale record has a negative impact, which indicates that active users and inactive users impact product development differently. From column (2), which controls for monthly active users and accumulated copies sold, it is evident that the aggregate count of user reviews has a positive impact on the likelihood of a product update in the following month, but the coefficient magnitude of active users shrinks. The coefficient of log-transformed review count indicates that a ten percent increase in the number of posted reviews is expected to lead to a 0.315 percent increase in the update likelihood. Column (3) substitutes the total review counts with the counts of reliability and design reviews, and the magnitude of the coefficient on reliability

review counts is statistically insignificant at conventional levels, while the coefficient on the count of design reviews is only significant at 10% level: if the count of monthly design reviews increases by ten percent, the producer is expected to increase the update likelihood by 0.152 percent.

Table 6 Baseline Regression: Impact of User Reviews on Product Development

	(1) *Update*	(2) *Update*	(3) *Update*	(4) *Update*	(5) *Update*	(6) *Update*
Age	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Promotion	0.093***	0.094***	0.094***	0.093***	0.094***	0.093***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
log (Owners)	-0.042***	-0.039***	-0.041***	-0.042***	-0.041***	-0.042***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
log (Users)	0.052***	0.028**	0.036***	0.034***	0.037***	0.034***
	(0.009)	(0.012)	(0.011)	(0.012)	(0.011)	(0.011)
Positive	-0.150***	-0.164***	-0.159***	0.078		
	(0.044)	(0.045)	(0.044)	(0.048)		
log (Count)		0.033***				
		(0.009)				
log (Reliability)			0.007	0.014	0.007	0.025
			(0.008)	(0.033)	(0.008)	(0.018)
log (Design)			0.016*	0.099***	0.015	0.083***
			(0.009)	(0.032)	(0.009)	(0.019)
Reliability Score					(0.019)	(0.02)
					(0.016)	(0.028)
Design Score					-0.075***	0.073**
					(0.022)	(0.031)
*log (Reliability) * Positive*				-0.009		
				(0.04)		
*log (Design) * Positive*				-0.098***		
				(0.037)		
log (Reliability)						-0.022
** Reliability Score*						(0.019)
log (Design)						-0.078***
** Design Score*						(0.02)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Game FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13230	13230	13230	13230	13230	13230

Standard errors in parentheses, clustered at game level
* p<0.10, ** p<0.05, *** p<0.01

It is also interesting that the ratings of the reviews matter in columns (1) to (4). This is consistent that less satisfactory products should be invested in more attention for further polish. I further investigate the different dimensions of ratings generated by machine learning algorithms to investigate whether they interact with the count of the reviews to motivate the producers to

learn. In column (4), I examine whether the sentiment *Positive* interacts with the two types of reviews in terms of impacts on product development. Interaction between the variable *Positive*, which captures the overall satisfaction level of the users in the focal month, and the count of reliability reviews is not significant, but the interaction with design review counts is significant and negative. In other words, producers are more likely to react to the design reviews when users express negative feelings through reviews.

In column (5), I dissemble the overall ratings *Positive* into *Reliability Score* and *Design Score* to address the concern that the overall score *Postive* is dominated by the evaluation of design and hence feedbacks on reliability defect cannot be reflected when the design is positively evaluated. I also interacted dimension ratings with review counts respectively in column (6), but the interaction term is only significant for design reviews, not reliability reviews. In Figure 2, for each game, I divide the observations into 5 quantile bins regarding the overall ratings and examine the coefficient of two types of reviews, respectively. It is evident that design reviews can significantly trigger the update when the ratings are low.

While it is surprising that firms do not react to reliability reviews, it addresses the endogeneity concern of simultaneity; that is, the producers and users identify certain defects simultaneously, in which case the coefficient on reliability reviews should be significant. In fact, video game tests by producers generally can only check the current feature but not how it interacts with others, such as personal hardware configurations. Therefore, producers hardly identify bugs that have not been reported in the first place since "bugs might be caused by a specific interaction of game states and properties that may not be evident to whoever experiences the bug" (Truelove et al. 2021). For the same reason, tackling the reliability reviews is challenging since sometimes bugs arise out of specific hardware setups and depend on more than one huge loosely connected system having a certain state(Murphy-Hill et al. 2014).

Given the average likelihood of updates is 0.241, producers do not publish updates monthly. This means that updates might be a response to the reviews from the two or three months before, in addition to the previous month. Therefore, in Table 7, I use different lags between the independent variable and the dependent variable to address this concern. It is evident here that using a one-month lag is appropriate since it yields the highest R-squared value, and using a two-month lag also yields similar results regarding our main variables of interest. Using three months or four months is not appropriate here since even the impact of the count of users and overall ratings would not be reflected. In Table A1 in the Appendix, I also replicate the main regression models using a 2-month lag, and despite the difference in magnitude, the main results persist. In Table A2 in the Appendix, I add a 2-month lag to the main regression models. Both AIC and BIC decrease, but

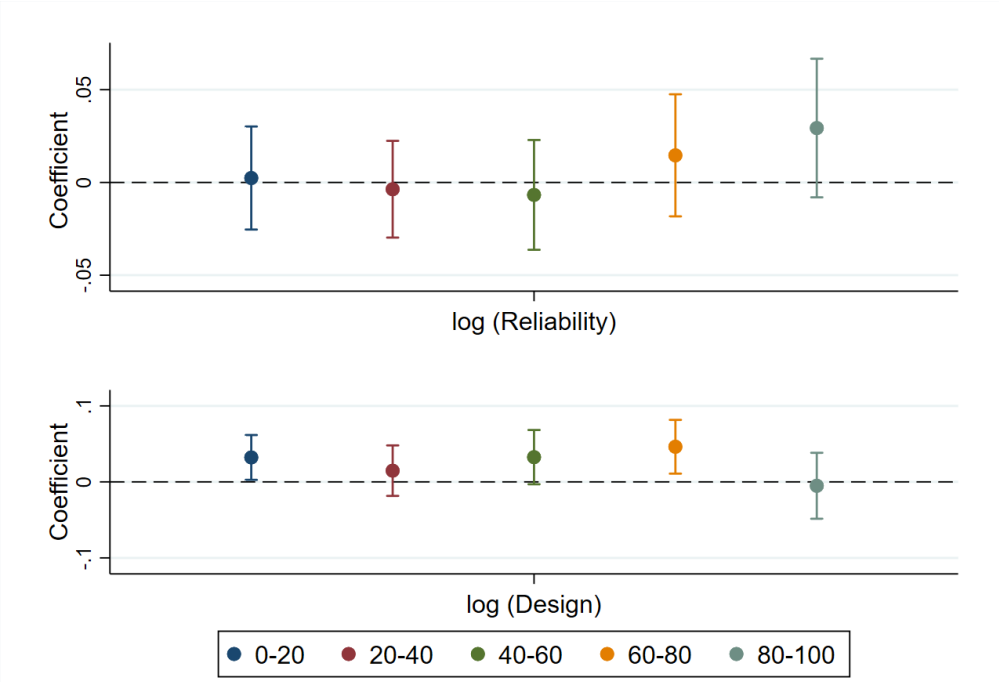


Figure 2 **Coefficients of Review Counts by Ratings**

Table 7 **Optimal Lag Selection**

	(1) *Update*	(2) *Update*	(3) *Update*	(4) *Update*
Age	-0.011***	-0.006***	-0.005***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Promotion	0.094***	-0.053***	-0.059***	-0.047***
	(0.016)	(0.010)	(0.008)	(0.010)
log (Owner)	-0.039***	-0.032***	-0.023***	-0.020**
	(0.009)	(0.008)	(0.008)	(0.008)
log (Users)	0.028**	0.023**	0.011	0.001
	(0.012)	(0.011)	(0.011)	(0.011)
log (Count)	0.033***	0.021**	0.020**	0.028***
	(0.009)	(0.009)	(0.009)	(0.009)
Ratings	-0.164***	-0.149***	-0.058	-0.056
	(0.045)	(0.043)	(0.041)	(0.044)
Time &Game FE	Yes	Yes	Yes	Yes
Lag	1 Month	2 Months	3 Months	4 Months
R^2	0.138	0.122	0.115	0.111

Standard errors in parentheses, clustered at game level
* p¡0.10, ** p¡0.05, *** p¡0.01

the magnitude of the coefficient only changes slightly. The main result that the count of reviews has a positive impact on subsequent update holds.

The result is in line with the marketing studies where ratings have a significant impact on other consumers' behavior (Chevalier et al. 2018, Moe and Trusov 2011). I argue that producers also make decisions about product development when they find consumers are unsatisfied. What is unknown to them is the preference of the customers in terms of the game design, so they pay more attention to the volume of the reviews and the textual information in the reviews. Therefore, for producers, the importance of reviews is the information embedded in the high volume of the design reviews, which represents the consumers' taste when ratings fail to meet their expectations. More reviews can provide information about the direction of the product development when they attempt to cater to frustrated users. Overall, the results show that different dimensions of reviews have heterogeneous impacts on product development activities. The producers do not respond to reliability-related reviews, whereas design-related reviews are relevant for product improvement when product ratings are low.

4.2. Heterogeneity across producers and mechanisms

I further investigate the heterogeneity of producers in Table 8. Column (1) replicates the baseline regression in Table 6. In Column (2), I interact the counts with the dummy *Firm*, which equals 1 when the game is produced by large firms and equals 0 when produced by an indie game studio or individual developers.The coefficient on the interaction term with reliability review counts is significant and positive at the 5% level, suggesting that individuals and studios are less likely to take advantage of the reliability reviews for subsequent product development compared to firms. Figure 3 shows the coefficients for individuals, studios, and firms, respectively. Individual developers do not learn from both types of reviews, and firms react differently from studios. Like the whole sample population, studios respond to design reviews, while firms respond to reliability reviews.

In column (3), I interact with the count of two-dimension reviews with producer age respectively and find that the interaction with reliability reviews count is significantly positive. This indicates young firms underperform in meeting users' reliability-related demand, and the reason is, again, modification surrounding game reliability is challenging and costly. Similarly, in column (4), I interact with the experience of the producers, and in column (5), I interact with the collaboration indicator. Producers learn from the user reviews regarding reliability as they have more project experience and when they are collaborating with others. By accumulating domain experience and partnering with other developers, producers, including small studios, can learn about reliability from user reviews in addition to the design that they generally focus on. Together, these results show that large firms can effectively address reliability-related reviews regardless of ratings, whereas both types of firms are equally capable of addressing design-related reviews, particularly when products receive negative ratings.

Table 8 OLS - Heterogeneity of Producers

	(1) *Update*	(2) *Update*	(3) *Update*	(4) *Update*	(5) *Update*
log (Reliability)	0.007	-0.008	-0.011	-0.001	-0.008
	(0.008)	(0.011)	(0.013)	(0.009)	(0.010)
log (Design)	0.016*	0.014	0.022	0.019*	0.027**
	(0.009)	(0.012)	(0.014)	(0.010)	(0.011)
Firm log (Reliability)*		0.035**			
		(0.016)			
*Firm * log (Design)*		0.008			
		(0.018)			
*Producer Age * log (Reliability)*			0.001**		
			(0.001)		
*Producer Age * log (Design)*			-0.001		
			(0.001)		
Experience log (Reliablity)*				0.001***	
				(0.000)	
Experience log (Design)*				-0.000	
				(0.000)	
Collaboration log (Reliability)*					0.061***
					(0.018)
Collaboration log (Design)*					-0.045**
					(0.019)
Time FE	Yes	Yes	Yes	Yes	Yes
Game FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	13230	13230	13230	13230	13230

Standard errors in parentheses, clustered at game level
* p<0.10, ** p<0.05, *** p<0.01

4.3. Mechanism

I first investigate the underlying mechanism by examining the details of product updates. However, since not all producers provide descriptions of their updates, I focus on a smaller subsample. Additionally, as producers typically release updates that include both design-related and reliability-related features in a single update, we do not directly link updates to user reviews here. This is because each update simultaneously addresses both design and reliability issues. Instead, our focus is on firms' capabilities. In Table 9, I compare firms and studios in terms of update content using a t-test. To measure the focus of updates, I use a dictionary-based approach that categorizes updates into design-related and reliability-related issues. Specifically, I calculate the percentage of update descriptions that address each type of issue. Since many keywords could apply to both categories, I use a strict dictionary where each term is assigned exclusively to one category, ensuring that the sum of the percentages remains below 100%. The results in Table 9 clearly show that firms produce a higher proportion of reliability-related updates, indicating stronger capabilities in this area. Additionally, I provide figures in Appendix (Fig A1 and Fig A2) illustrating the trend

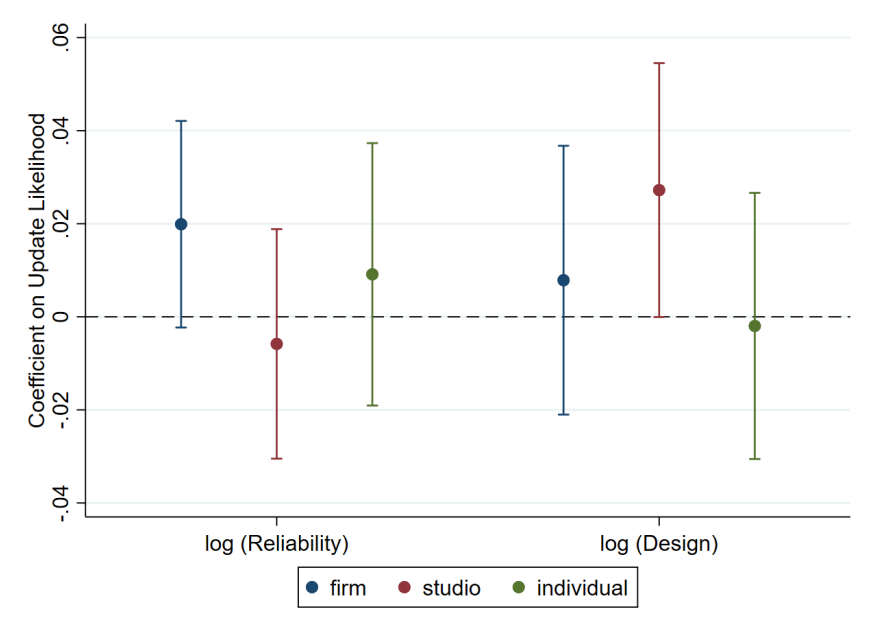


Figure 3 Coefficients of Review Counts by Producer Type

of these two update types since the game's launch (without controls). The figure demonstrates that the superior performance of large firms is not merely a result of early-stage testing, where producers may identify and fix bugs due to initial quality assessments. Instead, the pattern suggests a continuous learning process, in which firms optimize their products over time based on user feedback. Moreover, I examined the content length and update intervals between the two types of producers. Firms and non-firms do not differ in update frequency, and non-firms even produce more content in their updates. Therefore, the capability to address reliability issues is not merely a consequence of publishing updates in general but rather a reflection of firms' specific capability in handling reliability-related issues.

Table 9 T-test: Difference in Capability Across Types of Producers

	Firm=1	Firm=0	T-Statistic
# Reliablity Items/# All Items	0.110	0.012	-2.474**
# Design Items/# All Items	0.124	0.085	-1.215
Items in Updates	223.12	267.45	5.901***
Length in Descriptions (Word Count)	14.17	17.77	7.096***
Intervals between Two Consecutive Updates (in Days)	44.785	47.549	0.823

* p<0.10, ** p<0.05, *** p<0.01

Another concern is that producers do not necessarily develop updates ex post based on incoming user feedback but instead preemptively create a stock of update content and strategically time its release based on observed user engagement. Specifically, when a surge in user reviews occurs, firms may deploy an update to increase loyalty and signal responsiveness. To address this concern, I categorize reviews as useful or non-useful based on two commonly used criteria. The underlying

assumption is that if the effect is driven by a preemptively developed update stock, producers should be indifferent to the distinction between useful and non-useful reviews.

In Table 10, I examine whether the usefulness of reviews influences their impact on update decisions. I employ two widely used approaches to assess review helpfulness: length (Columns (1) to (3)) and user votes (Columns (4) to (6)). Reviews are classified as helpful if they either exceed 20 words in length or have been voted "helpful" by at least one other user. The word-count approach is frequently used in the marketing and information management literature, as prior research has shown that longer reviews are more likely to be perceived as informative and helpful (Mudambi and Schuff 2010, Kim et al. 2006, Salehan and Kim 2016). The second approach is employed by the Steam platform, where both users and producers can sort reviews based on the number of user endorsements.

In Columns (1) and (4), I regress the likelihood of updates on the two types of reviews, distinguishing between useful and non-useful reviews using the aforementioned criteria. The coefficient for the two types of reviews increases from 0.07 in Table 6 to 0.023–0.025 for reliability-related reviews and from 0.016 to 0.040–0.043 for design-related reviews. This indicates that the volume of reviews influences update decisions by providing valuable information.

In the remaining columns, I examine the effect separately for large firms (Columns (2) and (5)) and small studios (Columns (3) and (6)). Table 8 shows that, in general, large firms are more responsive to reliability-related reviews, and this effect persists even when focusing exclusively on useful reviews. The coefficient of useful reliability-related reviews is higher in Column (2) than in Column (3), suggesting a stronger impact for large producers.

Moreover, I identify a significant negative effect of non-useful design reviews on the likelihood of updates. This finding reinforces my argument regarding the learning mechanism by demonstrating that information overload and disorganized discussions dilute valuable insights, making it more difficult for producers to extract meaningful signals. This outcome would not occur if reviews influenced updates through alternative mechanisms. The effect is particularly pronounced for design-related reviews in the case of large firms.

I further investigate why firms are more likely to respond to reliability-related reviews. One explanation is that users are less tolerant of reliability issues when the product is from a large incumbent compared to a product from a small studio. In other words, users may expect large firms to provide more reliable products while holding smaller startups to a lower standard. To verify this mechanism, I adopt a disaggregated approach by analyzing individual reviews to understand how various factors influence user attitudes.

In Table 12, I regress different review features on the overall attitude expressed in the focal reviews. Specifically, in Columns (1) and (2), I examine whether mentioning reliability, design, and

Table 10 OLS - Usefulness of Reviews

	(1) Update	(2) Update	(3) Update	(4) Update	(5) Update	(6) Update
Promotion	0.095***	0.032	0.133***	0.095***	0.034*	0.132***
	(0.016)	(0.020)	(0.022)	(0.016)	(0.020)	(0.022)
Age	-0.010***	-0.012***	-0.009***	-0.010***	-0.012***	-0.009***
	(0.002)	(0.003)	(0.002)	(0.001)	(0.003)	(0.002)
Ratings	-0.115***	-0.177**	-0.085*	-0.088**	-0.116	-0.072
	(0.043)	(0.077)	(0.049)	(0.043)	(0.080)	(0.049)
log (Owner)	-0.038***	-0.044**	-0.034***	-0.037***	-0.042**	-0.033***
	(0.009)	(0.019)	(0.010)	(0.009)	(0.018)	(0.010)
log (Users)	0.035***	0.071***	0.017	0.037***	0.073***	0.020
	(0.011)	(0.020)	(0.014)	(0.011)	(0.020)	(0.013)
log(Useful Reliability)	0.023***	0.025**	0.016*	0.025***	0.031***	0.016*
	(0.007)	(0.011)	(0.008)	(0.008)	(0.011)	(0.009)
log(Useful Design)	0.043***	0.049***	0.043***	0.040***	0.054***	0.035***
	(0.009)	(0.015)	(0.011)	(0.010)	(0.017)	(0.011)
log(Non-Useful Reliability)	-0.012	-0.004	-0.017	-0.014*	-0.009	-0.017
	(0.008)	(0.011)	(0.010)	(0.008)	(0.012)	(0.011)
log(Non-Useful Design)	-0.022**	-0.033***	-0.014	-0.020**	-0.039***	-0.009
	(0.009)	(0.013)	(0.011)	(0.010)	(0.013)	(0.013)
Constant	0.607***	0.371	0.694***	0.576***	0.291	0.682***
	(0.119)	(0.253)	(0.128)	(0.120)	(0.246)	(0.131)
N	13230	5112	8118	13230	5112	8118
r2	0.141	0.171	0.138	0.141	0.173	0.137
Year FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Game FE	Y	Y	Y	Y	Y	Y
Measurement of Informativeness		Length			User Vote	
Producer	All	Firm	Studio	All	Firm	Studio

Standard errors in parentheses, clustered at game level. *p¡0.1 **p¡0.05 ** p¡0.01

price affects users' evaluations. After controlling for the timing of the review (i.e., whether users play the game one month after launch versus ten years later) and the duration of gameplay, I find that mentioning reliability is negatively correlated with ratings, whereas mentioning design is not correlated with ratings. This suggests that users take good design and quality for granted but are far less forgiving of unreliable products.

However, what is particularly interesting is the heterogeneity across producer types. When interacting these review dimensions with firm type in Columns (3) and (4), I find that users hold large firms to a higher standard. For studios and individual developers (*Firm=0*), users are also likely to associate design aspects with positive reviews, encouraging these firms to prioritize design improvements. However, for large firms (*Firm=1*), I find no correlation between mentioning design and positive comments, while reliability concerns have a greater negative impact compared to smaller producers.

These results indicate that user attitudes, as reflected in ratings, are not uniform across products. For small studios, users reward good design and penalize unreliable products. In contrast, for large firms, users expect a reliable product with good design from the outset. As a result, they do not

reward good design but heavily penalize unsatisfactory reliability. Together with the aggregated findings, these results suggest that producers adjust their learning trajectories in response to user evaluation patterns, effectively leveraging user reviews to guide strategic decision-making.

Table 11 Heterogeniety of Different Dimensions on User Attitude

	(1) Positive	(2) Positive	(3) Positive	(4) Positive
Reliabilty	-0.068***	-0.465***	-0.030*	-0.229*
	(0.010)	(0.06)	(0.017)	(0.13)
Design	0.008	0.058	0.0384***	0.340***
	(0.006)	(0.050)	(0.013)	(0.114)
Price	-0.027**	-0.200**	-0.027**	-0.204**
	(0.013)	(0.091)	(0.013)	(0.090)
Time Since Launch	-0.0003	-0.003	0.0003	0.003
	(0.001)	(0.007)	(0.001)	(0.008)
PlayTime	0.000	0.000	0.000	0.000
	0.000	0.000	(0.000)	(0.000)
Reliablilty * Firm			-0.068**	-0.388**
			(0.032)	(0.182)
Design * Firm			-0.065***	-0.535***
			(0.017)	(0.137)
Constant	0.904***	2.569***	0.892***	2.406***
	(0.027)	-0.04	(0.024)	-0.08
Model	OLS	Logit	OLS	Logit
Year FE	Y	Y	Y	Y
R2/LL	0.024	-3.21E+06	0.027	-3.20E+06
N	7937756	7937756	7937756	7937756

Standard errors in parentheses clustered at game level.
* p¡0.10, ** p¡0.05, *** p¡0.01

4.4. Robustness checks

One concern is that the producers may decide to save the ideas for new products instead of publishing updates when faced with complaints about reliability and quality. In fact, there is evidence that the comments on the previous generation of a product have a significant impact on the current generation (Li et al. 2022). Therefore, it is necessary to rule out the alternative possibility that producers may shift focus onto new products instead of updating the game after negative feedback.

I address the concern by investigating the probability of publishing DLC. The majority of the DLCs, an extension of games to which the consumer already has access, require a small payment (Ozuem et al. 2017). DLCs are not aimed at fixing glitches at all. DLCs contain large-scale paid

content, including new game modes, maps, and expansions. DLCs are a cyclical form of commodification, and companies are reliant on DLCs for revenue continuity since they fill the gaps between sequels (Lizardi 2012). Updates barely alter the core game or its storyline and they are installed free and automatically. But DLC has a separate webpage, and producers market themselves independently of the original game. Some games are free-to-play but provide DLC for customers with a certain charge. Therefore, DLC and updates identify different facets of product strategy: updates capture developments of existing products, while DLCs focus on generating revenue by developing and marketing new products. Therefore, the usage of DLC can be viewed as an alternative for game updates, and hence, I use it to run the robustness check.

In Table 12, I run a bivariate probit model in column (3) and compare it with univariate OLS models to address this concern. If producers avoid publishing updates but save ideas to develop new DLC as a response to the online reviews, differences across the coefficients would be observed across models. I only consider the sample of studio producers who take advantage of design reviews to publish updates. Columns (1) and (2) of the table show the result of OLS models for DLC and updates for the selected sample, respectively, and column (3) shows the bivariate probit model. The significance does not change, suggesting that the potential interaction between the decision to publish a DLC and an update does not undermine the results already reported. To facilitate comparison, I note that the average marginal effect of log (Design) on updates is 0.023 percent, compared with the 0.025 obtained in the OLS specification.

Table 12 Robustness Check: DLC vs Updates

	(1)	(2)	(3)	
	Update	*DLC*	*Update*	*DLC*
Model	OLS	OLS	Bivariate Probit	
Age	-0.010***	-0.002*	-0.047***	-0.029*
	(0.002)	(0.001)	(0.010)	(0.017)
Promotion	0.132***	0.019**	0.964***	0.548***
	(0.022)	(0.007)	(0.093)	(0.105)
log (Users)	0.017	0.011**	0.185***	0.236***
	(0.014)	(0.005)	(0.060)	(0.089)
log (Owners)	-0.036***	-0.010**	-0.247***	-0.214**
	(0.010)	(0.005)	(0.059)	(0.089)
Positive	-0.117**	-0.034	-0.640**	-0.683
	(0.051)	(0.023)	(0.277)	(0.418)
log (Reliability)	-0.004	-0.003	-0.013	-0.036
	(0.011)	(0.004)	(0.053)	(0.058)
log (Design)	0.025**	-0.001	0.126**	-0.010
	(0.011)	(0.005)	(0.060)	(0.072)

Standard errors in parentheses clustered at game level
*p<0.10, ** p<0.05, *** p<0.01"

Besides, to strengthen the belief that the above results are credible, I perform a placebo treatment test in Table 13, using reviews mentioning the price of the products. The concern is that the count of reviews is capturing other possible market performance records instead of a source of knowledge. The test is informative because user reviews about price should not affect the likelihood of updates after controlling for promotions, but the purported endogeneity would operate in a similar way. After controlling for the actual treatment, which guarantees the independent variables are not contaminated (Eggers et al. 2021), I replicate two main results in column (1) and column (3) and two corresponding placebo tests in column (2) and column (4). Both the significance and the magnitude of the coefficients do not change.

Table 13 Robustness Check - Placebo Test

	(1) *Update*	(2) *Update*	(3) *Update*	(4) *Update*
log (Reliability)	0.014	0.012	-0.011	-0.013
	(0.033)	(0.035)	(0.013)	(0.013)
log (Design)	0.099***	0.095***	0.022	0.019
	(0.032)	(0.034)	(0.014)	(0.015)
log (Price)		0.011		0.006
		(0.037)		(0.011)
*log (Reliability) * Positive*	-0.009	-0.008		
	(0.040)	(0.041)		
*log (Design) * Positive*	-0.098***	-0.097**		
	(0.037)	(0.040)		
*log (Price) * Positive*		-0.006		
		(0.045)		
*log (Reliability) * Firm Age*			0.001**	0.002**
			(0.001)	(0.001)
*log (Design) * Firm Age*			-0.001	-0.000
			(0.001)	(0.001)
*log (Price) * Firm Age*				-0.000
				(0.001)
Time FE	Yes	Yes	Yes	Yes
Game FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	13230	13230	13230	13230

Standard errors in parentheses clustered at game level
*p<0.10, ** p<0.05, *** p<0.01”

Another concern is that the result is not driven by producers or users but by the nature of games. It is possible some features of the games require the producer to come up with updates on

a frequent basis. For example, single-player games differ from multi-player games in many ways, including revenue-generating strategy, demands of complementary assets (contract or ownership of internet servers), direct network effect, etc. These features may confound the relationship between user reviews and updates. To address the concern, I generated a series of dummies associated with game features and then interacted with the review counts, respectively. The results are shown in Table A3, and none of the features is significant.

5. Conclusion

Platforms and complementors are increasingly enthusiastic about knowledge in the user community to understand user preferences, and user reviews are generally the most direct way of gaining knowledge from users. This study investigates whether producers learn from their users and which type of user knowledge is more informative for them. Using a dataset from the video game platform *Steam*, I find that updates of producers are positively correlated with the count of user reviews under certain circumstances. Specifically, when the product is poorly rated, producers are more likely to respond with an update to the game and learn from the design-related reviews but not the reliability reviews. They also benefit from the number of helpful reviews but not all reviews. Moreover, I identify that large firms can effectively take advantage of the reliability reviews compared to individual game developers and studios. As reliability is the bottleneck for users to evaluate products of large firms but not small studios, large firms invest more in addressing concerns with reliability.

The existing literature on user communities focuses on the interaction among the community members. The creation, enhancement, and selection of the user-generated contents and involvement are meaningful for generating a positive feedback loop (e.g. Claussen et al. 2021, Reimers and Waldfogel 2021), and my work links the elicitation of information within the loop to the other side of the market: producers. While both researchers and practitioners have recognized the importance of user reviews on producers regarding marketing strategies, I show the possibility of treating the user as a knowledge source in addition to the roles of product buyers to prolong the life of products. Platform owner has great power when entering the competition over the complementors because they know which product space is more popular (Zhu and Liu 2018), and the entry of platform owners shifts complementor innovation to unaffected areas (Wen and Zhu 2019). My result complements the story by adding that when faced with competition with strong competitors, small studios choose not to compete in reliability since they must somehow overlook the developed consumers (regardless of their dissatisfaction with reliability) and delve into new content by differentiation. Overall, this paper points to the dynamic paths of complementors of the platform in terms of product development and shows that if they can seek help through existing but unorganized user review, they can at least earn a better position in the competition.

There remain several avenues for further research. First, I revealed that the platform-hosted user community could be used by the producers to generate ideas for incremental product development. I consider the scenario where producers are making a commitment to the games and investigate under such assumption whether user reviews would be a great motivation and assistance. As I restrict the sample to the successful games and rule out the alternatives by examining the development of DLC, it would highly be the case that the producers continue to work on those successful games. However, it is also true that other producers with less successful games would launch new projects when the games received negative reviews. It is interesting to ask under what circumstance the producers would continue to improve the unwelcomed products in contrast to diversifying their products. Second, Foerderer (2020) finds that the interfirm knowledge spillover among platform complementors leads to more innovation measured by app updates. Similarly, producers, especially large ones, can get access to the reviews of competitors' products and absorb them for their own benefit, if necessary, since the information is public. From the platform owners' perspectives, this would create more value through the perfection of games. However, as the learning capability is disproportionately biased towards large corporate, the interaction between large firm production teams and small firms' user groups might drain the users from the small firms in the long run. Whether it is beneficial for small firms remains unclear. Last, future research could examine the relationship in other industry sectors. It would be interesting to see whether and how reviews on physical products impact generational product development.

Appendix

Table A1 OLS - Different Lag Use

	(1)	(2)	(3)	(4)	(5)	(6)
	Update	*Update*	*Update*	*Update*	*Update*	*Update*
Promotion	0.094***	-0.053***	0.093***	-0.053***	0.093***	-0.053***
	(0.016)	(0.009)	(0.016)	(0.009)	(0.016)	(0.009)
Age	-0.011***	-0.006***	-0.011***	-0.006***	-0.011***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ratings	-0.159***	-0.146***	0.078	0.060	-0.161***	-0.148***
	(0.044)	(0.043)	(0.048)	(0.044)	(0.043)	(0.042)
log (Owner)	-0.041***	-0.034***	-0.042***	-0.034***	-0.040***	-0.033***
	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)	(0.008)
log (Users)	0.036***	0.027***	0.034***	0.024**	0.038***	0.028***
	(0.011)	(0.010)	(0.012)	(0.010)	(0.011)	(0.010)
log (Reliability)	0.007	0.005	0.014	0.048	-0.008	-0.012
	(0.008)	(0.008)	(0.033)	(0.032)	(0.011)	(0.011)
log (Design)	0.016*	0.011	0.099***	0.054*	0.014	0.012
	(0.009)	(0.009)	(0.032)	(0.030)	(0.012)	(0.011)
*log (Reliability) * Ratings*			-0.009	-0.053		
			(0.040)	(0.038)		
*log (Design) * Ratings*			-0.098***	-0.049		
			(0.037)	(0.034)		
*log (Reliability) * Firm*					0.035**	0.039**
					(0.016)	(0.016)
*log (Design) * Firm*					0.008	0.001
					(0.018)	(0.017)
Game & Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Lag	1 Month	2 Months	1 Month	2 Months	1 Month	2 Months
R^2	0.137	0.122	0.141	0.125	0.140	0.124

Standard errors in parentheses, clustered at game level
* p¡0.10, ** p¡0.05, *** p¡0.01

Table A2 Robustness Check - Adding Lag

	(1) $Update_{t+1}$	(2) $Update_{t+1}$	(3) $Update_{t+1}$	(4) $Update_{t+1}$
$Promotion_t$	0.094***	0.095***	0.094***	0.095***
	(0.016)	(0.016)	(0.016)	(0.016)
Age_t	-0.011***	-0.011***	-0.011***	-0.010***
	(0.001)	(0.002)	(0.001)	(0.002)
$Ratings_t$	-0.164***	-0.257***	-0.159***	-0.252***
	(0.045)	(0.066)	(0.044)	(0.065)
$log(Owner_t)$	-0.039***	-0.033***	-0.041***	-0.034***
	(0.009)	(0.009)	(0.009)	(0.009)
$log(Users_t)$	0.028**	0.023	0.036***	0.029**
	(0.012)	(0.015)	(0.011)	(0.014)
$log(Count_t)$	0.033***	0.031**		
	(0.009)	(0.012)		
$log(ReliabilityCount_t)$			0.007	0.007
			(0.008)	(0.011)
$log(DesignCount_t)$			0.016*	0.016
			(0.009)	(0.012)
$Ratings_{t-1}$		0.124***		0.121***
		(0.039)		(0.039)
$log(Owner_{t-1})$		0.030		0.032
		(0.020)		(0.020)
$log(Users_{t-1})$		-0.002		-0.007
		(0.009)		(0.009)
$log(Count_{t-1})$		-0.016*		
		(0.009)		
$log(ReliabilityCount_{t-1})$				-0.002
				(0.008)
$log(DesignCount_{t-1})$				-0.009
				(0.008)
Time FE	Yes	Yes	Yes	Yes
Game FE	Yes	Yes	Yes	Yes
AIC	8927.440	8487.161	8934.525	8491.203
BIC	9099.716	8688.768	9114.290	8707.743

Standard errors in parentheses, clustered at game level
* p¡0.10, ** p¡0.05, *** p¡0.01

Table A3 OLS - Interaction of Game Features

	(1) *Update*	(2) *Update*	(2) *Update*	(3) *Update*	(4) *Update*	(5) *Update*	(6) *Update*	(7) *Update*	(8) *Update*	(9) *Update*
log (Reliability)	0.007	-0.012	0.010	0.009	0.004	0.014	-0.001	0.009	0.011	0.014
	(0.008)	(0.022)	(0.011)	(0.013)	(0.010)	(0.009)	(0.011)	(0.010)	(0.009)	(0.009)
log (Design)	0.016*	0.023	0.010	0.0001	0.015	0.018*	0.024**	0.016	0.011	0.012
	(0.009)	(0.023)	(0.011)	(0.014)	(0.011)	(0.010)	(0.012)	(0.010)	(0.010)	(0.010)
*Game Feature**		0.025	-0.006	-0.003	0.016	-0.026	0.021	-0.007	-0.019	-0.073***
log (Reliability)		(0.023)	(0.016)	(0.017)	(0.017)	(0.020)	(0.017)	(0.017)	(0.021)	(0.024)
*Game Feature **		-0.009	-0.010	0.025	0.008	-0.007	-0.022	-0.000	0.035	0.036
log (Design)		(0.025)	(0.016)	(0.017)	(0.019)	(0.021)	(0.016)	(0.022)	(0.024)	(0.030)
Game Feature		Singleplayer	Multiplayer	Action	RPG	Simulation	Adventure	Casual	Stargegy	Racing
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Game FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13230	13230	13230	13230	13230	13230	13230	13230	13230	13230

Standard errors in parentheses, clustered at game level
* p¡0.10, ** p¡0.05, *** p¡0.01

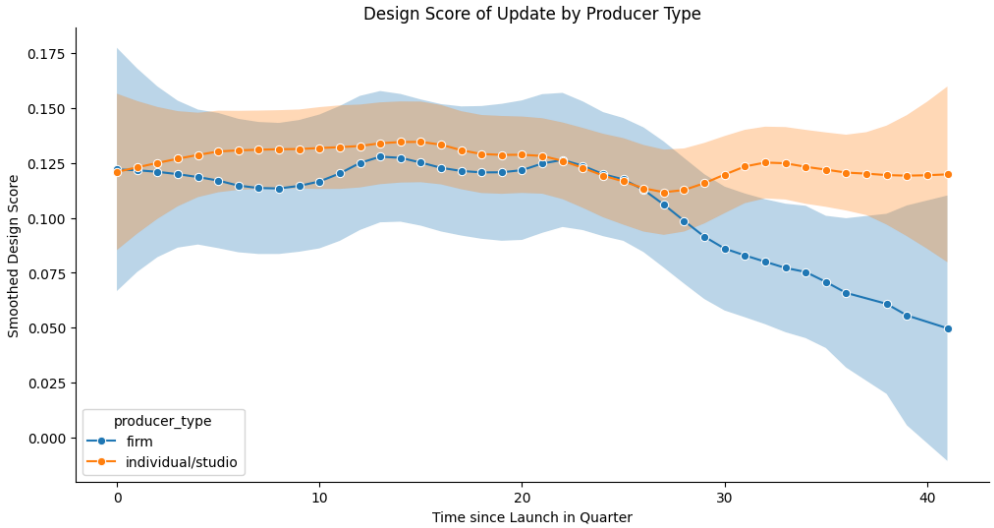


Figure A1 Percentage of Updates Related to Deisgn

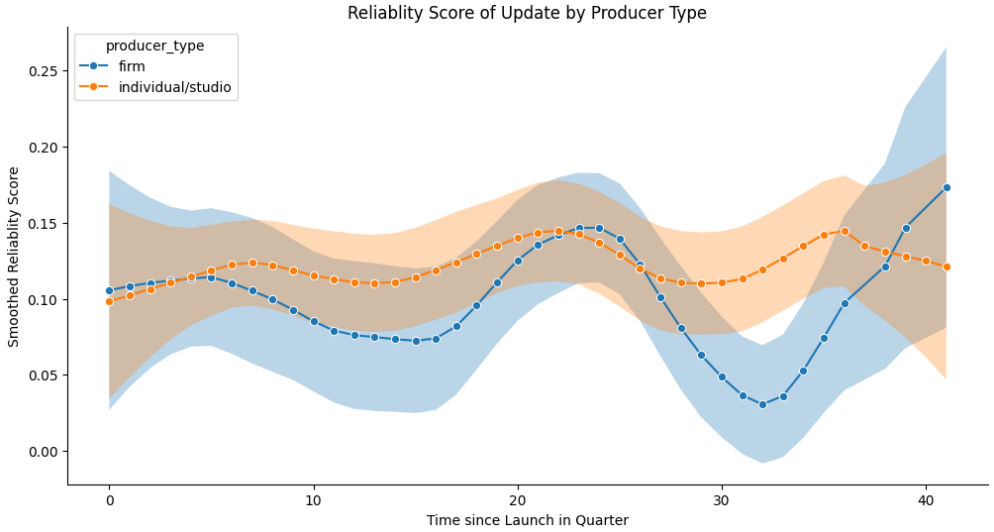


Figure A2 Percentage of Updates Related to Reliability

References

Antorini YM, Muñiz Jr AM, Askildsen T (2012) Collaborating with customer communities: Lessons from the lego group. *MIT Sloan Management Review* .

Aral S, Weill P (2007) It assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation. *Organization science* 18(5):763–780.

Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management science* 57(8):1485–1509.

Ba S, Pavlou PA (2002) Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly* 243–268.

Baldwin C, Von Hippel E (2011) Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization science* 22(6):1399–1417.

Bayus BL (2013) Crowdsourcing new product ideas over time: An analysis of the dell ideastorm community. *Management science* 59(1):226–244.

Bertschek I, Kesler R (2022) Let the user speak: Is feedback on facebook a source of firms' innovation? *Information Economics and Policy* 60:100991.

Chatterji AK, Fabrizio K (2012) How do product users influence corporate invention? *Organization Science* 23(4):971–987.

Chatterji AK, Fabrizio KR (2014) Using users: When does external knowledge enhance corporate product innovation? *Strategic Management Journal* 35(10):1427–1445.

Chen PY, Hong Y, Liu Y (2018) The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments. *Management Science* 64(10):4629–4647.

Chen Y, Xie J (2005) Third-party product review and firm marketing strategy. *Marketing science* 24(2):218–240.

Cheng X, Bala H, Yang M (2024) Engaging users on social media business pages: The roles of user comments and firm responses. *MIS Quarterly* 48(2).

Chevalier JA, Dover Y, Mayzlin D (2018) Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science* 37(5):688–709.

Claussen J, Ferreira P, Grad T, Sen A (2021) (how) does ugc impact professionals? evidence from local news. *Evidence from Local News (April 26, 2021)* .

Criscuolo P, Nicolaou N, Salter A (2012) The elixir (or burden) of youth? exploring differences in innovation between start-ups and established firms. *Research Policy* 41(2):319–333.

Day MY, Lin YD (2017) Deep learning for sentiment analysis on google play consumer review. *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, 382–388, URL `http://dx.doi.org/10.1109/IRI.2017.79`.

Dellarocas C (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science* 49(10):1407–1424.

Eggers AC, Tuñón G, Dafoe A (2021) Placebo tests for causal inference. *Unpublished manuscript. https://pelg. ucsd. edu/Eggers_2021. pdf* .

Fisher G, Kotha S, Lahiri A (2016) Changing with the times: An integrated view of identity, legitimacy, and new venture life cycles. *Academy of management review* 41(3):383–409.

Foerderer J (2020) Interfirm exchange and innovation in platform ecosystems: Evidence from apple's worldwide developers conference. *Management Science* 66(10):4772–4787.

Gans JS, Goldfarb A, Lederman M (2021) Exit, tweets, and loyalty. *American Economic Journal: Microeconomics* 13(2):68–112.

Garriga H, Von Krogh G, Spaeth S (2013) How constraints and knowledge impact open innovation. *Strategic Management Journal* 34(9):1134–1144.

Gimenez-Fernandez EM, Sandulli FD, Bogers M (2020) Unpacking liabilities of newness and smallness in innovative start-ups: Investigating the differences in innovation performance between new and older small firms. *Research policy* 49(10):104049.

Goes P, Tu Y, Tung YA (2013) Seller heterogeneity in electronic marketplaces: A study of new and experienced sellers in ebay. *Decision Support Systems* 56:247–258.

Hollenbeck B, Moorthy S, Proserpio D (2019) Advertising strategy in the presence of reviews: An empirical analysis. *Marketing Science* 38(5):793–811.

Jeppesen LB, Frederiksen L (2006) Why do users contribute to firm-hosted user communities? the case of computer-controlled music instruments. *Organization science* 17(1):45–63.

Jeppesen LB, Laursen K (2009) The role of lead users in knowledge sharing. *Research policy* 38(10):1582–1589.

Karanam A, Agarwal A, Barua A (2024) Machine learning based product design: The case of mobile apps. *Forthcoming at Information Systems Research* .

Kim SM, Pantel P, Chklovski T, Pennacchiotti M (2006) Automatically assessing review helpfulness. *Proceedings of the 2006 Conference on empirical methods in natural language processing*, 423–430.

Kleis L, Chwelos P, Ramirez RV, Cockburn I (2012) Information technology and intangible output: The impact of it investment on innovation productivity. *Information Systems Research* 23(1):42–59.

Kong D, Yang J, Duan H, Yang S (2020) Helpfulness and economic impact of multidimensional rating systems: Perspective of functional and hedonic characteristics. *Journal of Consumer Behaviour* 19(1):80–95.

Lee D, Hosanagar K (2021) How do product attributes and reviews moderate the impact of recommender systems through purchase stages? *Management Science* 67(1):524–546.

Lettl C, Rost K, Von Wartburg I (2009) Why are some independent inventors 'heroes' and others 'hobbyists'? the moderating role of technological diversity and specialization. *Research Policy* 38(2):243–254.

Li L, Gopinath S, Carson SJ (2022) History matters: the impact of online customer reviews across product generations. *Management Science* 68(5):3878–3903.

Lizardi R (2012) Dlc: Perpetual commodification of the video game. *Democratic Communiqué* 25(1):3.

Miller KD, Fabian F, Lin SJ (2009) Strategies for online communities. *Strategic management journal* 30(3):305–322.

Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *Journal of Marketing Research* 48(3):444–456.

Mudambi SM, Schuff D (2010) Research note: What makes a helpful online review? a study of customer reviews on amazon. com. *MIS quarterly* 185–200.

Murphy-Hill E, Zimmermann T, Nagappan N (2014) Cowboys, ankle sprains, and keepers of quality: How is video game development different from software development? *Proceedings of the 36th International Conference on Software Engineering*, 1–11.

Nickerson JA, Wuebker R, Zenger T (2017) Problems, theories, and governing the crowd. *Strategic Organization* 15(2):275–288.

Ogawa S, Piller FT (2006) Reducing the risks of new product development. *MIT Sloan management review* .

Ozuem W, Borrelli M, Lancaster G (2017) Leveraging the co-evolution of offline and online video games: An empirical study. *Journal of Strategic Marketing* 25(1):75–97.

Panico C, Cennamo C (2022) User preferences and strategic interactions in platform ecosystems. *Strategic Management Journal* 43(3):507–529.

Reimers I, Waldfogel J (2021) Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings. *American Economic Review* 111(6):1944–1971.

Rietveld J, Ploog JN (2022) On top of the game? the double-edged sword of incorporating social features into freemium products. *Strategic Management Journal* 43(6):1182–1207.

Salehan M, Kim DJ (2016) Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems* 81:30–40.

Schneider C, Weinmann M, Mohr PN, vom Brocke J (2021) When the stars shine too bright: The influence of multidimensional ratings on online consumer ratings. *Management Science* 67(6):3871–3898.

Schweisfurth TG, Raasch C (2015) Embedded lead users—the benefits of employing users for corporate innovation. *Research policy* 44(1):168–180.

Shah S, Nagle F (2019) Why do user communities matter for strategy? *Harvard Business School Strategy Unit Working Paper* (19-126).

Shah SK, Tripsas M (2007) The accidental entrepreneur: The emergent and collective process of user entrepreneurship. *Strategic entrepreneurship journal* 1(1-2):123–140.

Siering M, Deokar AV, Janze C (2018) Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. *Decision Support Systems* 107:52–63.

Sivam A, Dieguez T, Ferreira LP, Silva FJ (2019) Key settings for successful open innovation arena. *Journal of Computational Design and Engineering* 6(4):507–515.

Taeuscher K, Rothe H (2021) Optimal distinctiveness in platform markets: Leveraging complementors as legitimacy buffers. *Strategic Management Journal* 42(2):435–461.

Teece DJ, Pisano G, Shuen A (1997) Dynamic capabilities and strategic management. *Strategic management journal* 18(7):509–533.

Timoshenko A, Hauser JR (2019) Identifying customer needs from user-generated content. *Marketing Science* 38(1):1–20.

Truelove A, de Almeida ES, Ahmed I (2021) We'll fix it in post: what do bug fixes in video game update notes tell us? *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, 736–747 (IEEE).

Von Hippel E (1976) The dominant role of users in the scientific instrument innovation process. *Research policy* 5(3):212–239.

Von Hippel E (2005) Democratizing innovation: The evolving phenomenon of user innovation. *Journal für Betriebswirtschaft* 55:63–78.

Wen W, Zhu F (2019) Threat of platform-owner entry and complementor responses: Evidence from the mobile app market. *Strategic Management Journal* 40(9):1336–1367.

Wu L, Hitt L, Lou B (2020) Data analytics, innovation, and firm productivity. *Management Science* 66(5):2017–2039.

Xu Y, Ni J (2022) Entrepreneurial learning and disincentives in crowdfunding markets. *Management Science* 68(9):6819–6864.

Zhang H, Rao H, Feng J (2018) Product innovation based on online review data mining: a case study of huawei phones. *Electronic Commerce Research* 18:3–22.

Zhu F, Liu Q (2018) Competing with complementors: An empirical look at amazon. com. *Strategic management journal* 39(10):2618–2642.

Zhu K, Kraemer KL (2002) E-commerce metrics for net-enhanced organizations: Assessing the value of e-commerce to firm performance in the manufacturing sector. *Information systems research* 13(3):275–295.