Platform Monetization and Unintended Consequences on its Ecosystem: Evidence from a Two-sided Market for Books

KAI ZHU (MCGILL UNIVERSITY), QIAONI SHI (BOCCONI UNIVERSITY), SHRABASTEE BANERJEE (TILBURG UNIVERSITY)

Preliminary and incomplete: please do not circulate

How can a platform capture the value it creates for its users without damaging its ecosystem? In this study, we leverage a natural experiment on Goodreads.com to examine the potential intended and unintended consequences of monetizing a popular promotional program run by the platform: Goodreads Giveaways. Participating in this program was free for authors and publishers till January 2018, after which Goodreads enacted a policy change and began to charge a fixed participation fee. We collect large-scale data to analyze both the supply side (i.e., authors and publishers) and demand side (i.e., consumers) response to this monetization policy. We document several novel insights about the consequences of monetization that are above and beyond the traditional concern of network effects. Specifically, we find that Goodreads' monetization policy (i) increases supply concentration by increasing the representation of Big 5 publishers in the Giveaways marketplace, (ii) decreases product diversity by reducing participation from niche genres, and (iii) results in worse matches between consumers and products as measured by book ratings. Our findings highlight a more subtle and complex view of evaluating monetization and suggest that platforms need to counterbalance these effects by offering more flexible and nuanced incentive structures for different players in its ecosystem.

1 INTRODUCTION

Digital platforms have substantially changed markets for cultural goods by allowing creators to circumvent traditional intermediaries and directly market their products to potential customers. Concurrently, the information environment available to both consumers and creators has been dramatically improved via digitization: in addition to the information made accessible through recommendation algorithms and ratings systems, product success on digital platforms can inform creators and distributors about the appeal of new products before they reach a mass market [Peukert and Reimers, 2019].

All these features are exemplified by Goodreads.com, which is an online social platform where users search for books, track readings, write reviews, and connect with other book lovers. It has been a popular tool for book discovery, and is particularly valuable for its large repository of user-contributed reviews. In addition to providing these reviews, Goodreads has been running 'giveaway' contests (henceforth referred to as Giveaways) as a tool for authors and publishers to promote their works (see Figure 1). In short, Giveaways allows authors or publishers to distribute a pre-determined number of copies of their books to randomly chosen customers through a draw conducted on the website. Books are usually listed on the website for 1 to 2 months when consumers are asked to enter the giveaway, with winners being announced at the end of this period. Both pre-release as well as previously released books can be listed for Giveaways. Arguably, this program can boost a book's exposure, build an audience for the title, and help other readers discover and decide to read it. Over the past few years, Giveaways has become a core marketing tool for Goodreads.

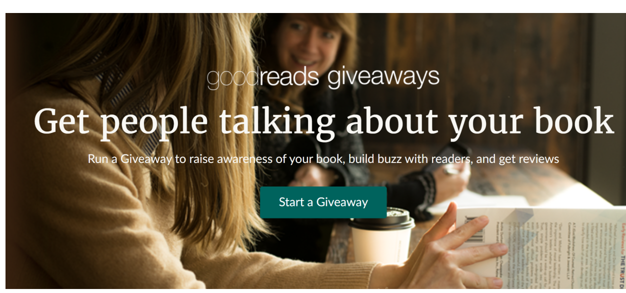

Fig. 1. An illustration of the Goodreads Giveaways program.

In January 2018, Goodreads started monetizing Giveaways by charging authors and publishers to list their books. Now, authors and publishers would have to pay to give away their own works. Creating a revenue stream is important for the sustainable growth of a platform. However, monetization is a challenging, multi-dimensional problem [Parker et al., 2016]. While one benefit of monetization could be to prevent authors from gaming the system and repeatedly exposing the same book, it also brings forth concerns of diversity and the representation of smaller, independent authors.[1] In particular, there is no clear guidance on how a platform can monetize the value it creates for its users without hurting network effects as well as the health of its ecosystem, especially in the context of a platform like Goodreads that deals with cultural products and does not charge overall platform participation fees. So far, there has been limited work in the literature examining the effects of platform monetization in a holistic way.

The advantage of our study setting is that we observe the dynamics of both supply side (i.e. authors and publishers) and demand side (i.e. readers) of the two-sided Giveaways market. This

[1] https://www.theverge.com/2017/11/29/16714972/goodreads-giveaways-program-changing-standard-premium-tiers-authors

puts us in a unique position to understand how the governance choice of monetizing this major feature on Goodreads will impact its ecosystem for product discovery and promotion. In particular, we ask (i) how does monetization affect the mix of authors and publishers that participate in Giveaways? (ii) how is book diversity, as measure by genre, affected by monetization? (iii) how does monetization impact the effectiveness of Giveaways on the demand side (in terms of how readers rate and review participating books)?

Using extensive data on Giveaways participation, book and publisher characteristics, and finally consumer ratings, we uncover several novel consequences of platform monetization in the context of Goodreads. First, we find a large overall decrease (nearly 66%) in the number of hosted Giveaways post-monetization. To ensure this does not reflect general platform-level trends, we further examine external word-of-mouth generated on Twitter. Doing so, we see that post-monetization, there are fewer tweets mentioning 'Goodreads giveaways', as well as lesser buzz in terms of unique user engagements and re-tweets, while tweets mentioning just Goodreads remain unchanged. Next, examining the concentration in publishers participating in Giveaways over time using the Hirfindahl-Hirschman Index, we find a sharp, discontinuous increase post-monetization, indicating that the effect on publishers is asymmetric. Delving deeper into raw distributions, we find that the proportion of top publishers rises by about 25% at the expense of self-publishing (which declines by 17-18%). Next, we examine whether this shift in supply has consequences for the kinds of products offered in the market, namely, the types of books that enter Giveaways. We focus on genre as the primary characteristic, and using entropy-based concentration measures, find that post-monetization, genres that were underrepresented further shrink in proportion, whereas popular genres gain market share. This illustrates a "rich-gets-richer" phenomenon and may have serious consequences for diversity and equity. Finally, we turn to the demand side, i.e, consumer responses to the books that enter Giveaways. We find using an event-study model and well as triple differences (differences-in-difference-in difference) that post-giveaway, books experience a decline in average ratings while review volume increases. This effect is further exacerbated post-monetization, in line with literature on the 'Groupon effect' [Byers et al., 2012]. This offers initial evidence that consumer-book matches worsen post-monetization as a consequence of supply concentration and lesser product variety on the market. Taken together, our results provide a picture of the impact of platform monetization overall, and helps to identify conditions under which it may benefit or hurt participants.

2 RELATED LITERATURE

Optimal strategies for platform monetization have been a growing concern with the increasing prevalence of two-sided markets in the Internet age. Traditionally, the literature on monetization has focused on optimal platform fees that can attract both sides of the market while maintaining a viable revenue stream [Rochet and Tirole, 2003]. This is determined by factors such as competition, market thickness, transactions elasticities, and pass-through rates (i.e, the ability of one party to pass costs to the other). These questions are particularly important for modern matching platforms such as Uber and Airbnb. In a related vein, there are several papers on 'freemium' platform models, and their consequences for revenue, as well as the kinds of consumers attracted by free trials [Datta et al., 2015, Shi et al., 2019]. However, optimal monetization schemes remain less studied in the context of cultural products. For instance, compensation schemes for artists on Spotify remains an active area of research (e.g, [Towse, 2020]). There has also been recent work in the context of a novel-writing platform demonstrating that revenue models (revenue-sharing vs pay-by-the-word) lead authors to respond differently to competition in terms of content novelty [Wu and Zhu, 2022]. Understanding these dynamics in more empirical contexts could help platforms make better informed decisions about monetization. Further, there is a dearth of literature dealing with monetization of specific platform features with the overall participation remaining free, as is the case in our application.

Our research also examines supply concentration and the resultant loss in product diversity in terms of books genres. Diversity has long been a critical scientific concept as well as an important societal focus in research [Cowell, 2000, Gini, 1921, Shannon, 1948]. In the context of digital platforms, previous works have demonstrated that consumption diversity is strongly associated with long-term user metrics [Anderson et al., 2020, Waller and Anderson, 2019]. This begs a central question in platform strategy: how to steer user behavior so that they can have a diverse consumption pattern while maintaining engagement [Anderson et al., 2020, Hansen et al., 2021]. A lot of current research attention has been devoted to understanding the impact of algorithmic recommendations on consumption diversity [Anderson et al., 2020, Hansen et al., 2021, Holtz et al., 2020]. This literature has found that recommendation systems often increase user engagement at the cost of lowering consumption diversity, which may be detrimental to long-term platform success. Despite this concern, we still have limited understanding of how consumption diversity may affect platform ecosystems and through which mechanism diversity plays a key role in the growth and evolution of platforms. Consumption diversity is also linked closely to the idea of product variety in general. Specifically in the context of books, it has been shown that increased product variety in online bookstores enhanced consumer welfare by $731 million to $1.03 billion in the year 2000, which is at least five times as large as the consumer welfare gain from increased competition and lower prices in this market [Brynjolfsson et al., 2003]. By this token, we expect diversity of offerings to play an important role in retaining and attracting consumers to Goodreads, which relies fundamentally on user content for its functioning.

In the competitive digital landscape for cultural products with thousands of products are fighting for consumers' attention (and money), it is also of fundamental importance to understand how platforms themselves can enable consumers to find the right kinds of products. There is some evidence in the literature that user reviews on Goodreads serve mostly a matching purpose: tracking the behavior of users over time reveals an increasing degree of specialization as they gather experience on the platform: they rate books with a lower average and number of ratings, while focusing on fewer genres. Thus, they become less similar to their average peer [Bondi, 2019]. This implies that incentives affecting the rating system (such as monetization) can affect the efficiency with which consumers match with books. Further, the matching process is likely to be hampered when the 'long tail' of genres shrinks as a result of supply concentration.

3 NATURAL EXPERIMENT AND DATA

To understand the impact of monetizing Giveaways, we collected extensive information on books, authors, publishers, and giveaway contests. In particular, we collected (1) the full set of giveaways hosted on the website from 2008 till 2020, (2) book, author and publisher level metadata and (3) star ratings and text reviews associated with each book.[2] This yields a total of 295,816 giveaway events for 201,573 books, and close to 90 million reviews and ratings. To rule out trend effects, we restrict our main estimation sample between Jan 2016 to Feb 2020 (before COVID hits), approximately 2 years before and 2 years after the monetization. We then exclude the books that were published more than 5 years before they participated in Giveaway, and the books that didn't get published within one year after participating.[3] This results in 82,552 books and 101,684 giveaway events.

[2] One drawback of our rating data is that Goodreads only allows for at most 3000 individual reviews and ratings per book to be scraped. However, we compare our collected sample to the raw numbers of 1-5 star ratings obtained via book-level metadata, and find very similar distributions, thus providing evidence that our sample is a valid characterisation of books' review distributions. This distribution is available in subsection A.1.

[3] Our results are also robust to different subsets of the data; these checks are available upon request.

In January 2018, Goodreads started to charge a Giveaways participation fee of $119 to $599. Will this fee impact how publishers or authors advocate their books via Goodreads? The short answer is yes. As shown in Figure 2, after January 2018, there is an immediate and large reduction in number of giveaway contests per month on Goodreads. From the figure, we observe that the average number of giveaways per month drops from around 3,000 to only 1,000 almost immediately after the policy change. This popular book discovery and promotion mechanism on the platform is suddenly used much less. This provides initial evidence that the monetary cost constitutes a significant obstacle for certain authors and publishers, and may have consequences for supply.

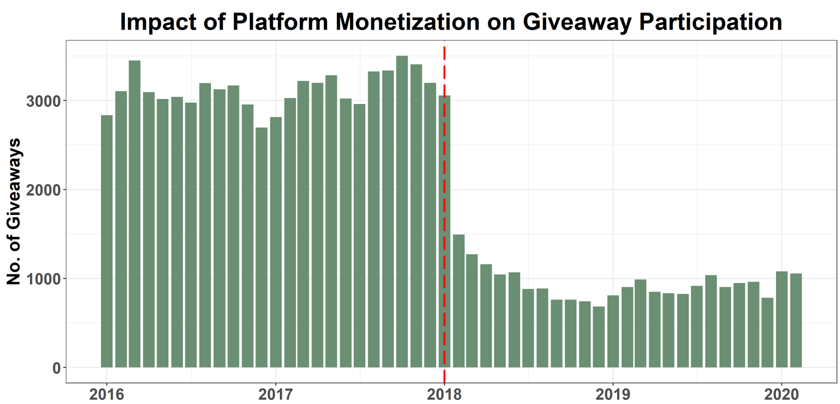

Fig. 2. Number of Giveaway contests drop sharply after monetization of Giveaways.

3.1 Word-of-mouth on Twitter

One major function of Giveaways is to get people talking about participating books (also exemplified in Figure 1). In addition to creating buzz and engagement on Goodreads itself, word-of-mouth on broader social media is central to creating awareness and raising the visibility of promoted books. As a check of spillover effects, and to gauge the impact of monetization on word-of-mouth outside of Goodreads, we thus provide some model-free evidence using Twitter data. Twitter's new API v2 endpoint of full-archive tweet counts allows us to retrieve the volume of tweets for a given query.[4]

Results are reported in Figure 3. In the lower panel, we observe a sharp drop in Twitter discussion volumes that mention both "Goodreads" and "Giveaways" around the time that monetization policy taking effects. Note that the change in word-of-mouth volume is not due to any platform-wide impact. The volume of Tweets discussing "Goodreads" only experiences no significant change around that time as shown in the upper panel of Figure 3. In other words, this shock is specific to the Giveaways program. This unintended consequence is not desirable from the standpoint of the platform, which promotes its main marketing tool, i.e. Giveaways, around having the power to create word-of-mouth.

Further, not only does the total amount of word-of-mouth on Twitter drop, other aspects of Twitter engagement also decrease. We further collect the full set of tweets that mentioned both "Goodreads" and "Giveaways" and summarize relevant tweet characteristics of this sample. We find that fewer unique users are talking about the Giveaway program and also lesser amplification

[4]See the official documentation of the endpoint: https://developer.twitter.com/en/docs/twitter-api/tweets/counts. It is academic research access only.

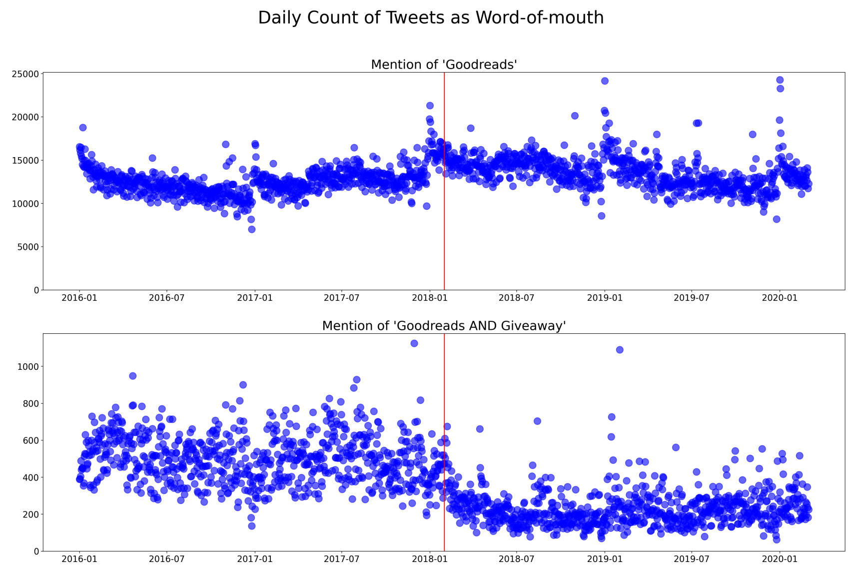

Fig. 3. Daily count of tweets as word-of-mouth. We see a decline in tweets mentioning both Goodreads and Giveaways post-monetization.

(i.e. retweets) for those tweets, as shown in Figure 4. This provides us an initial peek into how monetizing Giveaways may backfire for the platform. We will delve deeper into its impact on the platform ecosystem in the next few sections.

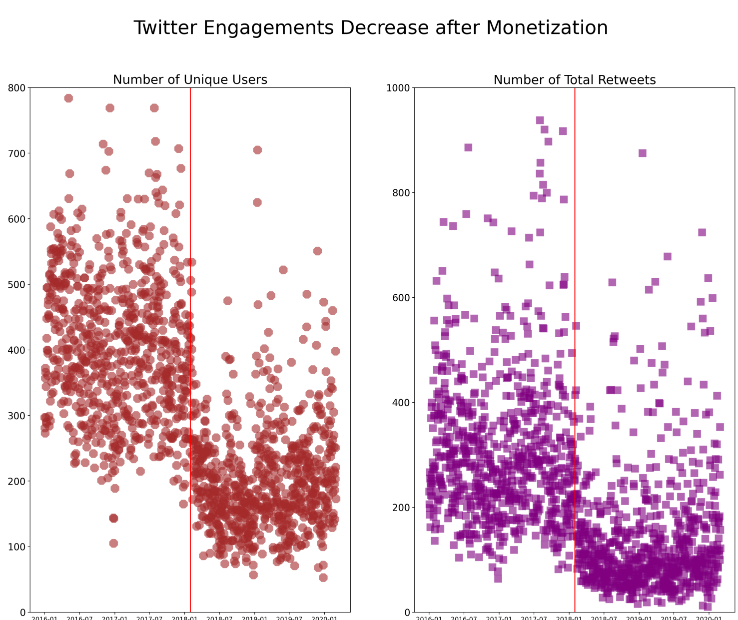

Fig. 4. Twitter engagement decreases after monetization.

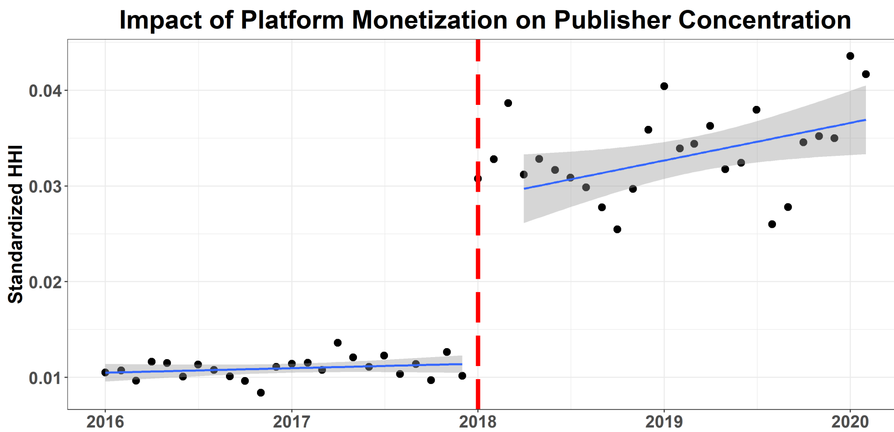

Fig. 5. The effect of giveaway monetization on the publisher concentration. We can observe a sharp increase in HHI, indicating that the publisher concentration increases significantly after monetization.

4 SUPPLY CONCENTRATION: IMPACT ON PUBLISHERS AND AUTHORS

While we observe an immediate drop in the total number of giveaway participations on Goodreads after Jan 2018, did monetization affect participation of different publishers in the same way? To understand this question, we narrow our focus on the publishers who participate in Giveaways.[5] First, we calculate the Herfindahl-Hirschman Index (HHI) for publishers in every year-month of our data. HHI, given by $\sum_{i=1}^{n} s_i^2$ is a measure of supply concentration, where s_i represents the market share of the publisher i. The larger the HHI is, the more concentrated publishers are [Hirshman, 1964]. Figure 5 shows that the HHI indeed surged after monetization, which indicates a concentration of power in the hands of a few publishers. In other words, monetization not only led to a reduction in the overall participation of books on Giveaway, but also impacted the mix of the types of participating publishers. Arguably, smaller publishers, who have fewer tools to market their books, are impacted the most by the policy change given their limited resources.

To further explore this hypothesis, in Figure 6, we now plot the monthly number and proportion of books participating in Giveaways for two categories of publishers: (i) Big Five publishing houses and (ii) self-publishing houses. Big Five publishing houses include five major publishers, Penguin Random House, Hachette, HarperCollins, Simon & Schuster, and MacMillan, and self-publishing houses produce books that are published directly by its authors. While Big Five publishing houses produced more than half of English-language books[6], self-publishing books often have a smaller market share. However, in raw terms, the number of books both kinds of publishers list on Giveaways is almost the same pre-monetization in Figure 6, indicating that self-publishing houses participate in Giveaways disproportionately more than Big 5 publishers.

However, this changes post-monetization. Figure 6 shows that the Giveaway participation of books published by both the Big Five and self-publishing houses dropped after monetization; however, a comparison of the two bar charts indicates a much larger impact of monetization on self-publishing books. Note that the proportion of books by the Big Five increased sharply from less than 20% to more than 30%, and the proportion of self-publishing books decreased immediately after monetization. This exploration suggests that monetization leads to supply concentration.

[5]Details on how this information is obtained from the raw data are provided in subsection A.2.
[6]https://en.wikipedia.org/wiki/Publishing#Book_publishing

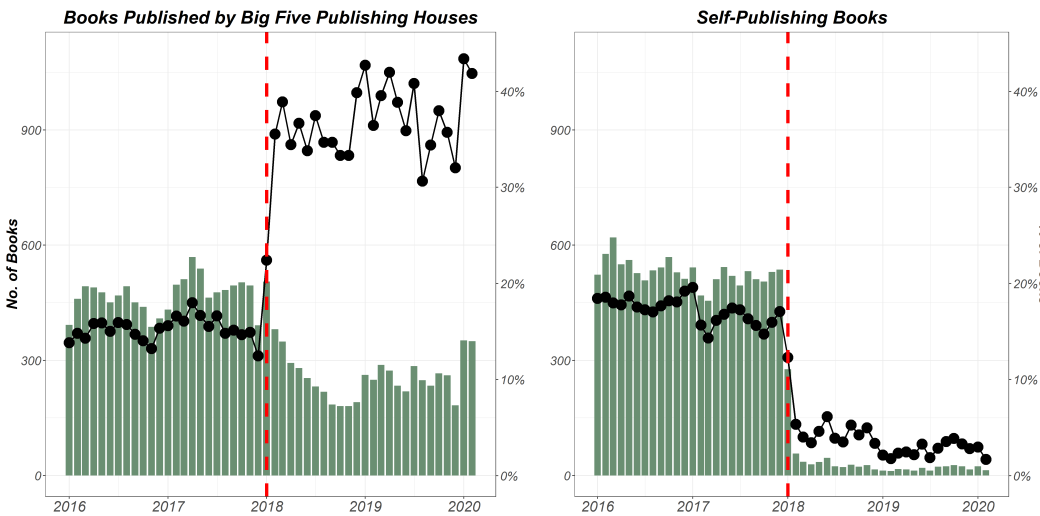

Fig. 6. The monthly number and proportion of books participating in Giveaways. The bar chart represents the number of books participating in Giveaways every month, while the line chart indicates the proportion. The number of books published by the Big Five and the number of self-publishing books decrease after monetization. However, the line chart shows the proportion of books by Big Five among all books in Giveaways almost double after Jan 2018.

Motivated by these results, we now conduct a regression analysis separately to demonstrate the effect of being a Big 5 publisher (relative to non-big 5). Notably, non-big 5 publishers excludes self-publishing entities in this analysis: since their mode of operation is drastically different for regular publishing, they may not constitute an appropriate baseline. We focus the analysis on the publisher level data by aggregating the monthly number of books participating on Giveaways for each publisher. We estimate regressions of the form:

$$y_{it} = \beta_0 + \beta_1 \times \text{Post-Monetization}_t + \beta_2 \times \text{Big-Five}_i + \beta_3 \times \text{Post-Monetization}_t \times \text{Big-Five}_i + \delta_t + \epsilon_{it}, \quad (1)$$

where Post-Monetization$_t$ is a dummy variable that equals 1 if time t is after Goodreads changed its Giveaway participation policy, Big-Five$_i$ indicates if publisher i is one of the Big-Five publishing houses, and δ_t denotes month fixed effects. We run the above model using two dependent variables y_{it}: (1) the number of total books for publisher i participating in Giveaways at time t, which is denoted as N_{it}; (2) the proportion of books on Giveaways at time t by publisher i, calculated as $Prop_{it} = \frac{N_{it}}{\sum_j N_{jt}}, j \neq i, Prop_{it} \in (0, 1]$. The result in Table 1 shows that although the overall monthly number of books on Giveaways by the Big Five decreased after monetization ($\beta = -43.335, p < 0.01$), the proportion of books by the Big Five publishing houses increased ($\beta = 0.026, p < 0.01$). This indicates that, compared to other publishers, the Big Five publishing houses are less impacted by the change in monetization policy. We find exactly analogous result for self-publishing houses, i.e, their participation decreases post-monetization, both in raw and proportional terms. These results are available in subsection A.2.

Finally, we examine how the mix of books on the market changes in terms of publisher characteristics before vs after monetization. Again, we are interested in Big 5 publishers and self-publishing houses as our primary examples. We estimate two regressions of the form:

Table 1. The impact of monetization on books published by Big Five publishing houses.

	Dependent variable:	
	No. of Books N_{it}	Proportion of Books Prop_{it}
	(1)	(2)
Post Monetization	-0.102^{***}	-0.00001^{***}
	(0.002)	(0.00000)
Big Five Publishing Houses	94.403^{***}	0.036^{***}
	(0.088)	(0.0001)
Post Monetization × Big Five Publishing Houses	-43.335^{***}	0.026^{***}
	(0.126)	(0.0001)
Month Fixed Effect	Yes	Yes
Observations	719,523	719,523
R^2	0.672	0.664
Note:		*p<0.1; **p<0.05; ***p<0.01

$$P_{jt} = \beta_0 + \beta_1 \times \text{Post-Monetization}_t + \delta_t + \epsilon_{jt} \tag{2}$$

where P_{jt} indicates whether a given book j in giveaway month t comes from a Big 5 publisher or is self-published respectively. We also add fixed effects for genre and month. Results are reported in Table 2. We see that for a given book, the likelihood that it will be from a Big 5 publisher rises by 11% following monetization. Conversely, the likelihood that it will be from a self-publishing house decreases by 5%, thus backing our claims so far.

Table 2. The book level impact of monetization on publisher proportions.

	Dependent variable:			
	Big Five Book		Self-Publishing Book	
	(1)	(2)	(3)	(4)
Post-Monetization	0.13^{***}	0.11^{***}	-0.07^{***}	-0.05^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Month Fixed Effect	Yes	Yes	Yes	Yes
Genre Fixed Effect	No	Yes	No	Yes
Observations	95,864	95,864	95,864	95,864
R^2	0.02	0.06	0.01	0.06
Note:				*p<0.1; **p<0.05; ***p<0.01

5 PRODUCT DIVERSITY: AN ANALYSIS OF BOOK GENRE

What is the implication of supply concentration on products offered on the market? To examine this, we turn to the books participating in Giveaways and investigate how they are affected by monetization. A particular aspect of books that we are interested in is genre diversity. A diverse selection of books attracts readers with different interests and satisfies potential variety-seeking behavior. As a result, more diversity is associated with larger cross-side network effects for readers as the supply side collectively becomes more appealing. Therefore, product diversity, as measured by diversity of book genre, is arguable a desired property for the ecosystem of platforms and users within it (the value of a 'long tail' in online assortments has been well-established in the literature, e.g [Brynjolfsson et al., 2003]).

To examine whether the monetization of Giveaways has an impact on genre diversity, we first infer genres for each book based on which shelves they are added to most frequently on Goodreads.[7] We then adopt Shannon entropy as our diversity measure over genre distribution. The concept of entropy originated from Shannon Information Theory [Shannon, 1948] and later has been widely used as a measure of diversity in multiple disciplines (also referred to as the Shannon diversity index). It is a quantitative measure that reflects how many different types there are and how individuals are distributed among those types. For each month, we compute the value of Shannon entropy for the collection of books that participate in Giveaways during that month. In Figure 7, we plot monthly standardized entropy of book genres over time. There is a sharp and significant drop in standardized entropy in the few months after January 2018, which indicates that the genre diversity of books decreased significantly after the monetization of the giveaway program. We obtain similar results also controlling for potential temporal dynamics (reported in subsection A.3).

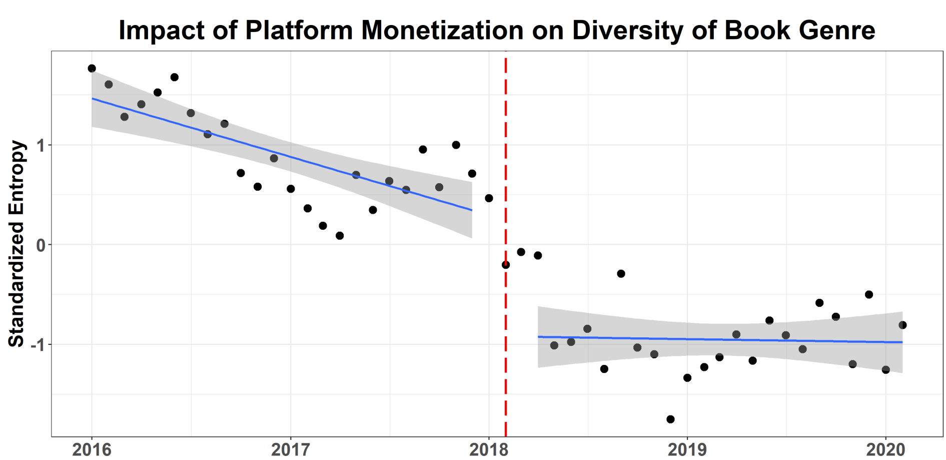

Fig. 7. The effect of giveaway monetization on the genre diversity of participating books. We can observe a sharp and significant drop in standardized entropy after treatment, indicating that the genre diversity of books decreased significantly after monetization.

Why did genre diversity decrease after monetization? To investigate the mechanism behind this drop, we further dive into the impact of monetization for each specific book genre. The key observation here is that a few popular genres become more dominant in genre proportion after monetization while the proportion of niche genres further shrink. In Figure 8, we first plot a few illustrative examples of genre proportion change before and after monetization. The plot has three examples for popular genres and three examples for niche genre. In the first row of the plot, we have *thriller*, *mystery*, and *historical fiction* - all of which have a relatively high proportion among

[7]More details about this process can be found in subsection A.3.

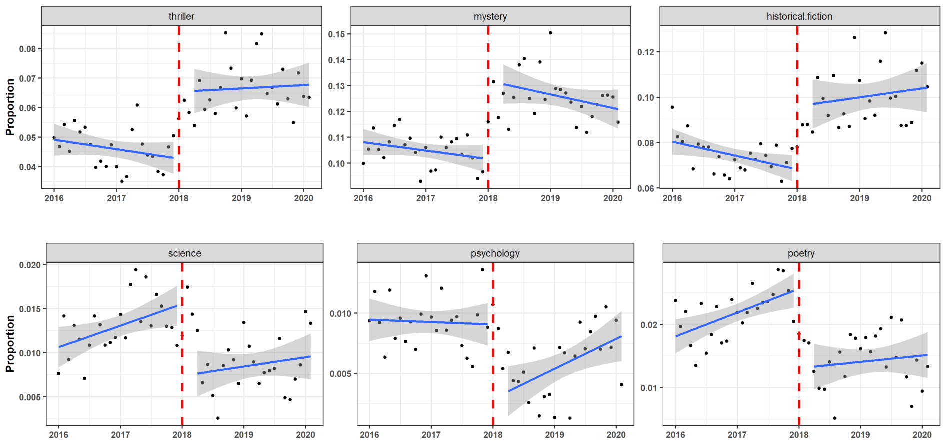

Fig. 8. Heterogeneous impact on book genres: rich gets richer and poor gets poorer.

giveaway books before monetization, and their proportions become even larger after. In the second row of the plot, we see the baseline proportion of genres *science*, *psychology*, *poetry* was low before monetization and becomes even lower after monetization.[8]

To follow up on this observation, we also conduct regression analysis with data of all 50 genres to examine heterogeneous impact of monetization across genre types. To do so, we estimate the following:

$$\begin{aligned} \text{GenreProportion}_{gt} = \beta_1 \times \text{Post Monetization}_{gt} \times \text{Top 25\% Genre}_g + \\ \beta_2 \times \text{Post Monetization}_{gt} \times \text{Bottom 25\% Genre}_g + \gamma_m + \epsilon_{gt} \end{aligned} \tag{3}$$

where the dependent variable is the genre proportion of genre g at giveaway month t. We classify genres into three categories based on their quartiles in the distribution of genre proportion during the pre-monetization period and add 0/1 indicators accordingly: top quartile (i.e. top 25% genre - popular group), bottom quartile (i.e. bottom 25% genre - niche group), and anything in between (i.e. quantile of the genre is between 25% and 75% - middle group). The model estimates are consistent with the visual examples and support the bipolar effect on book genres. The results are shown in Table 3. In column (1), the model estimates indicate that the proportions of popular genres increase by 1.2% on average after monetization while the proportion of niche genres decrease by 0.1% on average after monetization. Both estimates are statistically significant and in comparison with the middle group. The niche genres have a low baseline and hence even a small change in absolute terms could mean a big impact. To gauge the importance of change in proportion relative to the genres' own baseline, we also compute the percentage change in proportion and use it as an alternative dependent variable in the model.[9] In Column (2) of Table 3, we see that the proportion of niche genres decrease by as much as 26% percent - a big drop relative to its own baseline. On the other hand, a 1.2% increase in raw proportion translates to an 11% increase after monetization for popular genres.

[8] We also provide the full set of proportion change plots for all 50 genres in subsection A.3.

[9] For example, if romance has a genre proportion of 33% in Nov 2019 and its average proportion in the pre-period is 30%, then the percent change data point for Nov 2019 would be 33%/30% = 110% - i.e. it has a 10% increase relative to its own pre-period baseline.

Table 3. The impact of monetization on genre proportions of popular and niche books, in terms of both raw and percent changes.

	Dependent variable:	
	Raw Proportion	Percentage Change
	(1)	(2)
Post Monetization × Top 25% Genre	0.012***	0.110***
	(0.001)	(0.010)
Post Monetization × Bottom 25% Genre	−0.001***	−0.262***
	(0.0001)	(0.039)
Month Fixed Effect	Yes	Yes
Genre Fixed Effect	Yes	Yes
Observations	2,350	2,350
R^2	0.969	0.188
Note:		*p<0.1; **p<0.05; ***p<0.01

To summarize, we find evidence that monetizing Giveaways leads to a large decrease in the genre diversity of books offered through the program. The drop in diversity can be explained by the fact that genres that are already popular before monetization gain market share at the cost of genres of low popularity, which have an even lower proportion after monetization. Overall, these results highlight that monetization leads to a rich-gets-richer and poor-gets-poorer phenomenon in genre diversity of books. This is potentially a consequence of only a selected set of publishers and authors could still take advantage of the giveaway program after the monetization (as demonstrated in the previous section about supply concentration). In addition, even for those publishers and authors who still use Giveaway after monetization, they might change their behavior and only put books that are more mainstream and of general interest. Both composition change and behavior change of the supply side may result in a decrease in genre diversity, a result we hope to explore in more detail in the future.

6 DEMAND MISMATCH: IMPACT ON BOOK REPUTATION

Finally, we examine whether the above supply and product shifts induced by monetization have an effect on consumer demand (proxied by reviews and ratings). We estimate regressions of the form:

$$r_{jt} = \alpha_j + \gamma_t + \beta_1 \times \text{Post-Giveaway}_{jt} + \beta_2 \times \text{Post-Giveaway}_{jt} \times \text{Post-Monetization}_j + \epsilon_{jt} \quad (4)$$

where r_{jt} denotes in turn (i) the average rating and (ii) the review volume of book j in month t.[10] We further include book fixed effects to account for time invariant changes, and year-month fixed effects to account for general fluctuations in demand. Standard errors are clustered at the book level. Post-Giveaway is an indicator variable set to 1 after the giveaway date, and Post-Monetization is an indicator variable set to 1 for books that participate in the giveaway program after monetization. The parameter of interest is β_2, which can be interpreted as the effect of monetization on giveaway participation after January 2018. For our main analysis, we restrict ourselves to a 2 year window of rating arrival around the Giveaway participation date (12 months before and after) to minimize

[10]Results using disaggregated average ratings are very similar and reported in subsection A.4.

the impact of long term trend effects. However, our results remain consistent with the inclusion of all data points - these results are reported in subsection A.4. Overall, we find that Giveaway participation lowers average ratings and raises review volume, even in the absence of monetization (Table 4 and Table 5). These effects are consistent with the “Groupon effect”[Byers et al., 2012], wherein deals lead to greater uptake of the product at the cost of attracting lower ratings. One possible mechanism is a mismatch in preferences between reader and book when books are offered for free (a similar pattern of results has also found by [Zegners, 2017].) Additionally, we find that average ratings go down by about 0.07 stars post-monetization (Table 4 column 2), whereas the volume of reviews increases by 14 (Table 5 column 2). This demonstrates an exaggerated “Groupon effect” that comes about due to monetization.

Table 4. The impact of giveaway and monetization on avg ratings, measured at the year-month level for 12 months before and after giveaway participation.

	Dependent variable:	
	Monthly Avg Rating	
	(1)	(2)
Post Giveaway	-0.215^{***}	-0.192^{***}
	(0.003)	(0.004)
Post Giveaway × Post Monetization		-0.068^{***}
		(0.008)
Number of books	80498	80498
Overall mean rating	3.73	3.73
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	961,426	961,426
Adjusted R^2	0.512	0.512
Note:	*p<0.1; **p<0.05; ***p<0.01	

These results point to the fact that in general, Giveaway does not facilitate the best preference match between consumers and books, and the supply side changes described above lead to even worse matches, although book adoption (and hence review volume) increases.

Next, we specifically look at the number of posted reviews that contain text, to quantify whether text reviews and the length of these reviews change post-monetization. Consistent with the analysis above, we find that the number of text reviews also go up post-monetization (Table 6). However, looking at the text of posted reviews, we find that reviews tend to be shorter by about 38 characters after monetization (Table 7 estimates a specification with observations at the review level; an additional control is added for the number of days since the release date of the book). This points to the fact that reviewers may be less engaged and invested with the content they are reviewing, and leave shorter reviews that are more negative in valence.

6.1 Triple differences analysis

The above analysis implicitly constructs a control group based on the sample of books that participate in Giveaway pre-monetization. In other words, the post-monetization effect sizes are

Table 5. The effect of giveaway and monetization on review volume, measured at the year-month level for 12 months before and after giveaway participation.

	Dependent variable:	
	Monthly Num. of Rating	
	(1)	(2)
Post Giveaway	13.959***	9.322***
	(0.237)	(0.239)
Post Giveaway × Post Monetization		14.513***
		(0.501)
Number of books	81058	81058
Mean rating count	13.36	13.36
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	1,427,841	1,427,841
Adjusted R^2	0.156	0.158
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 6. The effect of giveaway and monetization on text review volume, measured at the year-month level for 12 months before and after giveaway participation.

	Dependent variable:	
	Monthly Num. of Text Reviews	
	(1)	(2)
Post Giveaway	1.713***	2.847***
	(0.107)	(0.094)
Post Giveaway × Post Monetization	3.550***	
	(0.199)	
Number of books	81058	81058
Overall mean text reviews	4.14	4.14
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	1,427,841	1,427,841
Adjusted R^2	0.177	0.176
Note:	*p<0.1; **p<0.05; ***p<0.01	

interpreted relative to this sample. However, this empirical strategy may not be valid if the evolution in ratings of pre- vs post-monetization participants differ for unobservable reasons.[11] To address this issue, we collect additional data that enables the creation of an independent control group based on which we can conduct a triple-differences analysis (difference-in-differences-in-differences, or DDD).

[11]Note that qualitative observable differences in pre vs post-monetization partipants (such as genre and publisher type) do not lead to a violation of identification assumptions in and of themselves, as long as parallel trends assumptions are valid.

Table 7. The impact of giveaway and monetization on review length, measured at the day level.

	Dependent variable:			
	Review length			
	(1)	(2)	(3)	(4)
Post Giveaway	-141.621^{***}	-153.277^{***}	-123.453^{***}	-137.915^{***}
	(1.826)	(1.987)	(1.996)	(2.260)
Time Since Release		0.108^{***}		0.097^{***}
		(0.006)		(0.005)
Post Giveaway × Post Monetization			-49.970^{***}	-38.870^{***}
			(2.849)	(2.918)
Number of books	77990	77990	77990	77990
Overall mean text length	743.14	743.14	743.14	743.14
Book FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	9,343,382	9,343,374	9,343,382	9,343,374
Adjusted R^2	0.068	0.068	0.068	0.068
Note:				*p<0.1; **p<0.05; ***p<0.01

Conceptually, DDD takes place in two DD steps. First, we compute a DD for books that participate in Giveaway after monetization, similar to Equation 4. Then, we adjust this DD for unobserved differences by subtracting from it the DD after monetization for books that do not participate in Giveaways.

6.2 Constructing control group using Goodreads' "Readers also enjoyed" feature

To construct the control group, we scrape details of a set of books that appear alongside each book that participates in Giveaway (henceforth referred to as focal books). These books appear under the banner 'Readers also enjoyed' (henceforth referred to as similar books), illustrated in Figure 9. To prevent many-to-many mapping, we randomly associate each scraped similar book with a single focal book. Each focal book, on the other hand, may be associated with multiple similar books. We then estimate a triple difference specification using matched-pair book fixed effects.

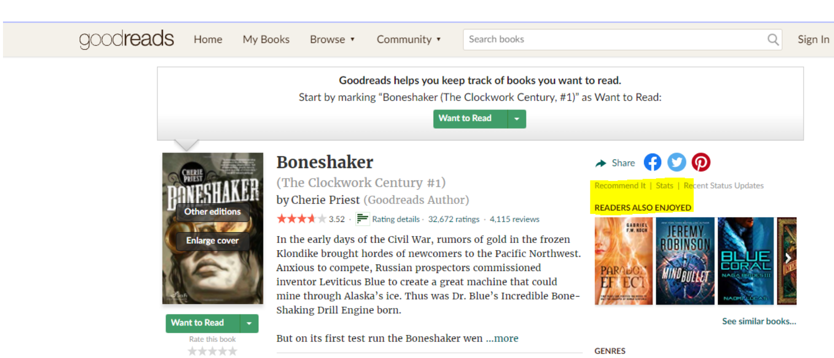

Fig. 9. An illustration of the 'Readers also enjoyed' feature.

The triple difference estimator can be computed as the difference between two difference in-differences estimators. Despite this, the triple difference estimator does not require two parallel trend assumptions to have a causal interpretation. The intuition is that the difference between two biased difference-in-differences estimators will be unbiased as long as the bias is the same in both estimators. In that case, the bias will be differenced out when the triple difference is computed. This requires only one parallel trend assumption, in ratios, to hold.

In this case, we consider three separate indicator variables: Treated (=1 if the book participates in a giveaway), Post-Giveaway (=1 for all periods after Giveaway participation) and Post-Monetization (=1 for books that participate in Giveaway after January 2018). The estimate of interest for us is the effect of giveaway participation on books that participate in a giveaway after monetization. This is given by β_5 in the equation below:

$$\begin{aligned} \mathrm{r}_{jt} &= \alpha_j + \gamma_t + \beta_1 \times \text{Post-Giveaway}_{jt} + \beta_2 \times \text{Post-Giveaway}_{jt} \times \text{Treated}_j + \\ &\beta_3 \times \text{Post-Monetization}_{jt} \times \text{Treated}_j + \beta_4 \times \text{Post-Giveaway} \times \text{Post-Monetization} + \\ &\beta_5 \times \text{Post-Giveaway} \times \text{Post-Monetization} \times \text{Treated}_j + \epsilon_{jt} \end{aligned} \tag{5}$$

Estimating this specification, we again find that average ratings of books that participate in Giveaway post-monetization are lower by about 0.75 stars, whereas review volume increases by 23 (Table 8 and Table 9).[12] Hence, even after considering a more stringent functional form that accounts for omitted confounders, our main results hold, and effect sizes are in fact larger.

7 CONCLUSION

Platform monetization is a pressing challenge for modern internet businesses, especially when coupled with promoting diversity in both consumption and production. In this paper, we try to examine the impact of monetization from various lenses in the context of Goodreads.com. We examine a natural experiment where Goodreads monetizes its main marketing and promotion tool for product discovery, the Giveaways program. Examining both the supply and demand side of the Giveaways marketplace, we first find that the mix of publishers is skewed in favour of more established entities after monetization - the representation of Big 5 publishing houses increases at the expense of self-publishing houses. This change in concentration trickles down to the diversity of book genres offered through the program: we find that more popular genres gain at the expense of less popular ones. Finally, on the demand side, we find that the promotional effects of Giveaways are amplified further post-monetization - books are adopted more and hence attract more ratings, but these ratings are more negative on average. Taken together, we demonstrate novel and unintended consequences of platform monetization in the context of cultural products. Going forward, our focus will be to uncover more dimensions of diversity and how they are impacted (e.g race and gender of authors). It has been established that platform businesses may have implicit biases against minority users (e.g [Edelman and Luca, 2014]). Examining what such biases might mean in our context will be particularly relevant. Exploring the text of consumer reviews can also help us better understand ways in which supply concentration can lead to consumption mismatch.

[12]For these results, we focus on the entire data and not just a 12 month window before vs after monetization, due to the presence of a control group that can account for trend effects. This is what leads to differences in the sample sizes across the tables.

Table 8. The effect of giveaway on average ratings, measured at the year-mon level, estimated using a triple-difference

	Dependent variable:
	Avg Rating
Post Giveaway	-0.196^{***}
	(0.005)
Post Giveaway × Treated	0.216^{***}
	(0.005)
Post Giveaway × Post Monetization	0.192^{***}
	(0.008)
Post Monetization × Treated	0.565^{***}
	(0.016)
Post Giveaway × Post Monetization × Treated	-0.749^{***}
	(0.014)
Number of books	81622
Overall average rating	3.45
Book FEs	Yes
Year-month FEs	Yes
Observations	6,516,591
Adjusted R^2	0.324
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9. The effect of giveaway on review volume, measured at the year-mon level, estimated using a triple-difference

	Dependent variable:
	Review Volume
Post Giveaway	9.697^{***}
	(0.180)
Post Giveaway × Treated	-19.550^{***}
	(0.142)
Post Giveaway × Post Monetization	-9.371^{***}
	(0.353)
Post Monetization × Treated	-6.810^{***}
	(0.386)
Post Giveaway × Post Monetization × Treated	23.127^{***}
	(0.457)
Number of books	81622
Overall review volume	20.74
Book FEs	Yes
Year-month FEs	Yes
Observations	6,516,591
Adjusted R^2	0.119
Note:	*p<0.1; **p<0.05; ***p<0.01

REFERENCES

Ashton Anderson, Lucas Maystre, Ian Anderson, Rishabh Mehrotra, and Mounia Lalmas. 2020. Algorithmic Effects on the Diversity of Consumption on Spotify. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 2155–2165.

Tommaso Bondi. 2019. Alone, together: Product discovery through consumer ratings. *NYU Stern School of Business* (2019).

Erik Brynjolfsson, Yu Hu, and Michael D Smith. 2003. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management science* 49, 11 (2003), 1580–1596.

John W Byers, Michael Mitzenmacher, and Georgios Zervas. 2012. The groupon effect on yelp ratings: a root cause analysis. In *Proceedings of the 13th ACM conference on electronic commerce*. 248–265.

F A Cowell. 2000. Chapter 2 Measurement of inequality. In *Handbook of Income Distribution*. Vol. 1. Elsevier, 87–166.

Hannes Datta, Bram Foubert, and Harald J Van Heerde. 2015. The challenge of retaining customers acquired with free trials. *Journal of Marketing Research* 52, 2 (2015), 217–234.

Benjamin G Edelman and Michael Luca. 2014. Digital discrimination: The case of Airbnb. com. *Harvard Business School NOM Unit Working Paper* 14-054 (2014).

Corrado Gini. 1921. Measurement of inequality of incomes. *The economic journal* 31, 121 (1921), 124–126.

Christian Hansen, Rishabh Mehrotra, Casper Hansen, Brian Brost, Lucas Maystre, and Mounia Lalmas. 2021. Shifting Consumption towards Diverse Content on Music Streaming Platforms. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (Virtual Event, Israel) *(WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 238–246.

AO Hirshman. 1964. The paternity of an index. 54, 5 (1964), 761–762.

David Holtz, Ben Carterette, Praveen Chandar, Zahra Nazari, Henriette Cramer, and Sinan Aral. 2020. The Engagement-Diversity Connection: Evidence from a Field Experiment on Spotify. In *Proceedings of the 21st ACM Conference on Economics and Computation* (Virtual Event, Hungary) *(EC '20)*. Association for Computing Machinery, New York, NY, USA, 75–76.

Geoffrey G Parker, Marshall W Van Alstyne, and Sangeet Paul Choudary. 2016. *Platform revolution: How networked markets are transforming the economy and how to make them work for you.* WW Norton & Company.

Christian Peukert and Imke Reimers. 2019. Digital disintermediation and efficiency in the market for ideas. *Available at SSRN 3110105* (2019).

Jean-Charles Rochet and Jean Tirole. 2003. Platform competition in two-sided markets. *Journal of the european economic association* 1, 4 (2003), 990–1029.

Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.

Zijun Shi, Kaifu Zhang, and Kannan Srinivasan. 2019. Freemium as an optimal strategy for market dominant firms. *Marketing Science* 38, 1 (2019), 150–169.

Ruth Towse. 2020. Dealing with digital: the economic organisation of streamed music. *Media, Culture & Society* 42, 7-8 (2020), 1461–1478.

Isaac Waller and Ashton Anderson. 2019. Generalists and Specialists: Using Community Embeddings to Quantify Activity Diversity in Online Platforms. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1954–1964.

Yanhui Wu and Feng Zhu. 2022. Competition, contracts, and creativity: Evidence from novel writing in a platform market. *Management Science, forthcoming* (2022).

Dainis Zegners. 2017. Building an Online Reputation with Free Content: Evidence from the E-book Market. (2017).

A APPENDIX

A.1 Goodreads ratings data and its representativeness

A limitation of collecting ratings data on Goodreads.com is that the full set of historical ratings are not visible at the same time. To circumvent this, we utilize a detailed algorithm of data collection as follows:

(1) Navigate to a book URL from the provided list.
(2) Scroll to "community reviews"
(3) Capture the following aggregate fields: Avg rating, total ratings, total reviews (ie ratings with text)
(4) Click on "more filters" and capture the rating distribution
(5) Check if the total number of ratings is <=300. If so, no need to apply additional filters. Just scrape the 10 pages of reviews displayed.
(6) If total number of ratings >300 but <=1500, we should use star rating filters:
  - "Go to more filters"
  - Click on a star rating
  - Scrape 10 pages of reviews as in Step 5.
(7) Go back to Step 6b to specify the next star rating. Repeat this for 1,2,3,4 and 5 stars one at a time. Thus, max num of pages to scrape = 5 star ratings levels * 10 pages = 50; max num of reviews = 50*30 = 1500
(8) If total number of >1500, we should use star rating AND order filters:
  - "Go to more filters"
  - Click on a star rating
  - Click on "sort order"→"oldest first"
  - Scrape 10 pages of reviews
  - Click on "sort order"→"newest first"
  - Scrape 10 pages of reviews as in Step 4
  - Go back to Step 6b to specify the next star rating. Repeat this for 1,2,3,4 and 5 stars one at a time.

Goodreads does enable us to see the summary statistics of submitted ratings, i.e, the absolute number of 1,2,3,4 and 5 stars that a book has accumulated (Figure 10). By comparing these to the distribution of ratings captured by our scraper, we can determine the representativeness of our sample. We eventually find qualitatively similar numeric ratings in our context, as shown in Figure 11.

A.2 Publisher Level Analysis

We have 16326 publishers in the raw data. Since the publisher information is self-reported on Giveaways, books may have different publisher information even if they are published by the same publisher. For example, two books published by Random House may have listed their publisher as "Random House" or "random house" (cases differ), or Alan Simon Books may have listed their publisher as "Alan Simon Books" or "Alan Simon". Therefore, we performed the following process to clean the publisher variable so that books published by the same publisher can be clubbed together. The following pre-processing steps were involved:

- Lower the case
- Remove words that may be omitted in same cases while exist for other cases. For example, books published by lulu.com may have the publisher information as "lulu" or "lulu.com". We remove words/phrases like "publishing", "group", "books", ".com" etc.

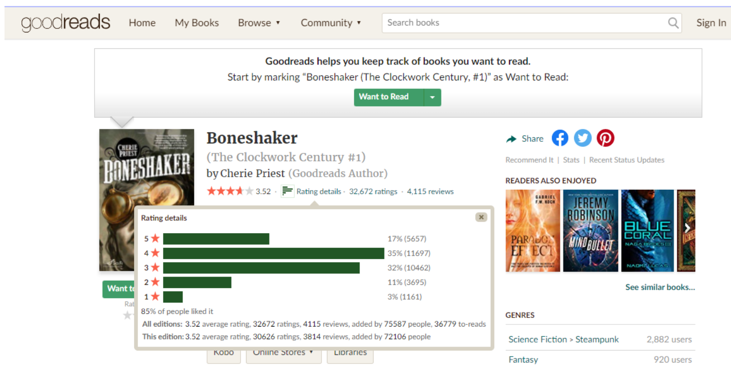

Fig. 10. Summary statistics available from Goodreads on review distributions of a given book.

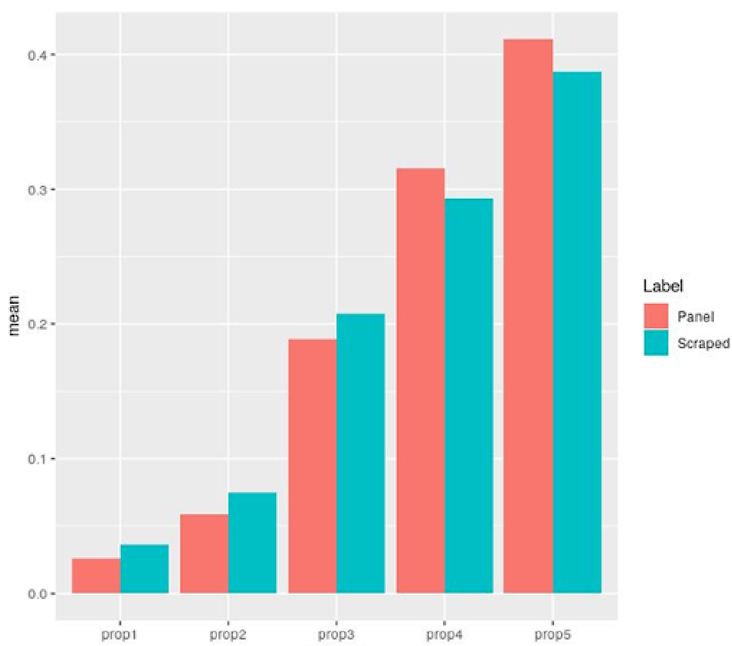

Fig. 11. Comparison of rating proportions obtained for each star level from the scraped data vs the Goodreads 'panel' of summary statistics. We see that the true distribution very closely tracks our collected sample.

- Remove punctuations, white space (some inputs have more than one white space between words)

The above cleaning process leaves us with 15322 publishers.

Moreover, Big 5 publishing houses have several subsidiaries under their umbrellas. To identify if a given publisher in our dataset belongs to the Big 5, we do the following:

- Collect all the divisions/imprints of the Big Five Publishing Houses (321 in total).
- Preprocess the division/imprint information as before: lower the case, remove potentially omitted words, remove punctuation/white space.

- Create a dummy equal to 1 if a publisher is one of the 321 imprints/divisions.

Self-publishing entities are identified and tagged in a similar way based on a manual assessment of the book market.[13]

A.2.1 Self-publishing proportion post-monetization. Here, we examine how monetization impacts publishers' participation on Giveaway, in particular among self-publishing and non-self-publishing houses. We estimate regressions of the form:

$$y_{it} = \beta_0 + \beta_1 \times \text{Post-Monetization}_t + \beta_2 \times \text{Self-publishing}_i + \beta_3 \times \text{Post-Monetization}_t \times \text{Self-publishing}_i + \epsilon_{it}, \quad (6)$$

with all variables defined as before. Consistent with what we found in Figure 6, the result of Table 10 Model (1) shows that the publishers that provide self-publishing books participated less as they decreased the monthly total number of books on Giveaway by 33. Moreover, the result of Model (2) indicates that these self-publishing service providing publishers, compared to other publish houses, dropped more books proportionally on Giveaway as the their proportion decreased ($\beta = -0.005, p < 0.01$) after monetization.

Table 10. Impact on Self-publishing Books

	Dependent variable:	
	No. of Books	Proportion of Books
	(1)	(2)
Post Monetization	-0.116^{***}	0.00000
	(0.007)	(0.00000)
Self-Publishing	40.600^{***}	0.013^{***}
	(0.156)	(0.0001)
Post Monetization × Self-Publishing	-33.379^{***}	-0.005^{***}
	(0.223)	(0.0001)
Month Fixed Effect	Yes	Yes
Observations	720,134	720,134
R^2	0.089	0.056
Note:		*p<0.1; **p<0.05; ***p<0.01

A.3 Book Genre Analysis

We infer book genres based on the most popular shelves they were put on. We took up to the top two bookshelves for each book as its genres after some manual processing of the raw bookshelves data. The manual processing includes: 1. remove irrelevant shelves, e.g. currently-reading, kindle, and so on; 2. merge some bookshelves that have the same meaning but with different names, e.g. 'historical' and 'history', 'sci-fi' and 'science-fiction'. 3. Take the top 50 bookshelves across all books as the genres categories we will use.

[13]Drawing from this list, among others: https://en.wikipedia.org/wiki/List_of_self-publishing_companies

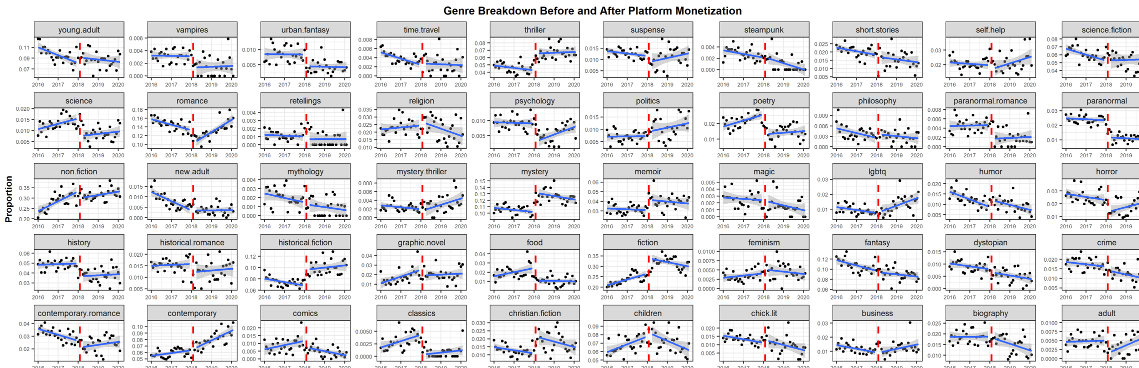

Fig. 12. Heterogeneous impact on book genres for all 50 genres

A.3.1 Entropy regressions. Here, we use a fixed effect regression model with standardized entropy as dependent variable. The regression includes an indicator for post monetization, which equals to 1 if the observation is after monetization and 0 otherwise. The regression also includes a linear term for time trend to account for common temporal trend across genres as well as month fixed effects to account for seasonality. The results are shown in Table 11. The estimated coefficient for post monetization indicator is negative and significant. It indicates a large reduction in genre diversity of books participating Giveaway program after monetization than before. Note that our model indicates that time trend account for part of the reduction in genre diversity. Nonetheless, the effect of monetization is still statistically significant and large in magnitude after accounting for that. Based on model estimates in column (2), genre diversity decreases by as large as 1.13 standard deviation after monetization.

$$\text{Entropy}_t = \beta_1 \times \text{Post Monetization}_t + \beta_2 \times \text{Time Trend}_t + \gamma_m + \epsilon_t \quad (7)$$

Table 11. Impact of Monetization on Genre Diversity

	Dependent variable:	
	Standardized Entropy	
	(1)	(2)
Post Monetization	-1.866^{***}	-1.134^{***}
	(0.134)	(0.280)
Time Trend		-0.028^{***}
		(0.010)
Month Fixed Effect	Yes	Yes
Observations	47	47
R^2	0.853	0.883
Note:	*p<0.1; **p<0.05; ***p<0.01	

A.4 Disaggregated rating analysis

Table 12. The effect of giveaway and monetization on average ratings, measured at the day level, full data.

	Dependent variable:			
	Avg Rating			
	(1)	(2)	(3)	(4)
Post Giveaway	−0.329***	−0.330***	−0.430***	−0.431***
	(0.010)	(0.010)	(0.008)	(0.008)
Time since release		0.002***		0.002***
		(0.0001)		(0.0001)
Post Giveaway × Post Monetization	−0.221***	−0.221***		
	(0.014)	(0.014)		
Number of books	81608	81608	81608	81608
Overall average rating	3.468	3.468	3.468	3.468
Book FEs	Yes	Yes	Yes	Yes
Year-month FEs	Yes	Yes	Yes	Yes
Observations	36,631,238	36,631,195	36,631,238	36,631,195
Adjusted R^2	0.166	0.166	0.165	0.165
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table 13. The impact of giveaway and monetization on avg ratings, measured at the year-month level, full data.

	Dependent variable:	
	Aggregated Avg Rating	
	(1)	(2)
Post Giveaway	−0.455***	−0.370***
	(0.005)	(0.006)
Post Giveaway × Post Monetization		−0.212***
		(0.010)
Number of books	81608	81608
Overall mean rating	3.58	3.58
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	2,544,641	2,544,641
Adjusted R^2	0.444	0.445
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 14. The effect of giveaway and monetization on review volume, measured at the year-month level, full data.

	Dependent variable:	
	Num. of Rating	
	(1)	(2)
Post Giveaway	3.037^{***}	1.341^{***}
	(0.150)	(0.156)
Post Giveaway × Post Monetization		4.402^{***}
		(0.268)
Number of books	81573	81573
Mean rating count	7.52	7.52
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	4,205,346	4,205,346
Adjusted R^2	0.153	0.153
Note:	*p<0.1; **p<0.05; ***p<0.01	