Optional Purchase Verification in E-commerce Platforms: More Representative Product Ratings and Higher-quality Reviews

(Authors' names blinded for peer review)

E-commerce platforms struggle to create and maintain high-quality reputation systems. One promising option is "purchase verification," which confirms that the user reviewing a product purchased the product from the platform. Previous works comparing platforms that require purchase verification with platforms that do not offer purchase verification found that review manipulation is easier in the latter. But what happens in platforms where purchase verification is optional? In such platforms, there is no monetary cost for posting fake reviews. Yet, optional purchase verification might introduce indirect costs for fake reviewers through expectation disconfirmation, hence positively affecting the reputation ecosystem of an e-commerce platform. To investigate, we use a quasi-experimental setup to analyze 336,043 book reviews. We find empirical evidence that introducing optional purchase verification reduces fake reviews, most of which are positive. This reduction of fake reviews results in lower, more representative product ratings and longer and more helpful reviews posted by more experienced reviewers. These new findings extend our understanding of how optional purchase verification can improve a platform's reputation ecosystem and suggest managerial interventions for platforms that have yet to develop a verification mechanism.

Key words: Online reputation ecosystems, Optional purchase verification, Quasi-experimental setup, Fake reviews

1. Introduction

The rapid growth of e-commerce marketplaces (e.g., Amazon) has triggered the offering of a large and diversified set of products (Statista 2018, Scrapehero 2018). To increase trust (Pavlou and Gefen 2004, Pavlou 2003, Awad and Ragowsky 2008) and reduce search cost (Bakos 2001) and information asymmetry (Akerlof 1978), e-commerce marketplaces have developed various reputation mechanisms based on user-generated content. These mechanisms, which include product ratings and reviews, significantly drive product sales (Bolton et al. 2004, Chevalier and Mayzlin 2006, Dellarocas 2003, Dellarocas et al. 2007, Kokkodis and Ipeirotis 2016, Gu et al. 2012, Chen and Lurie 2013, Reinstein and Snyder 2005, Chintagunta et al. 2010, Clemons et al. 2006, Han et al. 2018, Duan et al. 2008a,b, Hu et al. 2008, Cui et al. 2018, Archak et al. 2011, Zhao et al. 2013, Li et al. 2011, Lau et al. 2018). In fact, specific review characteristics have a stronger effect on sales than actual product information (Liu and Karahanna 2017), while even negative reviews can increase sales through product awareness (Berger et al. 2010).

Given their established economic value (Wu et al. 2015), it is not a surprise that sellers often submit fake reviews to benefit their products or hurt their competitors. Previous research has found both fraudulent positive and negative reviews in the hotel (Mayzlin et al. 2014) and restaurant (Luca and Zervas 2016) industries, estimating that as many as 16% of all reviews submitted are fake.

It is in the best economic interest of the platform to ensure that reviews are not only truthful but also high-quality. To this end, platforms have developed several mechanisms to reduce fake and low-quality reviews that might distort an item's reputation and mislead potential buyers (Lappas et al. 2016, Zhang et al. 2016, Luca and Zervas 2016). One such mechanism is purchase verification, which verifies that the user submitting the review has actually purchased the product on the platform, and, as a result, allows shoppers to identify reviews that are likely to be truthful (Kokkodis 2012). Many leading e-commerce platforms, including Amazon, Wayfair, and Target, now use purchase verification.

Despite this widespread use, the little research that has examined the effects of purchase verification on a platform's reputation ecosystem has focused primarily on situations where purchase verification is a prerequisite for reviewing (Mayzlin et al. 2014). However, for many platforms, verifying every purchase may be costly, noisy, or foreign to their business operations. Furthermore, requiring verification could hinder a platform's ability to quickly accumulate a critical mass of opinions and hence result in significantly lower review volume. After all, researchers have identified many legitimate reasons for users to review a product on a different platform than the one they used to purchase (Wasko et al. 2005, Bateman et al. 2011, Ren et al. 2012), and such non-verifiable truthful reviews can subsequently benefit both the reviewed products and the platform. In addition, since most users do not review (Hu et al. 2017), restricting truthful reviewers from participating by requiring purchase verification will result in the loss of valuable but already scarce sources of information.

These observations explain why many platforms (e.g., Amazon, Target) offer *optional purchase verification* (OPV), allowing both users who bought a product on the platform and those who did not to post a review. In these platforms, whether OPV decreases review manipulation remains an open question: absent a requirement to purchase, fraudulent reviewers can still post a review without an obvious cost. To investigate, we ask: *How does optional purchase verification affect the reputation ecosystem of an e-commerce platform?*

We argue that OPV will likely reduce fake reviews because of indirect costs related to the resulting disconfirmation between expected and realized product quality. Because e-commerce platforms offer products with no immediate substitutes (e.g., books), fake reviews are more likely to be overly positive (Luca and Zervas 2016). Hence, purchase verification will likely result in a reduction of net-positive fake reviews. This reduction will subsequently yield relatively lower, more representative product ratings, and longer, more descriptive reviews posted by more experienced reviewers.

Even further, this reduction of fake reviews will likely increase attention to longer truthful reviews, which will increase the average review helpfulness.

To test this theoretical framework we examine 336,043 book reviews from Amazon U.S. and Amazon U.K. The introduction of OPV at different times in the U.S. and U.K. markets forms a quasi-experimental setup that helps isolate the hypothesized effects. In line with our theory, we find that introducing purchase verification results in lower product ratings and reviews that are longer and are posted by more experienced reviewers. Empirical evidence further shows that OPV increases the overall review quality and review helpfulness. Extensive supportive analysis that includes (1) detection of reviews that Amazon has removed as fake, (2) deep learning models that identify fake reviews based on text, (3) models that estimate text authenticity, (4) comparison of review characteristics between fake and non-fake reviews, (5) evidence that positive fake reviews are followed by relatively negative verified reviews, and (6) analysis of the effect of OPV on the overall review volume provide substantial empirical support of the theorized mechanisms of action. Finally, an extended set of additional tests empirically eliminate alternative explanations that could be driving our observed results.

Our study extends the operations management literature on response bias and fake reviews (Chen et al. 2016, Kumar et al. 2018b, 2019, Lappas et al. 2016, Lee et al. 2018) by explaining how OPV affects both ratings and review characteristics in e-commerce platforms. By showing for the first time that introducing OPV has positive spillover effects on the overall reputation ecosystem of a platform, our findings can guide the business operations of multiple e-commerce and reputation platforms. Platforms that are already offering optional purchase verification can now better understand and measure their utility. Platforms that are currently operating without purchase verification (e.g., TripAdvisor, Yelp) have now encouraging empirical evidence that introducing optional purchase verification can have multi-dimensional positive effects on their reputation ecosystem.

2. Research context and hypotheses development

In the following paragraphs, we provide a brief background of the relevant work in user-generated content in digital and e-commerce platforms, and we hypothesize how the introduction of optional purchase verification can change review and reviewer characteristics.

2.1. Background

Recent studies in the interface of Operations Management-Information Systems (OM-IS) have answered the call to investigate user-generated content in digital platforms (Kumar et al. 2018c). Some focus on understanding and operationalizing user engagement in platforms that depend on content creation (e.g., Twitter, online communities, online reviews; Mallipeddi et al. 2021, Kokkodis et al. 2020, Kumar et al. 2018a). Others focus on the operational value of such content (Cui et al.

2018), and show that (1) proposing sellers' strategies that are consumer-specific yields better-quality reviews (Guan et al. 2020), (2) consumer and product characteristics moderate the effect of reviews on product choice (Ba et al. 2020), (3) scalable methods that use review text can identify and extract service dimensions (Mejia et al. 2020), and (4) building predictive models and performing big data analytics and sentiment analysis on user-generated content results in better sales predictions (Cui et al. 2018, Lau et al. 2018).

A stream of OM-IS research closely related to this work focuses on online review biases and manipulation. Specifically, to correct for reporting bias, Chen et al. (2016) model the generating process of online reviews through an inverse probability weighting scheme. Additional research uses machine learning to build meta-classifiers that identify deceptive reviews (Kumar et al. 2018b), and unsupervised hierarchical approaches to detect anomalies in reviewer behavior and potentially identify fake reviewers (Kumar et al. 2019). Finally, prior studies have also provided evidence of the existence of fake online reviews through text mining, sentiment, and econometric analyses (Lappas et al. 2016, Lee et al. 2018). Our work directly extends this line of OM-IS research by providing, for the first time, evidence that optional purchase verification is an effective mechanism against review manipulation, and by showing that its introduction has multi-dimensional positive spillover effects in the reputation ecosystem of a platform that yield more representative product ratings and longer and more helpful reviews posted by more experienced reviewers.

2.2. Hypothesis development

2.2.1. Combating fake reviews with purchase verification: Firms often have difficulty encouraging people to review products on online platforms (Chen et al. 2016). Some estimates suggest that only 0.1% of people actually review the products they purchase on Amazon (Hu et al. 2017). Because the vast majority of people who purchase a product do not provide reviews, sellers have the opportunity to manipulate the reputation of a product by injecting even modest numbers of fraudulent reviews (Lappas et al. 2016, Lee et al. 2018). Such fake reviews undermine the value of online reputation systems (Feng et al. 2012, Sussin and Thompson 2012). A considerable amount of review fraud originates from professional review-authoring companies that submit fake reviews on major review platforms in exchange for a fee (Lappas et al. 2016). The prevalence and success of these companies have even prompted government action that resulted in $350,000 in fines (Schneiderman 2014). Despite such ongoing efforts, many review manipulating companies are still active, prolific, and successful (Lappas et al. 2016).

To combat fraud and improve the quality of reviews, platforms have implemented various defense mechanisms. For instance, website moderators manually approve reviews on the TripAdvisor and Yelp platforms (Lappas et al. 2016). Suspicious reviews are placed on hold pending examination and can even be eliminated if the website's proprietary filtering process provides enough evidence. Businesses associated with fake reviews are penalized in the platform's rankings, excluded from press releases and top-10 lists, and may even have a relevant banner placed on their page.

Optional and required purchase verification offer another promising strategy for combating fake reviews and increasing trust in online reputation systems (Yuan et al. 2012). When required, purchase verification significantly reduces the number of fake reviews (Mayzlin et al. 2014). When optional, purchase verification reveals which of the available reviews are written by reviewers who have purchased the product on the focal platform, allowing users to distinguish between verified and non-verified reviews when determining their veracity (Kokkodis 2012, Anderson and Simester 2014, Kim et al. 2018b,a, He et al. 2020).

2.2.2. Types of fake reviews in e-commerce platforms: To understand how introducing OPV affects fake reviews in e-commerce platforms, we first need to better understand what types of fake reviews these platforms attract. In general, fraudulent reviews can either be (1) negative, intended to decrease the ratings of competitors' products, or (2) positive, intended to inflate a product's reputation (Mayzlin et al. 2014, Lappas et al. 2016, Lee et al. 2018). The type of fraudulent reviews depends considerably on the competitive nature of the product (Luca and Zervas 2016, Lee et al. 2018) or the market (Mayzlin et al. 2014). Fraudulent negative reviews are likely to be more common when competitor products or services are substitutes, as damaging the reputation of one's competitor product is likely to turn consumers to one's own product. For example, fraudulent negative reviews on hotels are more likely when competitors are geographically proximate and able to serve as a ready substitute. In contrast, fraudulent positive reviews are more likely when smaller hotels are trying to stand out and compete with larger competitors (Mayzlin et al. 2014).

In e-commerce settings where such immediate substitutes are improbable, fraudulent negative reviews are less likely to have the intended effect (Anderson and Simester 2014, Luca and Zervas 2016). However, fraudulent positive reviews are still likely to have an effect as products seek to differentiate themselves from and gain attention amidst the volume of other products available. The books domain, which is the focus of this study, is a characteristic example of a setting with a large number of competitive products, in which harming a competitor's reputation is unlikely to affect the sales of the fraudulent reviewer's book significantly. However, boosting one's own reputation can still be beneficial. In fact, a considerable number of authors on Amazon positively review their books to increase sales (Smith 2004).

2.2.3. Disconfirmation-induced reduction of fake reviews: Even though requiring purchase verification is a successful defense mechanism against fake reviews (Mayzlin et al. 2014), it is not immediately clear why *optional* purchase verification would discourage fake reviews in an e-commerce platform. When purchase verification is required, it introduces economic disincentives for creating fake reviews (Mayzlin et al. 2014). Since optional verification does not create such monetary disincentives for submitting fake reviews, *why would fake reviewers change their behavior after the introduction of OPV?* In fact, one could argue that, because optional purchase verification would naturally decrease the value of non-verified reviews (Anderson and Simester 2014), introducing OPV could create an incentive for fake reviewers to submit *more* fake reviews to compensate for the likely reduced influence of non-verified reviews on purchase decisions.

Response bias (i.e., who chooses to review a product) may provide some rationale of why introducing OPV could in fact reduce the number of fake reviews. In particular, underreporting bias (Hu et al. 2017) recognizes that the vast majority of people do not review products online. People who choose to review often do so because their realized experience with the bought product did not match the rosy picture characterized by the positively skewed reviews (expectation disconfirmation; Ho et al. 2017, Kokkodis and Lappas 2020). Introducing optional purchase verification can increase the value of such expectation disconfirmation incidents: if sellers choose to artificially inflate a product's reputation by submitting non-verified fake positive reviews, then—due to underreporting bias and the fact that not every buyer posts a review—they will also increase the likelihood of reviewers who experience disconfirmation to post negative reviews. As a result, attempts to inflate a product's reputation positively through fake non-verified reviews might consequently result in verified negative reviews from customers who bought the product, experienced it, and found discrepancies between their experience and the posted reviews. Given that verified reviews are relatively more valuable than non-verified ones (Anderson and Simester 2014), such negative verified reviews will hurt the product significantly more. Hence, and even though introducing optional purchase verification does not induce an extra monetary cost to sellers who post fake reviews, we argue that *OPV will end up reducing the amount of fake reviews* due to this indirect disconfirmation-induced cost.

2.2.4. Optional purchase verification effects on the reputation ecosystem: The reduction of fake reviews effectively alters the reputation ecosystem (i.e., product ratings, review text, review helpfulness) of an e-commerce platform. Since most fake reviews in such a platform are positive (Section 2.2.2), the disconfirmation-induced reduction of fake reviews suggests that products that have been artificially boosted through positive fake reviews will experience a drop in their average rating after the introduction of OPV. For instance, a product that receives a constant rate of positive fake reviews per year will experience a rate decrease after the introduction of OPV. Hence, all else being equal, this reduction of positive fake reviews will result in a relative *decrease* in product ratings:

HYPOTHESIS 1. *Introducing optional purchase verification in an e-commerce platform will result in lower, more representative product ratings.*

Besides product ratings, the reduction of fake reviews also alters the average textual characteristics of the posted reviews. In particular, fake—and paid—reviews tend to be shorter than truthful reviews (Li et al. 2014, Burtch et al. 2017). This is likely because truthful reviewers describe their genuine experiences, while fake reviewers need to generate and describe fake experiences (Li et al. 2014). Hence, due to the reduction of fake reviews that tend to be shorter, introducing OPV should all else being equal result in relatively longer product reviews.

HYPOTHESIS 2. *Introducing optional purchase verification in an e-commerce platform will result in longer product reviews.*

Nonverbal reviewer characteristics are highly predictive of fake reviews (Zhang et al. 2016). In fact, fake reviews are often posted by reviewers who review less and for shorter periods (lower experience) than truthful reviewers (Mukherjee et al. 2013). As a result, the reduction of such fake reviews due to the introduction of OPV will likely lead to a corpus of reviews posted by relatively more experienced reviewers:

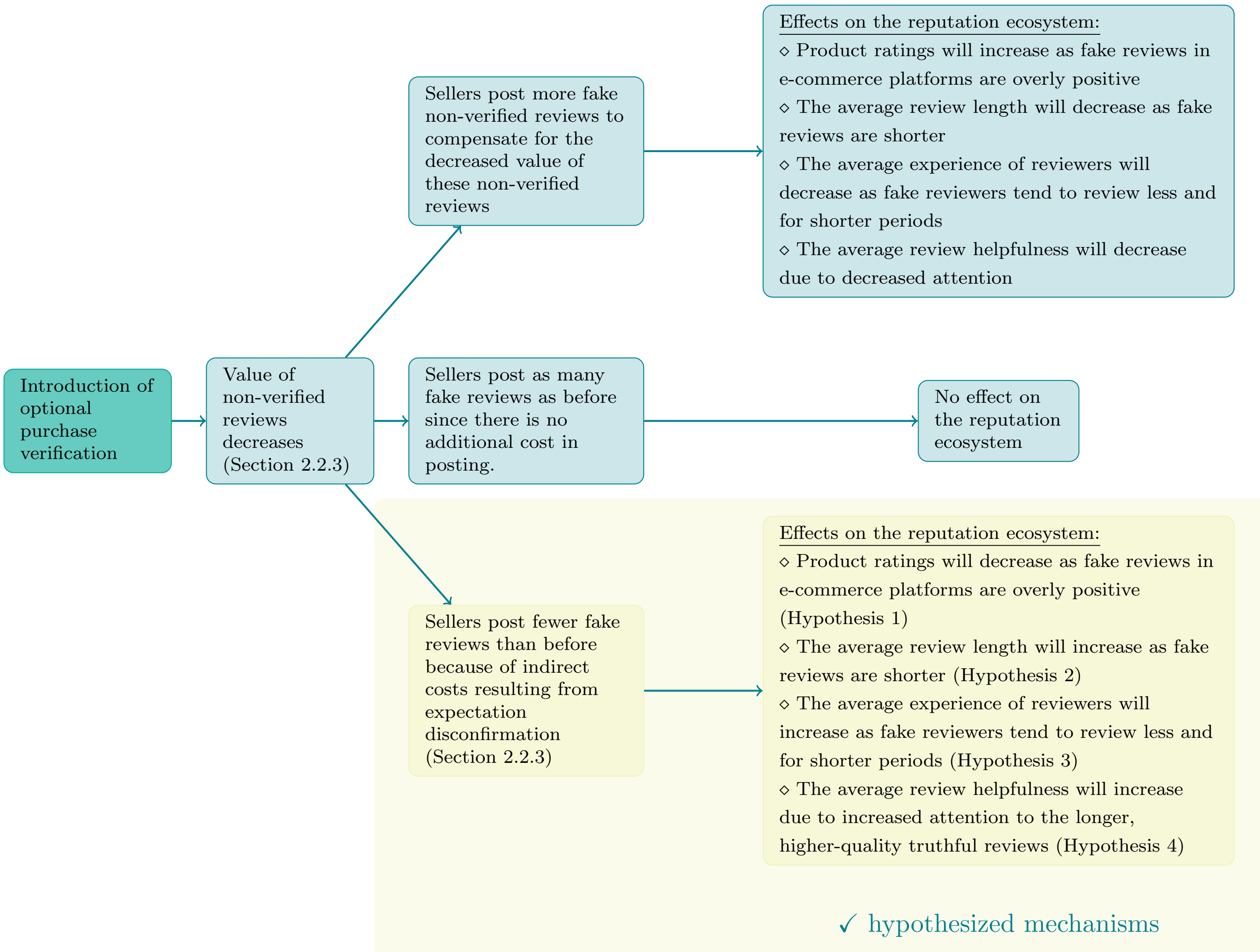
HYPOTHESIS 3. *Introducing optional purchase verification in an e-commerce platform will result in reviews written by more experienced reviewers.*

Finally, the reduced rate of fraudulent reviews may actually make it easier for consumers to pay attention to the truthful ones that remain (Hansen and Haas 2001, Eppler and Mengis 2004). Given that truthful reviews are often longer than fake ones (Li et al. 2014, Burtch et al. 2017), and since longer reviews tend to be more helpful (Mudambi and Schuff 2010), introducing OPV should also indirectly increase review helpfulness, as consumers can more easily pay attention to these longer, truthful reviews. Thus, we argue that introducing OPV will raise the overall quality of the reviews by reducing shorter fake reviews, which in turn will increase attention to the remaining longer, genuine, and more helpful reviews:

HYPOTHESIS 4. *Introducing optional purchase verification in an e-commerce platform will result in more helpful reviews.*

Figure 1 summarizes the hypothesized mechanisms of the expected effects of introducing optional purchase verification in an e-commerce platform, while it also illustrates how alternative mechanisms would result in different effects. Next, we discuss the empirical research setting that facilitates the investigation of these mechanisms and the testing of Hypotheses 1 to 4. (Sections 5.2 and 6.3 discuss alternative mechanisms that might overlapping with our primary theory presented in Figure 1.)

Figure 1 Overview of the hypothesized mechanisms



The Figure summarizes the hypothesized mechanisms of the expected effects of introducing optional purchase verification. It also shows how alternative underlying mechanisms would have resulted in different effects.

3. Research setting

Before describing our research setting and dataset, we formally define the focal problem as follows:

<u>Problem definition:</u> *Consider an e-commerce platform that does not offer any type of purchase verification. Our goal is to estimate the average effect of introducing optional purchase verification on product ratings and review characteristics.*

It is important to highlight that we do not focus on comparing verified with non-verified reviews—an interesting topic studied by Anderson and Simester (2014) and by Kim et al. (2018a,b). Instead,

Figure 2 **An example of an Amazon verified review header**

we focus on how the availability of OPV affects the reputation ecosystem of a platform, which includes both verified and non-verified reviews.

3.1. Quasi-experimental setup

We use data from Amazon, "*the largest internet-based retailer in the world by total sales and market capitalization*" (Barney 2014). Amazon allows users to submit reviews and rate products on a rating scale from one to five stars. To curate a high-quality reputation system, Amazon offers OPV: "*If you bought the product you are reviewing from Amazon.com, you [can] label the review as an Amazon Verified Purchase. The Amazon Verified Purchase label offers Amazon.com customers additional context and helps them better gauge the quality and relevance of a product review.*" (Amazon 2009, 2018). Figure 2 shows the header of a verified review. Figure 3 shows the header of a non-verified review.

The theoretical framework in Figure 1 argues that introducing OPV will affect product ratings and review characteristics through the reduction of fake reviews. To empirically test this framework, we design a quasi-experimental setup that uses the global dimension of Amazon and the seemingly exogenous (with respect to product ratings and review characteristics) timing differences in the introduction of OPV in different markets. In particular, Amazon U.S. introduced OPV in September 2009 (Amazon 2018). During the same period, Amazon U.K. was operating without purchase verification. Amazon U.K. finally introduced OPV in March 2012 (Amazon 2012). Amazon uses the same unique product identifiers (ASIN) across its platforms globally. This consistency allows for studying how product ratings and review characteristics *of the same products* evolve in these two different markets (U.S. vs. U.K.), before and after the introduction of OPV. This setup creates a quasi-experimental design (Adamopoulos et al. 2020, Huang et al. 2017, Chen et al. 2018, Chevalier and Mayzlin 2006, Mayzlin et al. 2014), where we assume that the introduction of the VP badge is exogenous to product ratings, review length, helpfulness, and reviewer experience.[1] Figure 4 visualizes this quasi-experimental setup and the hypothesized (Hypothesis 1) effects of introducing OPV on assigned ratings. Figure 11 in E-Companion J shows the actual visualization of the parallel trend of the assigned ratings as derived from our data.

[1] A similar setup that compares timing variations of different treatments across different Amazon markets was used in Adamopoulos et al. (2020).

Figure 3 An example of an Amazon non-verified review header

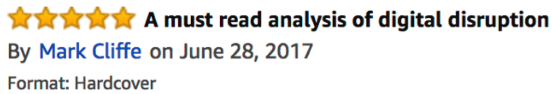
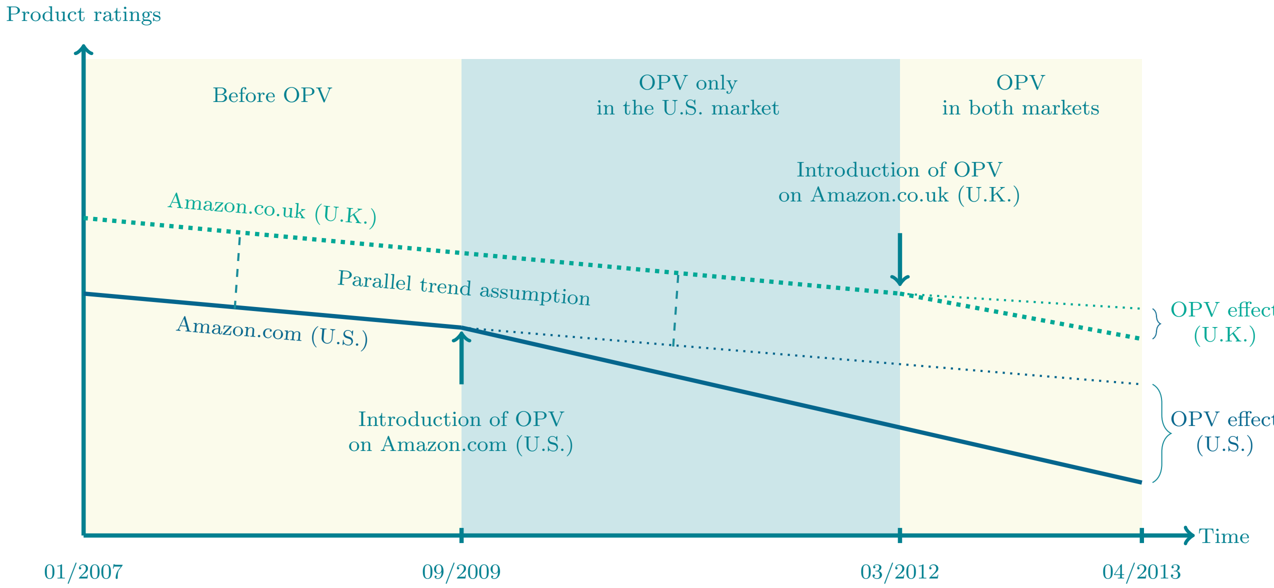


Figure 4 The focal quasi-experimental setting



The diagram shows a conceptualization of the hypothesized effect on product ratings. (Figure 11 in E-Companion J shows the actual parallel trend from our data). We consider 6.3 years of reviews. Before 09/2009, none of the markets were offering any type of purchase verification. Between 09/2009 and 03/2012, only Amazon U.S. was offering optional purchase verification (OPV). After 03/2012, both markets were offering OPV.

3.2. Focal, dependent, and control variables

Our analysis focuses on products that were available in both the U.K. and the U.S. markets before and after the introduction of OPV in the two markets. Specifically, we analyze 336,043 book reviews posted between 01/2007 and 04/2013 in the two markets. *All products in our data have reviews both before and after* the introduction of OPV.

3.2.1. Focal variable: We capture the focal treatment effect through the binary variable "OPV Introduction" that describes whether or not a review is posted after the introduction of OPV (Figure 4). Note that the treatment applies to a market as a whole. Hence, once OPV is introduced, its treatment applies to both verified and non-verified reviews.

3.2.2. Dependent variables: Hypotheses 1 to 4 require four dependent variables: assigned rating, review length, reviewer experience, and review helpful votes. The assigned rating and the review helpful votes are extracted directly from the posted reviews. The number of words in a review defines review length. The total number of a reviewer's posted reviews defines reviewer experience.

3.2.3. Control variables: To better isolate the effect of the introduction of OPV, we control for various observed and unobserved confounding factors. Product fixed effects eliminate the unobserved time-invariant effects of product peculiarities; market fixed effects control for unobserved systematic differences between the U.S. and the U.K.; Bi-weekly time fixed effects account for time trends.

Various *time-varying* covariates control for the observed population heterogeneity that affects our dependent variables. First, the time since the first posted review for each product ("Product tenure") controls for any tenure-specific effects. The current average of prior product ratings ("Accumulated product rating (simple average)") before the submission of a new review controls for trends in the perceived product quality as well as for effects of previously posted ratings (Moe and Trusov 2011). Similarly, the standard deviation between already assigned ratings ("Ratings deviation") captures observed disagreements within a product's reviewer population at the time of a new review. The total number of posted reviews ("Accumulated product reviews") right before submission controls for the observed popularity of each product.

Amazon does not estimate product ratings by simply averaging all assigned ratings.[2] Instead, it uses machine learning models that evaluate multiple criteria to estimate product ratings. Hence, to control for possible discrepancies between the "Accumulated product rating (simple average)" and the machine-learning predicted product rating of Amazon's that could correlate with our dependent variables we reverse-engineer Amazon's algorithm through predictive modeling—we discuss the details of this process in E-Companion E. The result is captured in the variable "Accumulated product rating (reverse engineered)") that measures the expected Amazon's algorithm rating at each point in time. (Note that our dependent variable "Assigned rating" is not affected by Amazon's algorithm, as it directly measures the user-assigned ratings.)

Another characteristic that might affect the process of posting reviews is the within-product review rankings. At each point in time, consumers might be intrigued to post a review depending on what reviews are being displayed on the product front page. To control for this variability, we build predictive models that estimate the likelihood of each review to be displayed on the front page ("Likelihood of top-ranked review"). E-Companion F presents the details of this approach.

Finally, when estimating the effect of introducing OPV on review helpful votes, we control for both the assigned rating and the review length. Both of these measures are visible when a review receives helpful votes, and, as we know from prior studies (Kim et al. 2006, McCallum et al. 1998, Yin et al. 2014), they significantly correlate with the perceived helpfulness of a review.

Table 1 shows the descriptive statistics of the focal, dependent, and control variables. Figure 12 in E-Companion J shows their correlations. Because we perform our analysis at the product level,

[2] `https://www.amazon.com/gp/help/customer/display.html?nodeId=GQUXAMY73JFRVJHE`

Table 1 Descriptive statistics of the dependent, focal, and control variables

		Mean	Median	StD	Min	Max
Focal variable	OPV introduction	0.46	0	0.5	0	1
Dependent variables	Assigned rating	4.2	5	1.1	1	5
	Review length	120	76	138	1	4173
	Reviewer experience	95	23	331	1	12490
	Review helpful votes	3.1	1	13	0	906
Time-varying control variables	Product tenure (bi-weekly)	73	70	48	0	162
	Accumulated product rating (simple average)	4.2	4.3	0.56	1	5
	Likelihood of top-ranked review	0.33	0.28	0.18	0	1
	Ratings deviation	0.96	1	0.39	0	2
	Verified reviews	0.27	0	0.43	0	1
	Accumulated product rating (reverse engineered)	4.3	4.4	0.3	2.8	4.8
	Review length	120	76	138	1	4173
	Accumulated product reviews	69	41	84	0	838
Fixed effects	Time	163 bi-weekly periods				
	Market	2 markets (U.K., U.S.)				
	Product	2,035 books				

336,043 book reviews over a period of 6.3 years (between 01/2007 and 04/2013) aggregated in 180,359 bi-weekly observations. The data comes from two markets (U.S. and U.K.). All products appear in both markets before and after the introduction of OPV. Descriptive statistics show bi-weekly aggregates.

we aggregate variables in bi-weekly observations.[3] Sensitivity analyses show results for monthly (Table 16), weekly (Table 17), and non-aggregated observations (Table 18).

4. Empirical evaluation

A difference-in-differences (DID) specification can empirically isolate the effect of introducing optional purchase verification on product ratings and review characteristics (Wooldridge 2010). For a product p at market m and time t we can estimate the effect of the introduction of OPV as follows:

$$\mathrm{DV}_{pmt} = P + M + T + \gamma OPV_{mt} + \boldsymbol{\beta X}_{pmt} + \varepsilon_{pmt}\,, \tag{1}$$

where $\mathrm{DV}_{pmt} \in$ {Assigned rating, Review length, Review helpful votes, Reviewer experience}, P, M, and T are the fixed effects of product, market (U.S., U.K.), and time (bi-weekly periods—see Tables 16, 17, and 18 for monthly, weekly, and non-aggregated observations), and $\boldsymbol{X}_{pmt}$ is the vector of the observed time-varying variables and the constant term. The coefficient of interest γ, is the difference-in-difference estimate of the effect of introducing OPV on the dependent variable.

[3] Aggregations are useful in our context as they help decrease potential popularity effects of products that receive significantly larger-than-average amounts of reviews.

Table 2 Effects of the introduction of OPV on product ratings and review characteristics

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.06^{*}		2.41		3.68		-0.93^{**}
		(0.03)		(2.24)		(8.32)		(0.36)
Accumulated product rating (simple average)		0.17^{***}		-4.32^{***}		-3.06		0.13
		(0.01)		(0.83)		(3.19)		(0.11)
Ratings deviation		0.01		-2.73^{***}		-5.60^{*}		-0.46^{***}
		(0.01)		(0.61)		(2.34)		(0.09)
Accumulated product reviews		0.00		-0.79		-4.36^{**}		-0.31^{**}
		(0.01)		(0.72)		(1.56)		(0.10)
Verified reviews		0.14^{***}		-5.80^{***}		-23.30^{***}		0.18^{***}
		(0.01)		(0.42)		(1.27)		(0.04)
Likelihood of top-ranked review		0.16^{***}		83.93^{***}		36.26^{***}		-0.12
		(0.01)		(0.46)		(1.25)		(0.09)
Accumulated product rating (reverse engineered)		-0.06^{***}		3.02^{**}		0.32		-0.11
		(0.01)		(1.09)		(2.88)		(0.16)
Review length								2.06^{***}
								(0.15)
Assigned rating								-1.31^{***}
								(0.06)
OPV introduction	-0.07^{***}	-0.24^{***}	13.52^{***}	8.00^{***}	16.07^{***}	37.76^{***}	1.71^{***}	1.25^{***}
	(0.01)	(0.01)	(1.30)	(1.20)	(3.62)	(4.08)	(0.14)	(0.14)
Observations	180359	180359	180359	180359	180359	180359	180359	180359
R^2	0.01	0.04	0.11	0.43	0.01	0.02	0.02	0.05

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

If $\gamma > 0$, then the introduction of OPV had an overall positive effect on assigned ratings, length, helpful votes, and reviewer experience. Otherwise, if $\gamma < 0$, the introduction of OPV had a negative effect on these dependent variables.

Our setting quasi-simulates a randomized control trial under the assumption that the introduction of OPV is exogenous to assigned ratings and review characteristics—an assumption that aligns with Amazon's announcement (Amazon 2009) and we also empirically test in E-Companion B.2. Furthermore, if our theoretical mechanism holds and OPV reduces fake reviews (most of which are positive), then the decision to introduce OPV is likely uncorrelated with any omitted variables that drive assigned ratings, review length, review helpful votes, and reviewer experience. Finally, as we empirically show in Section 5.1 and E-Companion B, it is unlikely that a trend in assigned rating, review length, review helpful votes, and reviewer experience triggered Amazon's decision to introduce OPV.

4.1. Results

Table 2 shows the results. For each of the four dependent variables we estimate two specifications: columns A1, B1, C1, and D1 show the effects of the introduction of OPV by only controlling for time, market, and product fixed effects. Columns A2, B2, C2, and D2 show the effects of the introduction of OPV under the complete specification of Equation 1. All columns support Hypotheses 1 to 4: Introducing optional purchase verification results in lower average product ratings (Columns A1-A2) and in longer reviews posted by more experienced reviewers that end up being more helpful (Columns B1-D2).

To estimate the magnitude of these effects, we use the following formula:

$$\text{OPV-driven percentage change (\%)} \in [\frac{\gamma}{\min(DV)} * 100, \frac{\gamma}{\max(DV)} * 100] \,.$$

By considering all specifications, we find that the OPV-driven percentage change on assigned rating ranges $\in$ [-24%, -1.4%]. Similarly, the effect on review length ranges $\in$ [0.2%,1,352%], on reviewer experience $\in$ [0.1%,3,776%], and on review helpful votes $\in$ $[0.1,\infty)$. The OPV-driven percentage changes on the mean of each dependent variable (based on the estimates of the complete specification of Equation 1) are -5.7% (Assigned rating), 6.7% (Review length), 39.7% (Reviewer experience), and 40.3% (Review helpful votes).

4.2. Empirical investigation of the underlying mechanisms

The empirical analysis (Table 2) supports Hypotheses 1 to 4. However, it does not provide evidence in support of an underlying mechanism that drives the results. Our theory argues that introducing optional purchase verification should reduce the amount of fake reviews, the majority of which tend to be positive in an e-commerce platform (Figure 1). To empirically investigate this mechanism, we first need to identify which reviews are likely to be fake. As a result, we perform the following analyses:

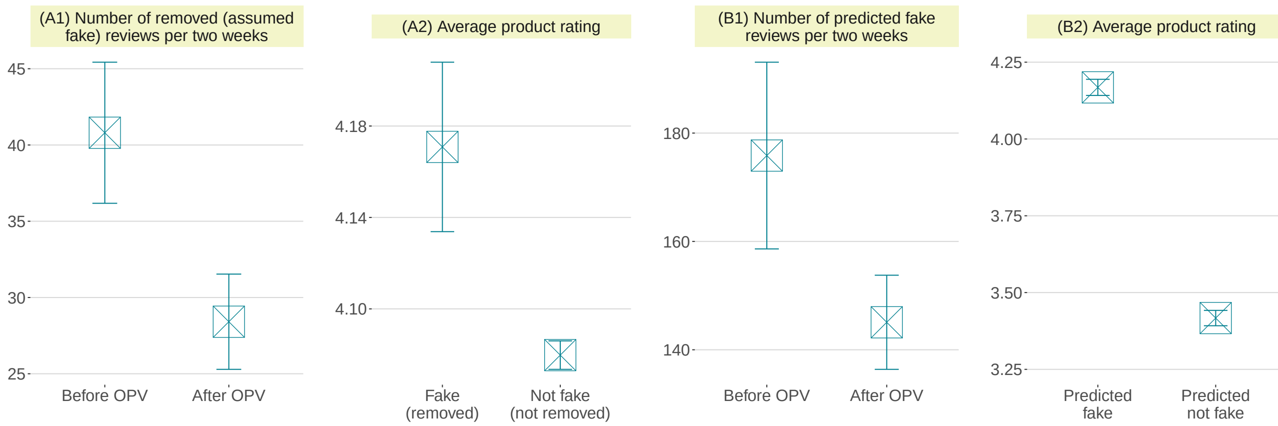
Detection of fake reviews: Over time, Amazon has been identifying and removing fake reviews (Wehner 2016). Since our dataset was collected between 2010 and 2013, we can identify the reviews that Amazon has since removed, and we can label them as "fake."

Prediction of fake reviews: We can build predictive models that estimate the probability of a review to be fake by using a unique, labeled Amazon reviews dataset for deception detection (Saxena 2018).

Text authenticity: We can estimate the text authenticity of each review through the LIWC framework (LIWC 2018) before and after the introduction of OPV.

We discuss the details and the results of these three approaches next.

Figure 5 **After the introduction of OPV, the platform observes fewer fake reviews which tend to be positive**

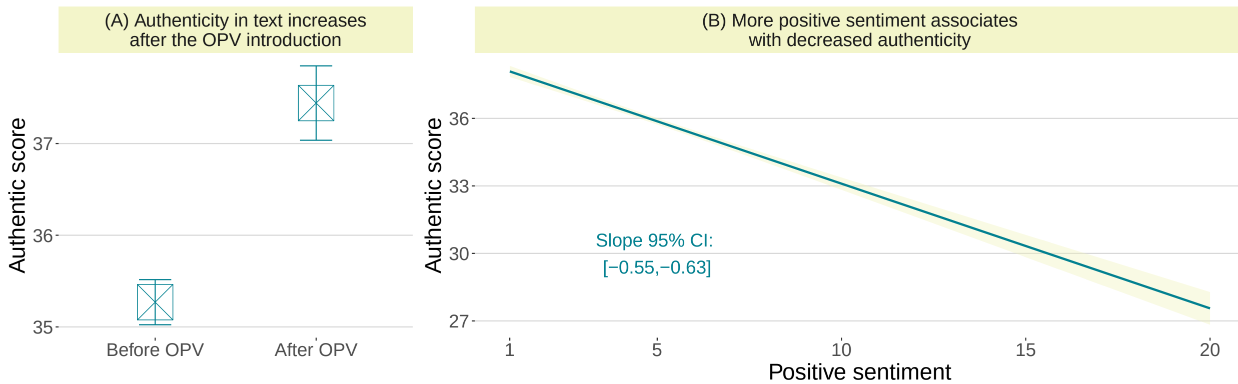


Figures A1 and A2 assume as fake reviews those that Amazon subsequently removed (Section 4.2.1). Figures B1 and B2 assume as fake reviews the ones annotated by predictive models (Section 4.2.2). Error bars show 95% confidence intervals.

4.2.1. Detection of fake reviews. We collected our main dataset between 2010 and 2013. Since then, Amazon has been allocating significant resources towards identifying and removing fake reviews (Wehner 2016). Hence, a straightforward way to identify fake reviews is to find out which reviews from our corpus has Amazon since deleted. Amazon removed 3,863 reviews, which we assume to be fake. Figure 5A1 shows that the bi-weekly frequency of appearance of fake reviews was higher before the introduction of OPV ($p < 0.001$). This indicates that proportionally more fake reviews were posted before the introduction of OPV. In addition, Figure 5A2 shows that the average product rating of the deleted (fake) reviews was higher ($p < 0.001$) than the product rating of the non-deleted ones, hence showing that the removed fake reviews were indeed more positive than the non-fake ones. As a result, if we assume that the removed reviews by Amazon were fake, then these fake reviews were, on average, more positive than non-fake reviews, and they were found more frequently before the introduction of OPV. This evidence provides empirical support for our main hypothesized mechanism described in Figure 1.

4.2.2. Prediction of fake reviews. Predicting whether or not a review is fake is an inherently difficult task (Luca and Zervas 2016, Kumar et al. 2018b, 2019). One of the main difficulties originates from the lack of ground truth data. Recently, researchers have published a dataset of labeled Amazon reviews for deception detection (Saxena 2018). The dataset includes 21,000 Amazon reviews, of which 10,500 are labeled as "fake." We use this data to learn models that predict the likelihood of a review to be fake according to the review text. We discuss this process in E-Companion G. Figures 5B1 and 5B2 show the results in terms of the number of fake reviews before and after the introduction of OPV and in terms of product ratings for fake and non-fake reviews.

Figure 6 The introduction of OPV correlates with more authentic and less positive reviews



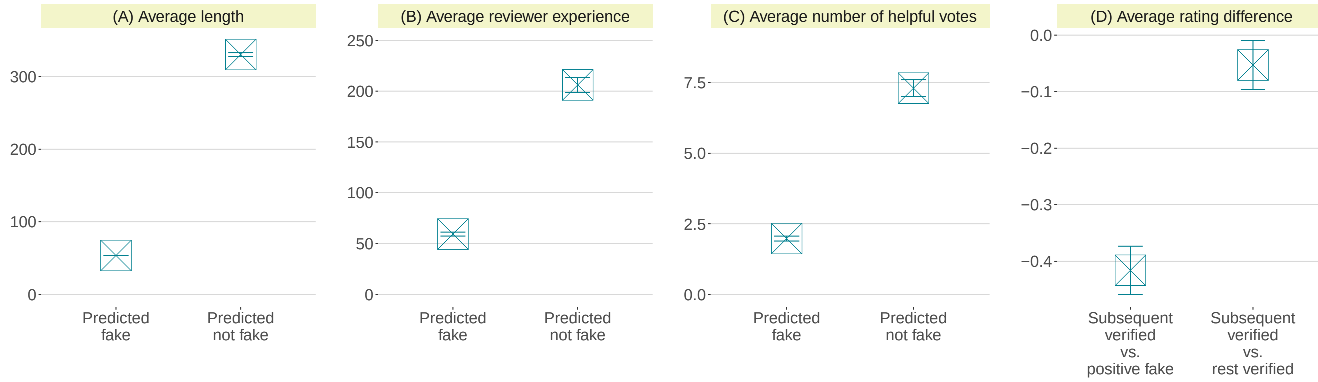
The review text was less authentic before the introduction of OPV (A). Furthermore, there is a negative association between text authenticity and positive review sentiment (B). Error bars show 95% confidence intervals. CI: Confidence Interval.

The results provide additional support to the hypothesized mechanism: before the introduction of OPV, the prediction task identified more fake reviews. In addition, the predicted fake reviews were more positive than the non-fake ones.

4.2.3. Text authenticity. Next, we focus on the raw review text. Recent developments in text analysis allow to detect concepts related to text authenticity. Specifically, we use the LIWC package to analyze the unstructured review text (LIWC 2018). LIWC adopts a dictionary-based approach and has been successfully used to estimate emotionality in various contexts (Sridhar and Srinivasan 2012, Hong et al. 2016, Goes et al. 2014, Yin et al. 2014). To capture authenticity, the LIWC package estimates the "Authentic" metric: "higher numbers are associated with a more honest, personal, and disclosing text; lower numbers suggest a more guarded, distanced form of discourse" (LIWC 2018). Figure 6A shows that after the introduction of OPV, the "Authentic" scores increased significantly ($p < 0.001$). In addition, Figure 6B shows that authenticity has a negative relationship ($p < 0.001$) with the review sentiment in the focal dataset. Hence, the two Figures provide further support to the main theoretical mechanism presented in Figure 1.

4.2.4. Fake reviews are shorter, posted by less-experienced reviewers, and are less helpful: Beyond the main mechanism that focuses on the reduction of positive fake reviews, our theoretical framework further claims that fake reviews are likely shorter, written by less experienced reviewers, and, as a result, are less helpful. Similar to before, we analyze these characteristics of a subset of reviews that our predictive model labels as fake and as truthful (E-Companion G). Figure 7 shows the results of this analysis. Predicted fake reviews are significantly ($p < 0.001$) shorter than predicted non-fake reviews (Figure 7A). Furthermore, the average reviewer experience of reviews predicted to be truthful is almost four times larger than that of reviewers who posted

Figure 7 Empirical evidence shows that fake reviews are shorter, less helpful, and posted by less-experienced reviewers. Positive fake reviews are followed by relatively more negative verified reviews.



Predicted fake reviews are shorter (A), posted by less-experienced reviewers (B), and they accumulate fewer helpful votes (C). Subsequent verified reviews are significantly more negative than (1) positive fake reviews and (2) the rest verified reviews of the focal product. Subsequent verified reviews include verified reviews posted up to three months after a positive fake review. Rest verified reviews include all other verified reviews of the product. Positive fake reviews are reviews predicted to be fake that assigned ratings greater than 3. Error bars show 95% confidence intervals.

fake reviews. Finally, on average, predicted truthful reviews are 2.5 times more helpful than predicted fake ones. As a result, these observations provide empirical support of the mechanisms that structure Hypotheses 2, 3 and 4.

4.2.5. Disconfirmation-driven negative verified reviews: Perhaps the hardest component of our mechanism to test is that the introduction of OPV added an indirect cost on fake reviews due to expectation disconfirmation (Section 2.2.3 and Figure 1). Conceptually, if expectation-disconfirmation indeed adds an indirect cost on posting fake positive reviews, then we should be able to observe in the data relatively more negative verified reviews following positive fake reviews. Simply put, positive fake reviews at time t should be followed by relatively more verified negative reviews at times $t+1$, $t+2$, etc. To test whether we observe this pattern in our data, we estimate for each product the average rating of the verified reviews posted within three months after a positive predicted fake review. We name these reviews as *"subsequent verified."* Then, we compare these ratings with the product-specific ratings of (1) the predicted positive fake reviews (i.e., fake reviews with a product rating greater than 3), and (2) all verified reviews excluding the *subsequent verified* ones ("rest verified"). For our theorised disconfirmation mechanism to hold, the subsequent verified ratings should be significantly lower than both the predicted fake ratings and the rest verified ones. Figure 7D shows exactly this relationship: subsequent verified ratings are on average 0.053 ($p < 0.05$) more negative compared with the rest verified reviews, and 0.41 ($p < 0.001$) more negative compared with the positive fake reviews.

Table 3 The volume of (mostly fake) reviews decreases after the introduction of OPV

	Volume		
	(A1)	(A2)	(A3)
Time (bi-weekly) FE	✓	✓	✓
Product FE	✓	✓	✓
Market FE	✓	✓	✓
Product tenure		-0.22^{**}	-0.23^{**}
		(0.09)	(0.09)
Accumulated product rating (simple average)		0.10^{***}	0.10^{***}
		(0.02)	(0.02)
Review length		-0.08^{***}	-0.08^{***}
		(0.01)	(0.01)
Assigned rating		0.08^{***}	0.08^{***}
		(0.01)	(0.01)
Accumulated product reviews		0.00	0.00
		(0.06)	(0.06)
Ratings deviation		0.06^{***}	0.06^{***}
		(0.02)	(0.02)
Verified reviews		0.06^{***}	0.06^{***}
		(0.01)	(0.01)
Likelihood of top-ranked review		-0.01^{*}	-0.02^{**}
		(0.01)	(0.01)
Accumulated product rating (reverse engineered)		0.06	0.06
		(0.05)	(0.05)
OPV introduction	-0.47^{***}	-0.51^{***}	-0.49^{***}
	(0.03)	(0.03)	(0.03)
OPV introduction × Fake review			-0.66^{***}
			(0.03)
Observations	180359	180359	180359
R^2	0.05	0.05	0.06

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p<0.001$, ** $p<0.01$, * $p<0.05$).

4.2.6. The effect of the introduction of OPV on the overall volume of reviews: Our hypothesized mechanism is also connected to the overall review volume. In particular, our theory implies that since the introduction of OPV reduces the number of fake reviews, it will all else being equal subsequently reduce the overall review volume. Table 3 tests this expected effect. The dependent variable in this regression is the bi-weekly aggregated received reviews per product. Similar to Table 2, we test specifications with and without the vector of time-varying control variables. Columns (A1) and (A2) show the results: as expected, the introduction of OPV reduced the product-specific volume of reviews.

On top of this analysis, we use the predicted probabilities of a review to be fake to estimate whether the observed effect is more pronounced on fake reviews. Based on the analysis in E-Companion G, we create a binary variable "Fake review" that describes whether a given review is predicted to be fake. Then, we interact this variable with the introduction of OPV and re-run the full specification of Equation 1 to see the combined effect on review volume. Conceptually, the coefficient of the interaction variable shows how much stronger the OPV effect is when the reviews are predicted to be fake. Column A3 shows the results: the negative coefficient of the interaction term suggests that the observed review volume reduction is more pronounced on likely fake reviews, therefore further supporting the hypothesized mechanisms in Figure 1.

Overall, six different independent tests provide empirical support for the proposed theory (Figure 1): introducing optional purchase verification reduces the amount of fake reviews, the majority of which tend to be positive in e-commerce platforms. As we discuss in Section 6.3, even though none of these tests is perfect, the fact that all point in the same direction is encouraging that our hypothesized theoretical mechanisms are indeed driving the observed results. Next, we empirically test and eliminate a set of alternative explanations.

5. Alternative explanations

Our study makes a series of assumptions regarding the focal quasi-experimental setting. At the same time, the analysis in Section 4.2 does not rule out alternative mechanisms that could also be driving the observed results. To provide additional empirical support for our theory, in this Section we:

- Test the validity of our quasi-experimental setup through various statistical tests and subsample analysis (Section 5.1, E-Companion B).
- Empirically test and eliminate alternative theoretical mechanisms that could partially explain the observed results (Section 5.2, E-Companion C).
- Test the sensitivity of the observed results under various groupings over time as well as through log-transformations of the dependent variables and alternative models (E-Companion H).

Table 4 summarizes these analyses, which we discuss in detail in the following paragraphs and the E-Companion.

5.1. Correctness of the quasi-experimental setup

5.1.1. Market-specific trends: A test that can provide some confidence in support of our DID identification strategy adds market-specific time trends to the list of controls (p.238, Angrist and Pischke 2008, Besley and Burgess 2004, Adamopoulos et al. 2020). Specifically, in our context, we estimate the following specification:

$$\mathrm{DV}_{pmt} = P + M + \mu * t + \gamma OPV_{mt} + \boldsymbol{\beta X}_{pmt} + \varepsilon_{pmt} \;, \tag{2}$$

Table 4 Investigation of alternative explanations

Analysis	H1	H2	H3	H4	Discussion
Alternative explanation: Imperfections of the quasi-experimental DID setting drive the results					
Solution: Test the assumptions of DID:					
Market-specific trends do not drive the results	✓	✓	✓	✓	Section 5.1.1, Table 5
Coincidental events do not drive the results (placebo regressions)	✓	✓	✓	✓	Section 5.1.2, Table 6
Granger causality holds (relative time regressions)	✓	✓	✓	✠	E-C B.1 Figure 10
the introduction of OPV was likely exogenous to the four dependent variables	✓	✓	✓	✓	E-C B.2, Table 13
Alternative explanation: Contamination through global accessibility of the U.S. reviews drives the results					
Solution: Subsample analysis:					
Analysis of U.S.-only reviews	✓	✓	✓	NS	Section 5.1.3, Table 7
Random selection of products (appear in one market)	✓	✓	✓	✓	Section 5.1.3, Table 8
Nearest neighbor matching of products	✓	✓	✓	✓	Section 5.1.3, Table 9
Products with early post-treatment U.K. reviews	✓	✓	✓	✓	Section 5.1.3, Table 10
Alternative explanation: Concurrent events drive the results					
Solution: Fixed effects on concurrent events	✓	✓	✓	✓	Section 5.2.1, Table 11
Alternative explanation: Systematic differences between the two markets drive the results					
Solution:					
Market and time FE control for unobserved static and within period systematic differences	✓	✓	✓	✓	Sections 4, 5.1.1, Tables 2, 5
No empirical evidence of imbalance between the number of fake reviews of the two markets	✓	✓	✓	✓	Section 5.2.2, Figure 8A
Differences in publication dates across the two markets do not drive the results	✓	✓	✓	✓	Section 5.2.3, Table 12
Alternative explanation: Fake reviewers changed their behavior					
Solution: Empirical evidence shows that fake reviewers did not change their behavior	NA	✓	NA	✓	Section 5.2.4, Figure 8B and Figure 8C
Alternative explanation: New restrictions on verified reviews drive the results					
Solution: Empirical evidence shows that these restrictions do not drive the results	✓	✓	✓	✓	Section 5.2.5, Figure 8D
Alternative explanation: Regular reviewers changed their behavior					
Solution: Empirical evidence shows that regular reviewers did not change their behavior	✓	✓	✓	✓	Section 5.2.6, Figure 9
Alternative explanation: Results are driven by products of different quality					
Solution: Empirical evidence shows that results hold across products of varying qualities, with some products being more affected than others	✓	✓	✓	✓	E-C C, Figure 14
Alternative explanation: Results are driven by the chosen aggregation thresholds and models					
Solution: Alternative aggregations and models:					
Regressions on weekly, monthly, and non-aggregated data	✓	✓	✓	✓	E-C H, 16, Tables 17, 18
Logarithmic transformations	✓	✓	✓	✓	E-C H, Table 19
Negative binomial regressions	✓	✓	✓	✓	E-C H, Table 20

✓: Hypothesis supported, or relevant assumption holds. ✠: There is some pre-trend significance, but post-trend effects are much stronger (Figure 10). NS: The effect of the introduction of OPV on review helpful votes is not statistically significant ($p > 0.05$). NA: Alternative explanation does not apply to the column dependent variable. E-C: E-Companion where μ is a market-specific dummy and t is the time trend. This specification allows treated (U.S.) and control (U.K.) markets to follow different trends, hence capturing any systematic time-changing differences between the two markets that could affect the results. "*It is heartening to find that the estimated effects of interest are unchanged by the inclusion of these trends, and discouraging otherwise*" (p.238, Angrist and Pischke 2008). Table 5 shows that the main results are robust to the inclusion of market-specific time trends. Despite controlling for market-specific trends, introducing OPV still yields lower product ratings, and longer and more helpful reviews by more experienced reviewers, hence suggesting that any market-specific trends do not drive our observed results (Angrist and Pischke 2008).

5.1.2. Placebo regressions: A different test runs placebo regressions that reveal whether the DID specifications pick up any coincidental spurious entry effects (Chan and Ghose 2014). Specifically, in our scenario, we randomly sample a period before introducing optional purchase verification in the U.S., and we assign a placebo the introduction of OPV in the middle of the sampled period. Table 6 in E-Companion A shows the results. For all dependent variables, the coefficients of the placebo OPV are not statistically significantly different than zero. Hence, the placebo regressions provide empirical evidence that the observed effects are not coincidental.

5.1.3. Potential contamination from browsing: A valid concern about our quasi-experimental setup is that there might be contamination across treatment and control markets. For instance, U.K. users can browse reviews on U.S. websites and, as a result, be affected by the introduction of OPV (treatment). To empirically test for this possibility, we:

- Estimate fixed-effect panel regressions for the U.S. market alone, before and after the introduction of OPV.
- Randomly select products so that they form independent sets across markets. Through this randomization, each product appears only in a single market.
- Match products based on their observed characteristics through a nearest neighbor approach. Then, we randomly choose matched pairs so that each product appears only in a single market. Through this matching, we can control for matched-pairs fixed effects across both markets (by replacing the product fixed effects of our main specification with matched-pairs fixed effects).
- Identify a group of products that had reviews in the U.K. before a review was posted in the U.S. in the post-treatment period.

Because the first approach focuses only on the U.S. (treated) market, contamination from browsing cannot exist. In the next two approaches, we expect contamination effects to be absorbed in the product fixed effects, as each product now only appears in one of the two markets. Finally, in our last test, because all reviews in the U.S. were posted in the pre-treatment period, contamination might only come from other products. Tables 7, 8, 9, and 10 in E-Companion A show the results. In all four scenarios, despite the significant subsampling, the observed effects remain qualitatively unchanged. The only exception is the effect of the introduction of OPV on review helpful votes on Table 7, which is positive but insignificant.

5.1.4. Additional tests of our quasi-experimental setup: E-Companion B provides additional tests (relative time regressions, visualization of the parallel trend, exogeneity of the introduction of OPV) that show additional support in favor of the correctness of the DID specification.

5.2. Empirical elimination of alternative mechanisms

Despite the presented empirical evidence in support of the theoretical framework, alternative mechanisms could work in parallel to partially drive the main results.

5.2.1. Concurrent events: A valid concern is that our specification of Equation 1 does not control for other concurrent events that cannot be potentially fully captured through the market and time fixed effects and that could have affected the studied dependent variables. To test for this possibility, we capture all relevant concurrent Amazon events—as listed in the History of Amazon Wikipedia page[4]—through dummy variables. In particular, we create dummies for the (1) launch of Amazon Kindle (November 2007), (2) introduction of Nook (October 2009), (3) introduction of iBook (January 2010), (4) launch of Azure (February 2010), (5) launch of Amazon video (February 2011), (6) collection of sales taxes in California (July 2011), (7) announcement of Kindle Fire (September 2011), (8) launch of Amazon Appstore (March 2011), (9) acquisition of Kiva systems (March 2012), (10) collection of sales taxes in Nevada and Texas (April 2012), and (11) announcement of Kindle Fire HD (September 2012). Table 11 shows the results, which are aligned with the main results presented in Table 2. Overall, these external events do not affect the hypothesized and observed effect of the introduction of OPV on the assigned review ratings, length, helpfulness, and observed reviewer experience.

5.2.2. Imbalance of fake reviews between the two markets: An alternative explanation argues that the two markets are systematically different in the way that they attract fake reviews (i.e., one market attracts more fake reviews than the other). Figure 8A shows that this is likely not the case. The two markets do not significantly differ in terms of predicted fake reviews (E-Companion G), as shown by the box plots that are not statistically significantly different.

5.2.3. Differences in publication dates: An alternative argument suggests that there might be a systematic lag of the book publication dates between the two markets (e.g., the U.S. market publishes books earlier than the U.K. market). In general, it might be more likely to post fake

[4] https://en.wikipedia.org/wiki/History_of_Amazon

reviews in the early period of a publication. Hence, the existence of a time-varying systematic difference that is not captured by the market fixed effects could be driving the results.

The market-specific trends analysis we discussed in Section 5.1.1 and the fact that all included products have reviews in all periods presented in Figure 4 are encouraging first steps that any time-varying systematic differences in publication dates do not drive our results.

Yet, by looking into our data, we observe a systematic bias in terms of review arrivals: 50% of U.S. books receive reviews 33.5 (or more) days earlier than U.K. books. To test whether this systematic difference drives our results, we randomly under-sampled books from the U.S. so that the average difference between publication days of the U.S. and the U.K. is zero. The results in Table 12 align with our main results, suggesting that this systematic difference is likely captured in the market fixed effects and does not drive our observed results.

5.2.4. Change in fake reviews: An alternative explanation argues that fake reviewers adapted their behavior after the introduction of OPV, such that they started providing longer and hence more helpful reviews. Figures 8B and 8C show that this is not the case: there is no significant difference between the length and the helpfulness of predicted fake reviews before and after the introduction of OPV.

5.2.5. Amazon restricts the number of posted non-verified reviews per week: After the introduction of OPV, Amazon started restricting the number of non-verified reviews that a person can post on the platform to five per week—it is unclear whether Amazon implemented this policy before or after 2016.[5] To test whether this decision affects our results, we check (1) whether there is a significant change in the average number of weekly non-verified reviews that a reviewer posts before and after the introduction of OPV, and (2) whether we have reviewers in our data that have posted more than 5 reviews after the introduction of OPV. Figure 8D shows the results: there is no statistically significant change in the number of posted non-verified reviews per week. Interestingly, multiple reviewers in our data posted more than five non-verified reviews after the introduction of OPV, suggesting that the aforementioned policy was likely not in place during our data collection period.

5.2.6. Change of regular-reviewer behavior: Finally, an alternative mechanism argues that optional purchase verification changed reviewer behavior, such that they started writing fewer, lengthier, and more critical reviews. As a result, the observed effects are not driven exclusively by the reduction of positive fake reviews.

To test this possibility, we create a dataset that captures reviewer behavior before and after the introduction of OPV. Practical constraints do not allow studying the behavior of all reviewers

[5] `https://sellerengine.com/amazon-verified-product-reviews/`

in our dataset before and after the introduction of purchase verification. Such an analysis would require us to collect $\sim$ 29,432,096 reviews. Instead, to investigate this alternative mechanism, we randomly choose 1,000 reviewers from our dataset who have posted reviews both before and after the introduction of OPV. Because they post multiple reviews over time, we name these reviewers as "regular reviewers."

By collecting the complete reviewing histories of these regular reviewers, we create a separate dataset of 20,462 reviews posted between 09/2008 and 09/2010 (1 year before and 1 year after the introduction of OPV in the U.S.). We then generate bi-weekly aggregations of reviewer behavior regarding the number of reviews posted, average product ratings, and the average length of posted reviews. Finally, we check whether there are significant differences in regular-reviewer behavior before and after the introduction of purchase verification.

Figure 9 shows the results. In Figure A, we find no evidence that reviewers post more reviews after the introduction of OPV. Similarly, in Figure B, there is a statistically insignificant *increase* in the average product ratings ($p > 0.05$). Finally, Figure 9C shows that reviewers do not post longer reviews after the introduction of OPV ($p > 0.05$). Overall, empirical evidence suggests that regular-reviewer behavior remains unchanged after the introduction of purchase verification. Combined with the fact that the volume of reviews decreases after the introduction of OPV (Section 4.2.6), this unchanged reviewer behavior suggests that the observed reduction in reviews comes indeed from our hypothesized decreased rate of fake reviews.

6. Contributions to research and practice

We explored the effects of introducing optional purchase verification on the reputation ecosystem of an e-commerce platform. Theoretically, we argued that, despite the non-obvious additional cost of posting fake reviews, introducing optional purchase verification will likely reduce fake reviews because of indirect costs related to disconfirmation between expected and realized product quality. Using a quasi-experimental setup and a large set of product ratings and reviews from Amazon, we found that, indeed, introducing optional purchase verification reduces the amount of fake reviews, hence resulting in lower product ratings and longer and more helpful reviews written by more experienced reviewers.

6.1. Contributions to research in the OM-IS interface

By investigating these novel effects of optional purchase verification, our research extends the work on user-generated content and digital platforms in the OM-IS interface (Kumar et al. 2018c,a, Mallipeddi et al. 2021, Kokkodis et al. 2020, Cui et al. 2018, Guan et al. 2020, Ba et al. 2020, Mejia et al. 2020, Lau et al. 2018), and, in particular, OM-IS research on response bias and fake reviews (Chen et al. 2016, Kumar et al. 2018b, 2019, Lappas et al. 2016, Lee et al. 2018). Closely relevant research to ours compared platforms that require verified reviews (e.g., Expedia) with platforms that do not offer any type of purchase verification (e.g., TripAdvisor) to find that fake reviews are more prevalent in platforms without purchase verification (Mayzlin et al. 2014). Other relevant works focused only on the period after the introduction of optional purchase verification and compared non-verified with verified reviews to find that the latter tend to be more positive (Anderson and Simester 2014) and to have a stronger effect on sales (Kim et al. 2018a,b, He et al. 2020). Our study is unique, as it is the first to investigate how *introducing* optional purchase verification (i.e., studying both the before and after periods) alters product ratings and review characteristics in e-commerce platforms.

Furthermore, no prior study has connected optional purchase verification with product ratings, review length, reviewer experience, and review helpfulness. Our research is the first to suggest that introducing purchase verification associates with a decrease in product ratings and an increase in review length, review helpfulness, and reviewer experience. Through a comprehensive empirical analysis, our work shows that these observations are a direct result of reducing fake reviews, the majority of which are positive. This new mechanism of action is relatively counter-intuitive, as optional purchase verification does not exclude non-verified reviews and hence it does not add a monetary cost in posting fake reviews. However, it creates an environment where non-verified reviews have less value than verified ones. This implicit cost discourages fake reviewers from posting non-verified reviews, as such non-verified reviews might trigger negative verified reviews due to expectation disconfirmation. As a result, our work is the first to explain why optional purchase verification will likely benefit all aspects of the reputation ecosystem of an e-commerce platform.

Aside from these main research contributions, our analyses on identifying fake reviews provides new empirical evidence that, indeed, the majority of fake reviews in platforms where products do not have immediate substitutes are positive. This observation complements previous findings that fake reviews can be both positive and negative in environments where local competition exists (Mayzlin et al. 2014), and provides additional context on when review manipulation tends to be positive (Lee et al. 2018).

Interestingly, our work also shows that the most significant impact of introducing optional purchase verification applies to reviewers who eventually do not post a review (discouraged fake reviewers). This result can inform and guide future work in e-commerce and other types of online platforms to consider not only how design choices affect people who use the platform but also how they affect those who may choose not to use the platform to begin with.

In terms of empirical methodology, our study uses a new quasi-experimental setup that allows the comparison of the state of the two platforms before and after the introduction of OPV. This setup helps towards isolating the hypothesized effects, while it also overcomes a significant limitation that hinders efforts to quantify the effect of defense mechanisms against fake reviews: the lack of a deterministic way of establishing if a given review is fake or not. Finally, by using deep learning and predictive modeling analytics to identify fake reviews and estimate product ratings and review rankings (Sections 4.2 and 5, E-Companion E, F, and G), we also contribute to the call for applications of machine learning and AI in business operations (Kumar et al. 2018c, Cui et al. 2018, Kumar et al. 2018b, Geva and Saar-Tsechansky 2021).

6.2. Contributions to practice

Given the documented benefits of allowing more customers to provide reviews (Wang et al. 2019), the results of this work have practical implications for the business operations of e-commerce platforms that either already host both verified and non-verified online reviews or are considering introducing optional purchase verification. First, we showed for the first time that OPV is still an effective mechanism for reducing fraudulent reviews, even though it does not require purchase verification to review. Because of this reduction of overly positive fake reviews, the average product rating decreases and likely becomes more representative of the true product quality. In addition, optional purchase verification resulted in higher-quality textual reviews, without significantly reducing the total number of good quality (truthful) reviews, which is an unavoidable cost of platforms that introduce required purchase verification.

The positive effects of optional purchase verification extend beyond the ones studied in this paper. A better reviewing environment that includes OPV allows for efficient ranking algorithms to increase attention on high-quality reviews. For instance, Amazon now uses OPV as a feature to algorithmically rank reviews and products.[6] As a result, by explicitly (through the reduction of fake reviews) and implicitly (through ranking algorithms that use OPV as a feature) eliminating information overload, optional purchase verification cannot only allow platforms to develop higher-quality reputation systems, but also generate fruitful conditions for platform operations to provide relevant incentives and attract new high-quality reviewers.

Finally, our work can guide platforms that currently do not offer any type of purchase verification (e.g., Yelp or TripAdvisor). Even though in these platforms the effects of OPV on product ratings might not be as negative—since these platforms on expectation receive more fake negative reviews than an e-commerce platform (Mayzlin et al. 2014)—the overall reputation ecosystem will likely improve by hosting more truthful, representative reviews. Hence, by showing that requiring purchase verification is not necessarily the only way to curate high-quality reputation systems, our work can motivate these platforms to allow optional purchase verification (e.g., through electronic receipt submission). More importantly, our work is the first show that such an action will succeed

[6] https://www.amazon.com/gp/help/customer/display.html?nodeId=GQUXAMY73JFRVJHE

in reducing fake reviews through the discussed disconfirmation-induced cost. Even if very few users end up submitting verified reviews, the disconfirmation-induced effects can have significant positive spillovers in the overall reputation ecosystems of these platforms.

6.3. Generalizability, limitations, and future directions

Generalizability: Our analysis focuses on book reviews. We chose books because a book's lifetime is significantly longer than products from other categories such as electronics. However, our hypothesized mechanism is product-type independent. Hence, we expect the mechanism to generalize to any type of product that attracts relatively more positive than negative fake reviews (Luca and Zervas 2016, Lee et al. 2018). An analysis of a small set of electronics reviews in Table 21 of E-Companion I further supports this argument.

Limitations: Despite the quasi-experimental setup, and although we have taken careful steps to provide empirical support to our theoretical arguments, our study relies on observational data and, as a result, has empirical limitations. For example, unobserved events that could be interfering with the results might not be fully captured by our time fixed effects and other control variables. Yet, the strong evidence (Sections 4.2 and 5) supporting our theory suggests that, at least, it is one of potentially multiple causal paths that drive the observed effects.

For instance, we argue that the introduction of OPV reduced the value of fake reviews. But other mechanisms could also be contributing to the decreased value of non-verified reviews:

- Increased ability to identify fake reviews: The OPV introduction could have improved Amazon's ability to identify and remove fake reviews because OPV could be a useful predictive feature in Amazon's fake-review detection algorithms. Even though generally possible, we argue that this mechanism likely does not affect our data and results. Specifically, our data collection happened between 2010 and 2013, while the first evidence that Amazon started removing fake reviews comes from 2016 (Wehner 2016). In fact, our analysis in Section 4.2.2 identified thousands of removed fake reviews. Together, these observations suggest that Amazon was not actively detecting and removing fake reviews during our collection period.
- Effects of the product ranking algorithm: Both Amazon's review and product ranking algorithms use OPV as a predictive feature.[7][8] Hence, fake reviewers could have realized that their reviews are not as effective as they used to since Amazon's algorithms started discounting them after the OPV introduction. Even though possible, the variables "Likelihood of top-ranked review" and "Verified reviews" in our analysis capture a significant portion of such anticipated ranking effects. Furthermore, similar to before, it is unclear if (and to what extend) Amazon was using these algorithms during our collection window between 2010 and 2013.

[7] https://sellerengine.com/amazon-verified-product-reviews/

[8] https://amzn.to/2QApUPe

Note that the catalyst in these two alternative scenarios is the same: the introduction of OPV improved the platform's ability to build better ranking models and restrict reviews and reviewers. A different type of study could examine whether these mechanisms interact with our main mechanism—a study that would survey businesses (perhaps anonymously) of how they alter their behavior in the presence of OPV.

A different type of limitation originates from the error propagation of our two predictive models described in E-Companion E and F. We believe that the benefits of using these algorithms outweigh these prediction errors, as they allow us to control for portions of the effects of Amazon's algorithms that it would have been impossible to control for otherwise. Similarly, the prediction of fake reviews described in E-Companion G is not perfect. Furthermore, the additional fake-detection approaches we use in Sections 4.2.1 and 4.2.3 also have limitations. However, the fact that all three independent and very different approaches support our hypothesized mechanism of Figure 1 increases our confidence that this mechanism drives the observed results.

As we discussed in Section 5.1.3, some U.K. customers could have been treated by browsing reviews on the U.S. platform which had introduced optional purchase verification first. Four different tests in Section 5.1.3 provide encouraging evidence that this effect is limited and does not drive our observed results. It is also important to mention that even if this contamination exists in a way not captured by our analysis, then the observed effects are likely underestimated: potential U.K. fake reviewers would have realized that U.K. customers can browse the U.S. website and get treated with the OPV, and hence they would have likely deterred to post fake reviews, decreasing the amount of fake reviews in the pretreatment period of the U.K. platform.

6.4. Conclusion

In conclusion, this paper is the first to study how introducing optional purchase verification affects the reputation system of an e-commerce platform. The presented empirical analysis of Amazon books showed evidence that introducing optional purchase verification reduces the number of fake reviews, most of which tend to be positive. This reduction of fake reviews results in lower, more representative product ratings, as well as in longer and more helpful reviews posted by more experienced reviewers. As a result, introducing optional purchase verification creates a better, more credible reviewing environment for both reviewers and consumers.

References

Adamopoulos, P., V. Todri, A. Ghose. 2020. Demand effects of the internet-of-things sales channel: Evidence from automating the purchase process. *Information Systems Research*, 32 (1), 238–267.

Akerlof, G. A. 1978. The market for lemons: Quality uncertainty and the market mechanism. In *Uncertainty in Economics*, pp. 235–251. Elsevier.

Amazon 2009. Verified purchase reviews. https://bit.ly/339qViU. [Online; accessed 10-March-2020].

Amazon 2012. Amazon verified purchase. https://bit.ly/2TSYvpn. [Online; accessed 10-March-2020].

Amazon 2018. About amazon verified purchase reviews. https://goo.gl/mcFnzb. [Online; accessed 10-March-2020].

Anderson, E. T., D. I. Simester. 2014. Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, 51 (3), 249–269.

Angrist, J. D., J.-S. Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

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

Awad, N. F., A. Ragowsky. 2008. Establishing trust in electronic commerce through online word of mouth: An examination across genders. *Journal of Management Information Systems*, 24 (4), 101–121.

Ba, S., Y. Jin, X. Li, X. Lu. 2020. One size fits all? the differential impact of online reviews and coupons. *Production and Operations Management*, 29 (10), 2403–2424.

Bakos, Y. 2001. The emerging landscape for retail e-commerce. *Journal of Economic Perspectives*, 15 (1), 69–80.

Barney, J. 2014. Amazon urges california referendum on online tax. https://on.ft.com/2TVE3nD. [Online; accessed 10-March-2020].

Bateman, P., P. Gray, B. Butler. 2011. The impact of community commitment on participation in online communities. *Information Systems Research*, 22 (4), 841–854.

Berger, J., A. T. Sorensen, S. J. Rasmussen. 2010. Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science*, 29 (5), 815–827.

Besley, T., R. Burgess. 2004. Can labor regulation hinder economic performance? evidence from india. *The Quarterly Journal of Economics*, 119 (1), 91–134.

Bolton, G. E., E. Katok, A. Ockenfels. 2004. How effective are electronic reputation mechanisms? An experimental investigation. *Management Science*, 50 (11), 1587–1602.

Burtch, G., Y. Hong, R. Bapna, V. Griskevicius. 2017. Stimulating online reviews by combining financial incentives and social norms. *Management Science*, 64 (5), 2065–2082.

Chan, J., A. Ghose. 2014. Internet's dirty secret: Assessing the impact of online intermediaries on hiv transmission. *MIS Quarterly*, 38 (4), 955–976.

Chen, H., Z. Zheng, Y. Ceran. 2016. De-biasing the reporting bias in social media analytics. *Production and Operations Management*, 25 (5), 849–865.

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

Chen, Z., N. H. Lurie. 2013. Temporal contiguity and negativity bias in the impact of online word of mouth. *Journal of Marketing Research*, 50 (4), 463–476.

Chevalier, J. A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43 (3), 345–354.

Chintagunta, P. K., S. Gopinath, S. Venkataraman. 2010. The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29 (5), 944–957.

Clemons, E. K., G. G. Gao, L. M. Hitt. 2006. When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, 23 (2), 149–171.

Cui, R., S. Gallino, A. Moreno, D. J. Zhang. 2018. The operational value of social media information. *Production and Operations Management*, 27 (10), 1749–1769.

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

Dellarocas, C., X. M. Zhang, N. F. Awad. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21 (4), 23–45.

Duan, W., B. Gu, A. B. Whinston. 2008a. Do online reviews matter? An empirical investigation of panel data. *Decision Support Systems*, 45 (4), 1007–1016.

Duan, W., B. Gu, A. B. Whinston. 2008b. The dynamics of online word-of-mouth and product salesan empirical investigation of the movie industry. *Journal of Retailing*, 84 (2), 233–242.

Eppler, M. J., J. Mengis. 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines. *The Information Society*, 20 (5), 325–344.

Feng, S., L. Xing, A. Gogar, Y. Choi. 2012. Distributional footprints of deceptive product reviews. In *Proceedings of the 6th International Conference on Web and Social Media*, pp. 98–105. AAAI.

Filippas, A., J. J. Horton, J. Golden. 2018. Reputation inflation. In *Conference on Economics and Computation*, pp. 483–484. ACM.

Geva, T., M. Saar-Tsechansky. 2021. Who is a better decision maker? data-driven expert ranking under unobserved quality. *Production and Operations Management*, 30 (1), 127–144.

Goes, P. B., M. Lin, C.-m. Au Yeung. 2014. "Popularity effect" in user-generated content: Evidence from online product reviews. *Information Systems Research*, 25 (2), 222–238.

Greenwood, B. N., R. Agarwal. 2015. Matching platforms and hiv incidence: An empirical investigation of race, gender, and socioeconomic status. *Management Science*, 62 (8), 2281–2303.

Greenwood, B. N., S. Wattal. 2017. Show me the way to go home: An empirical investigation of ride-sharing and alcohol related motor vehicle fatalities. *MIS Quarterly*, 41 (1), 163–187.

Gu, B., J. Park, P. Konana. 2012. The impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research*, 23 (1), 182–196.

Guan, X., Y. Wang, Z. Yi, Y.-J. Chen. 2020. Inducing consumer online reviews via disclosure. *Production and Operations Management*, 29 (8), 1956–1971.

Han, X., N. Yaraghi, R. Gopal. 2018. Winning at all costs: Analysis of inflation in nursing homes rating system. *Production and Operations Management*, 27 (2), 215–233.

Hansen, M. T., M. R. Haas. 2001. Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. *Administrative Science Quarterly*, 46 (1), 1–28.

He, J., X. Wang, M. B. Vandenbosch, B. R. Nault. 2020. Revealed preference in online reviews: Purchase verification in the tablet market. *Decision Support Systems*, 132, 113281.

Ho, Y.-C., J. Wu, Y. Tan. 2017. Disconfirmation effect on online rating behavior: A structural model. *Information Systems Research*, 28 (3), 626–642.

Hong, Y., N. Huang, G. Burtch, C. Li. 2016. Culture, conformity, and emotional suppression in online reviews. *Journal of the Association for Information Systems*, 17 (11), 2.

Hu, N., L. Liu, J. J. Zhang. 2008. Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management*, 9 (3), 201–214.

Hu, N., P. A. Pavlou, J. Zhang. 2017. On self-selection biases in online product reviews. *MIS Quarterly*, 41 (2), 449–471.

Huang, N., Y. Hong, G. Burtch. 2017. Social network integration and user content generation: Evidence from natural experiments. *MIS Quarterly*, 41 (4), 1035–1058.

Kim, D., K. Park, J.-H. Ahn. 2018a. Do verified consumer reviews affect sales? An empirical analysis of mixed review systems in the film industry. In *Proceedings of the European Conference on Information Systems (ECIS)*. AIS.

Kim, S. J., E. Maslowska, E. C. Malthouse. 2018b. Understanding the effects of different review features on purchase probability. *International Journal of Advertising*, 37 (1), 29–53.

Kim, S.-M., P. Pantel, T. Chklovski, M. Pennacchiotti. 2006. Automatically assessing review helpfulness. In *Conference on Empirical Methods in Natural Language Processing*, pp. 423–430. ACL.

Kokkodis, M. 2012. Learning from positive and unlabeled amazon reviews: Towards identifying trustworthy reviewers. In *International Conference on World Wide Web*, pp. 545–546. ACM.

Kokkodis, M. 2021. Dynamic, multi-dimensional, and skillset-specific reputation systems for online work. *Information Systems Research*, p. (Articles In Advance).

Kokkodis, M., P. G. Ipeirotis. 2016. Reputation transferability in online labor markets. *Management Science*, 62 (6), 1687–1706.

Kokkodis, M., T. Lappas. 2020. Your hometown matters: Popularity-difference bias in online reputation platforms. *Information Systems Research*, 31 (2), 412–430.

Kokkodis, M., T. Lappas, S. Ransbotham. 2020. From lurkers to workers: Predicting voluntary contribution and community welfare. *Information Systems Research*, 31 (2), 607–626.

Kumar, N., L. Qiu, S. Kumar. 2018a. Exit, voice, and response on digital platforms: An empirical investigation of online management response strategies. *Information Systems Research*, 29 (4), 849–870.

Kumar, N., D. Venugopal, L. Qiu, S. Kumar. 2018b. Detecting review manipulation on online platforms with hierarchical supervised learning. *Journal of Management Information Systems*, 35 (1), 350–380.

Kumar, N., D. Venugopal, L. Qiu, S. Kumar. 2019. Detecting anomalous online reviewers: An unsupervised approach using mixture models. *Journal of Management Information Systems*, 36 (4), 1313–1346.

Kumar, S., V. Mookerjee, A. Shubham. 2018c. Research in operations management and information systems interface. *Production and Operations Management*, 27 (11), 1893–1905.

Lappas, T., G. Sabnis, G. Valkanas. 2016. The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research*, 27 (4), 940–961.

Lau, R. Y. K., W. Zhang, W. Xu. 2018. Parallel aspect-oriented sentiment analysis for sales forecasting with big data. *Production and Operations Management*, 27 (10), 1775–1794.

Le, Q., T. Mikolov. 2014. Distributed representations of sentences and documents. In *International Conference on Machine Learning*, pp. 1188–1196.

Le, T. T., W. Fu, J. H. Moore. 2020. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics*, 36 (1), 250–256.

Lee, S.-Y., L. Qiu, A. Whinston. 2018. Sentiment manipulation in online platforms: An analysis of movie tweets. *Production and Operations Management*, 27 (3), 393–416.

Li, H., Z. Chen, B. Liu, X. Wei, J. Shao. 2014. Spotting fake reviews via collective positive-unlabeled learning. In *International Conference on Data Mining*, pp. 899–904. IEEE.

Li, X., L. M. Hitt, Z. J. Zhang. 2011. Product reviews and competition in markets for repeat purchase products. *Journal of Management Information Systems*, 27 (4), 9–42.

Liu, Q. B., E. Karahanna. 2017. The dark side of reviews: The swaying effects of online product reviews on attribute preference construction. *MIS Quarterly*, 41 (2), 427–448.

LIWC 2018. Linguistic Inquiry and Word Count. [Online; accessed: 28-November-2020].

Luca, M., G. Zervas. 2016. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, 62 (12), 3412–3427.

Mallipeddi, R. R., R. Janakiraman, S. Kumar, S. Gupta. 2021. The effects of social media content created by human brands on engagement: Evidence from Indian general election 2014. *Information Systems Research*, 32 (1), 212–237.

Mayzlin, D., Y. Dover, J. Chevalier. 2014. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104 (8), 2421–55.

McCallum, A., K. Nigam, others. 1998. A comparison of event models for naive bayes text classification. In *Workshop on Learning for Text Categorization*, volume 752, pp. 41–48. AAAI.

Mejia, J., S. Mankad, A. Gopal. 2020. Service quality using text mining: Measurement and consequences. *Manufacturing & Service Operations Management*.

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

Mudambi, S. M., D. Schuff. 2010. Research note: What makes a helpful online review? A study of customer reviews on amazon. com. *MIS Quarterly*, 34 (1), 185–200.

Mukherjee, A., V. Venkataraman, B. Liu, N. Glance. 2013. Fake review detection: Classification and analysis of real and pseudo reviews. Technical Report UIC-CS-2013–03, University of Illinois at Chicago.

Pavlou, P. A. 2003. Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce*, 7 (3), 101–134.

Pavlou, P. A., D. Gefen. 2004. Building effective online marketplaces with institution-based trust. *Information Systems Research*, 15 (1), 37–59.

Reinstein, D. A., C. M. Snyder. 2005. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The Journal of Industrial Economics*, 53 (1), 27–51.

Ren, Y., F. M. Harper, S. Drenner, L. Terveen, S. Kiesler, J. Riedl, R. E. Kraut. 2012. Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. *MIS Quarterly*, 36 (3), 841–864.

Saxena, A. 2018. Deception detection on amazon reviews dataset. https://bit.ly/2W5Cw15. [Online; accessed: 10-March-2020].

Schneiderman 2014. Attorney General Schneiderman announces agreement with 19 companies to stop writing fake online reviews and pay more than $350,000 in fines. https://on.ny.gov/38L4a6t. [Online; accessed: 10-March-2019].

Scrapehero 2018. How many products does amazon sell? https://bit.ly/39GzDI8. [Online; accessed 10-March-2019].

Smith, D. 2004. *Amazon reviewers brought to book.* The Guardian.

Sridhar, S., R. Srinivasan. 2012. Social influence effects in online product ratings. *Journal of Marketing*, 76 (5), 70–88.

Statista 2018. Quarterly share of e-commerce sales of total u.s. retail sales from 1st quarter 2010 to 2nd quarter 2018. https://bit.ly/2Q1cCHR. [Online; accessed 10-March-2020].

Sussin, J., E. Thompson. 2012. The consequences of fake fans, 'likes' and reviews on social networks. https://gtnr.it/2IyOUP5. [Online; accessed 10-March-2020].

Todri, V. 2020. The impact of ad-blockers on online consumer behavior. *Marketing Science (Forthcoming).*

Wang, Y., P. Goes, Z. Wei, D. Zeng. 2019. Production of online word-of-mouth: Peer effects and the moderation of user characteristics. *Production and Operations Management*, 28 (7), 1621–1640.

Wasko, M. M., S. Faraj, others. 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 29 (1), 35–57.

Wehner, M. 2016. Amazon deletes over 500,000 sketchy reviews in massive cleanup effort. *BGR*. [Online; accessed 08-May-2019].

Wooldridge, M. J. 2010. *Econometric analysis of cross section and panel data.* MIT Press.

Wu, C., H. Che, T. Y. Chan, X. Lu. 2015. The economic value of online reviews. *Marketing Science*, 34 (5), 739–754.

Yin, D., S. Bond, H. Zhang. 2014. Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, 38 (2), 539–560.

Yuan, W., Y. Hong, P. Pavlou. 2012. Do consumers trust online product reviews? An experimental study of biases in online product reviews. In *Americas Conference on Information Systems*.

Zhang, D., L. Zhou, J. L. Kehoe, I. Y. Kilic. 2016. What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems*, 33 (2), 456–481.

Zhao, Y., S. Yang, V. Narayan, Y. Zhao. 2013. Modeling consumer learning from online product reviews. *Marketing Science*, 32 (1), 153–169.

E-Companion

A Tables and Figures referenced in Section 5

Table 5 Market-specific trends do not drive the results

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Market-specific trends	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.06^{*}		3.16		2.59		-1.17^{**}
		(0.03)		(2.24)		(8.30)		(0.36)
Accumulated product rating (simple average)		0.16^{***}		-4.56^{***}		-2.72		0.20
		(0.01)		(0.83)		(3.19)		(0.11)
Accumulated product reviews		0.00		-0.23		-5.16^{**}		-0.49^{***}
		(0.01)		(0.76)		(1.64)		(0.11)
Ratings deviation		0.01		-3.07^{***}		-5.11^{*}		-0.35^{***}
		(0.01)		(0.61)		(2.35)		(0.09)
Verified reviews		0.14^{***}		-5.73^{***}		-23.40^{***}		0.15^{***}
		(0.01)		(0.42)		(1.26)		(0.04)
Likelihood of top-ranked review		0.16^{***}		84.04^{***}		36.11^{***}		-0.16
		(0.01)		(0.46)		(1.24)		(0.09)
Accumulated product rating (reverse engineered)		-0.06^{***}		3.08^{**}		0.23		-0.13
		(0.01)		(1.10)		(2.89)		(0.16)
Review length								2.07^{***}
								(0.15)
Assigned rating								-1.30^{***}
								(0.06)
OPV introduction	-0.07^{***}	-0.24^{***}	11.63^{***}	9.51^{***}	9.20^{**}	35.57^{***}	1.16^{***}	0.77^{***}
	(0.01)	(0.01)	(1.24)	(1.16)	(3.04)	(3.85)	(0.12)	(0.13)
Observations	180359	180359	180359	180359	180359	180359	180359	180359
R^2	0.01	0.04	0.11	0.43	0.01	0.02	0.03	0.06

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Table 6 Random coincidental factors do not drive the results

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.15^{**}		18.91^{*}		9.29		-7.80^{***}
		(0.05)		(7.76)		(21.59)		(1.91)
Accumulated product rating (simple average)		0.13^{***}		-2.59^{*}		3.89		0.13
		(0.01)		(1.12)		(8.35)		(0.25)
Accumulated product reviews		−0.01		-8.95^{**}		−9.10		-2.80^{***}
		(0.02)		(2.85)		(5.28)		(0.78)
Ratings deviation		0.01		-2.25^{**}		−0.81		−0.06
		(0.01)		(0.85)		(6.04)		(0.26)
Verified reviews		0.00		0.00		0.00		0.00
		(.)		(.)		(.)		(.)
Likelihood of top-ranked review		0.15^{***}		93.23^{***}		37.36^{***}		0.34
		(0.01)		(0.85)		(2.22)		(0.29)
Accumulated product rating (reverse engineered)		-0.11^{***}		4.00^{*}		−11.32		-2.15^{***}
		(0.01)		(1.75)		(11.76)		(0.60)
VP introduction						0.00		
						(.)		
Review length								2.39^{***}
								(0.44)
Assigned rating								-1.68^{***}
								(0.18)
Placebo OPV	−0.00	−0.02	−3.22	−4.21	11.26	4.91	0.03	0.47
	(0.02)	(0.02)	(3.45)	(2.77)	(9.54)	(9.51)	(0.54)	(0.54)
Observations	70340	70340	43324	43324	33776	33776	23770	23770
R^2	0.01	0.03	0.00	0.35	0.00	0.01	0.01	0.04

The table shows the results of placebo regressions. Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Table 7 Analysis of reviews from the U.S. platform

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-4.10^{**}		142.40		-661.27		12.27
		(1.41)		(170.53)		(449.26)		(14.52)
Accumulated product rating (simple average)		-0.39^{**}		3.30		10.54		-0.67
		(0.12)		(9.34)		(24.67)		(1.55)
Ratings deviation		0.00		-5.95		-1.27		-1.87
		(0.06)		(6.93)		(12.94)		(0.98)
Accumulated product reviews		-0.20^{**}		-0.88		30.80^{***}		-1.41^{*}
		(0.06)		(3.98)		(8.78)		(0.56)
Verified reviews		0.16^{***}		-15.00^{***}		-18.43^{***}		0.24^{*}
		(0.01)		(1.56)		(3.89)		(0.11)
Likelihood of top-ranked review		0.17^{***}		94.22^{***}		30.16^{***}		-0.02
		(0.01)		(1.19)		(2.15)		(0.14)
Accumulated product rating (reverse engineered)		0.02		6.21		-1.90		0.57
		(0.06)		(6.16)		(14.23)		(0.68)
Review length								1.16^{***}
								(0.20)
Assigned rating								-2.02^{***}
								(0.22)
OPV introduction	$-0.09.$	-0.17^{**}	16.27^{*}	20.58^{*}	29.23	57.35^{*}	0.94.	0.21
	(0.05)	(0.06)	(8.21)	(8.04)	(20.87)	(23.02)	(0.54)	(0.67)
Observations	21764	21764	21764	21764	21764	21764	21764	21764
R^2	0.00	0.04	0.00	0.33	0.00	0.01	0.00	0.04

This analysis includes only reviews from the U.S. platform. Since the U.S. was the first eventually treated platform, contamination from browsing cannot be present. Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$).

Table 8 Random selection of products that appear in a single market

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-1.32^{*}		120.39	243.30	237.48		13.30
		(0.63)		(65.15)	(215.28)	(214.33)		(8.47)
Accumulated product rating (simple average)		-0.03^{*}		1.46		2.92		-0.06
		(0.01)		(1.40)		(5.79)		(0.19)
Ratings deviation		-0.01		-1.77		-1.84		-0.79^{***}
		(0.01)		(0.99)		(3.52)		(0.14)
Accumulated product reviews		-0.06^{***}		6.41^{***}		12.82^{***}		-0.45^{*}
		(0.01)		(0.94)		(2.71)		(0.19)
Verified reviews		0.13^{***}		-5.60^{***}		-22.19^{***}		0.21^{***}
		(0.01)		(0.59)		(1.70)		(0.06)
Likelihood of top-ranked review		0.17^{***}		84.02^{***}		36.80^{***}		-0.03
		(0.01)		(0.60)		(1.67)		(0.10)
Accumulated product rating (reverse engineered)		0.02		0.36		-5.87		-0.15
		(0.01)		(1.48)		(4.39)		(0.21)
Review length								2.07^{***}
								(0.16)
Assigned rating								-1.33^{***}
								(0.09)
OPV introduction	-0.06^{***}	-0.23^{***}	13.28^{***}	7.85^{***}	12.91^{*}	31.79^{***}	1.64^{***}	1.07^{***}
	(0.02)	(0.02)	(1.97)	(1.86)	(5.22)	(5.75)	(0.23)	(0.23)
Observations	89763	89763	89763	89763	89763	89763	89763	89763
R^2	0.01	0.04	0.10	0.42	0.01	0.02	0.03	0.06

This subsample analysis randomly chooses products such that each product appears only in a single market. As a result, it removes relative spillover effects from users who were browsing the U.S. platform while buying a product from the U.K. one. Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p<0.001$, ** $p<0.01$, * $p<0.05$).

Table 9 Nearest neighbor matching of products

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-2.53^{*}		247.16^{*}		590.57	10.05	6.41
		(1.06)		(104.83)		(370.85)	(11.99)	(11.56)
Accumulated product rating (simple average)		-0.00		2.98		14.59		0.47
		(0.03)		(2.26)		(10.08)		(0.33)
Ratings deviation		-0.01		-2.07		9.45		-1.13^{***}
		(0.02)		(1.53)		(5.54)		(0.28)
Accumulated product reviews		-0.03		4.92^{*}		21.44^{***}		-0.40
		(0.03)		(2.09)		(4.73)		(0.34)
Verified reviews		0.13^{***}		-5.71^{***}		-23.05^{***}		0.17
		(0.01)		(0.97)		(3.11)		(0.09)
Likelihood of top-ranked review		0.16^{***}		84.95^{***}		39.33^{***}		0.11
		(0.01)		(1.03)		(2.73)		(0.20)
Accumulated product rating (reverse engineered)		0.00		-2.40		-0.57		-0.48
		(0.03)		(2.28)		(6.69)		(0.38)
Review length								1.91^{***}
								(0.34)
Assigned rating								-1.31^{***}
OPV introduction	-0.05	-0.21^{***}	14.66^{***}	6.77^{*}	26.85^{**}	44.33^{***}	1.93^{***}	1.33^{***}
	(0.03)	(0.04)	(3.21)	(2.85)	(8.55)	(9.73)	(0.37)	(0.34)
Observations	28969	28969	28969	28969	28969	28969	28969	28969
R^2	0.01	0.04	0.11	0.43	0.01	0.03	0.04	0.07

Each paired product appears in either the U.S. or the U.K. platform. Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Table 10 Subsample of products that had post-treatment reviews in the U.K. first

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.09^{*}		9.28^{*}		-17.72		-1.16
		(0.05)		(3.79)		(15.42)		(0.72)
Accumulated product rating (simple average)		0.14^{***}		-3.38^{***}		-2.82		0.16
		(0.01)		(0.93)		(3.97)		(0.13)
Ratings deviation		-0.00		-2.64^{***}		-5.81^{*}		-0.57^{***}
		(0.01)		(0.68)		(2.81)		(0.11)
Accumulated product reviews		-0.00		1.09		-3.04		-0.77^{***}
		(0.01)		(1.13)		(2.27)		(0.18)
Verified reviews		0.14^{***}		-6.36^{***}		-23.98^{***}		0.22^{***}
		(0.01)		(0.49)		(1.42)		(0.05)
Likelihood of top-ranked review		0.16^{***}		87.34^{***}		37.27^{***}		-0.11
		(0.01)		(0.52)		(1.39)		(0.11)
Accumulated product rating (reverse engineered)		-0.05^{***}		1.97		-0.06		-0.22
		(0.01)		(1.17)		(3.40)		(0.18)
Review length								2.23^{***}
								(0.18)
Assigned rating								-1.42^{***}
								(0.08)
OPV introduction	-0.06^{*}	-0.21^{***}	22.91^{***}	14.09^{***}	21.30^{*}	38.84^{***}	1.50^{***}	0.92^{**}
	(0.03)	(0.03)	(3.88)	(3.21)	(9.75)	(9.94)	(0.29)	(0.29)
Observations	141300	141300	141300	141300	141300	141300	141300	141300
R^2	0.01	0.04	0.08	0.41	0.01	0.02	0.03	0.06

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Table 11 Concurrent events do not drive the results

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Other events	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.06^{*}		2.39		3.84		-0.82^{*}
		(0.03)		(2.24)		(8.32)		(0.36)
Accumulated product		0.17^{***}		-4.33^{***}		-3.05		-0.13
rating (simple average)		(0.01)		(0.83)		(3.19)		(0.11)
Ratings deviation		0.01		-2.74^{***}		-5.58^{*}		-0.51^{***}
		(0.01)		(0.61)		(2.34)		(0.09)
Accumulated product reviews		0.00		-0.77		-4.38^{**}		-0.32^{***}
		(0.01)		(0.72)		(1.56)		(0.09)
Verified reviews		0.14^{***}		-5.79^{***}		-23.33^{***}		-0.07
		(0.01)		(0.42)		(1.27)		(0.04)
Likelihood of		0.16^{***}		83.94^{***}		36.24^{***}		0.94^{***}
top-ranked review		(0.01)		(0.46)		(1.25)		(0.06)
Accumulated product rating		-0.06^{***}		3.01^{**}		0.36		-0.00
(reverse engineered)	(0.01)	(0.01)		(1.09)		(2.88)		(0.15)
OPV introduction	-0.07^{***}	-0.24^{***}	13.44^{***}	7.55^{***}	17.08^{***}	38.78^{***}	1.72^{***}	1.67^{***}
	(0.01)	(0.01)	(1.30)	(1.19)	(3.66)	(4.10)	(0.14)	(0.15)
Observations	180359	180359	180359	180359	180359	180359	180359	180359
R^2	0.01	0.04	0.11	0.43	0.01	0.02	0.02	0.03

Other events include dummies that control for the introduction of Nook, iBooks, Azure, Amazon video, Amazon app store, California, Nevada, and Texas taxes, Kindle fire, Kindle fire HD, and the acquisition of Kiva. Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Table 12 Differences in publication dates across the two markets do not drive the results

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.04		3.81		-6.15		-0.40
		(0.04)		(3.35)		(11.92)		(0.49)
Accumulated product rating (simple average)		0.18^{***}		-3.74^{**}		-1.00		0.19
		(0.02)		(1.21)		(4.33)		(0.17)
Ratings deviation		0.00		-3.13^{***}		-4.27		-0.66^{***}
		(0.01)		(0.93)		(3.23)		(0.12)
Accumulated product reviews		0.01		-0.88		-3.23		-0.18
		(0.01)		(1.17)		(2.02)		(0.14)
Verified reviews		0.14^{***}		-5.65^{***}		-24.95^{***}		0.15^{*}
		(0.01)		(0.60)		(1.82)		(0.07)
Likelihood of top-ranked review		0.17^{***}		83.79^{***}		38.48^{***}		-0.19
		(0.01)		(0.67)		(1.82)		(0.12)
Accumulated product rating (reverse engineered)		-0.06^{***}		1.24		-0.26		-0.30
		(0.02)		(1.52)		(3.93)		(0.22)
Review length								2.01^{***}
								(0.19)
Assigned rating								-1.19^{***}
								(0.09)
OPV introduction	-0.06^{***}	-0.24^{***}	10.61^{***}	5.97^{***}	10.88^{*}	34.28^{***}	1.62^{***}	1.23^{***}
	(0.02)	(0.02)	(1.87)	(1.76)	(5.15)	(5.92)	(0.20)	(0.22)
Observations	87060	87060	87060	87060	87060	87060	87060	87060
R^2	0.01	0.05	0.11	0.43	0.01	0.02	0.02	0.05

The table shows results from a randomly chosen subset of U.S. reviews in a way that the difference in publication dates between the U.S. and the U.K. vanishes. Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Figure 8 The two markets had the same distributions of fake reviews before the introduction of OPV; fake reviews do not differ before and after the introduction of OPV; restrictions on the number of fake reviews per week do not drive the results.

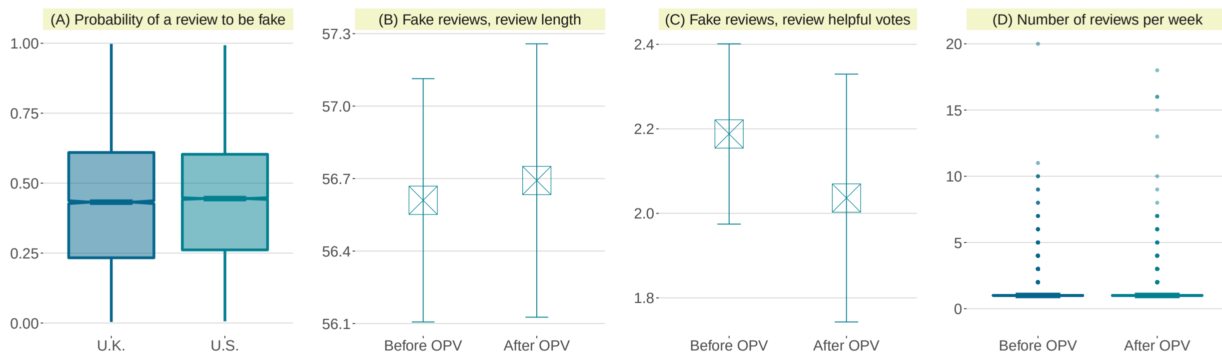
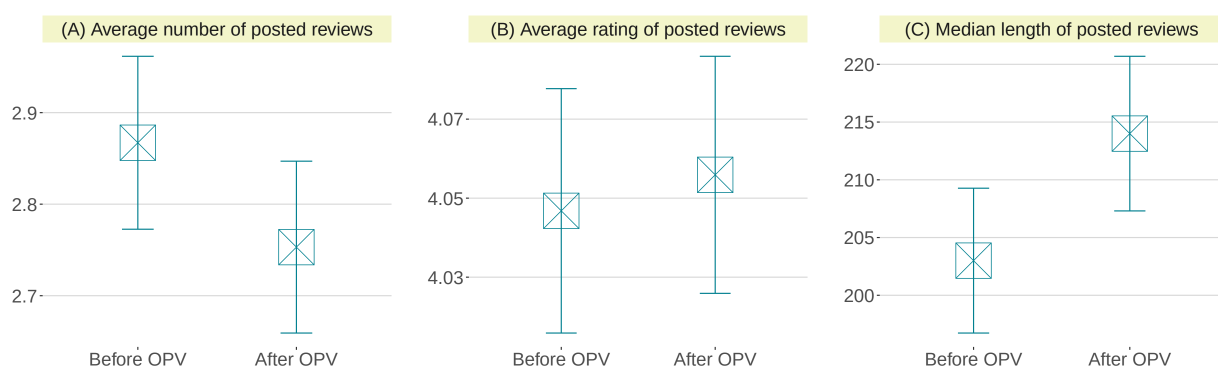


Figure A shows that the distributions of fake reviews across the two markets are almost identical (not statistically significantly different) before the introduction of OPV in the U.S. Figures B and C show that fake reviews have the same characteristics before and after OPV. Figure D shows that most reviewers in our data post 1 non-verified review per week; some post up to 20. The distributions are not statistically different before and after the introduction of OPV. The box plots visualize five summary statistics: the median, two hinges, two whiskers, and all "outlying" points individually.

Clarification: Even though our predictive model was trained on a completely different dataset of reviews, it is possible that it identifies similar text characteristics of fake reviews before and after the OPV. As a result, Figures (B) and (C) might be affected by algorithmic bias. Error bars show 95% confidence intervals.

Figure 9 Regular reviewers did not significantly change their behavior after the introduction of OPV



The Figure shows that there is no empirical evidence that regular reviewers (A) review more, (B) are more negative, and (C) post longer reviews after the introduction of OPV. Regular reviewers post reviews both before and after the introduction of OPV. Error bars show 95% confidence intervals.

B Additional evidence in support of our DID setting

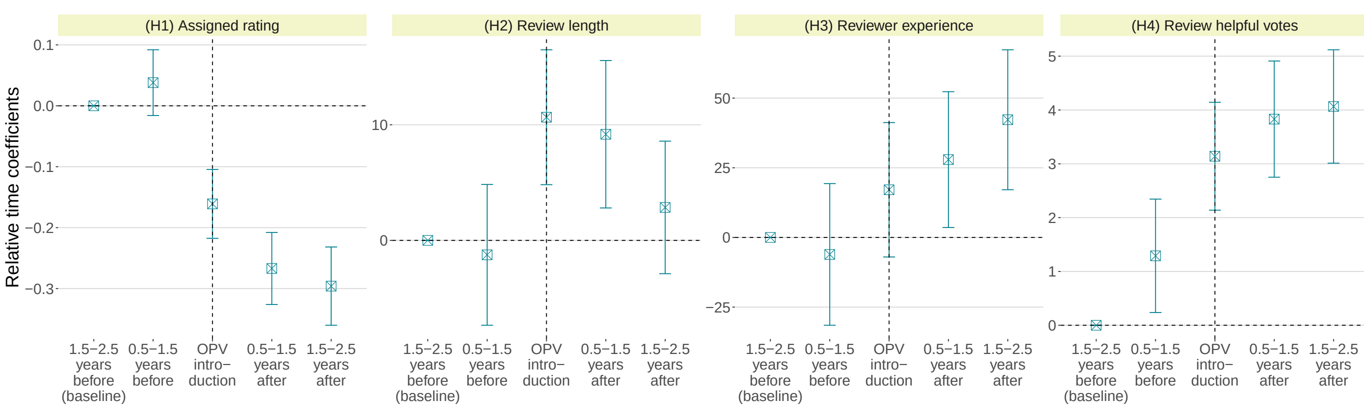
B.1 Relative time regressions

The core specification (Equation 1) shows an aggregate difference in differences approach between the same products in treated (OPV) and untreated (non-OPV) markets. A concern for this approach is the existence of homogeneity in the pretreatment trend between the OPV and the non-OPV markets. This concern exists because the two markets are in different countries (U.S., vs. U.K.), and as a result, unobservable and randomly distributed environmental factors can create heterogeneity in the pretreatment trends. To reduce this concern, and further support the argument that the introduction of OPV at different times in the two markets can be treated as an exogenous shock, we estimate the following relative time regression (Greenwood and Agarwal 2015, Greenwood and Wattal 2017, Todri 2020):

$$\mathrm{DV}_{pmt} = P + M + E + \zeta_m[\rho_{r-3}, \rho_{r-2}, ..., \rho_{r+2}][1^t_{r-3}, 1^t_{r-2}, ..., 1^t_{r+2}]' + \boldsymbol{\beta}\boldsymbol{X}_{pmt} + \varepsilon_i\,, \tag{3}$$

where ζ_m is a dummy variable that indicates whether or not market m will ever get treated (i.e.,, whether it will get OPV). Coefficients ρ_t are the relative time coefficients of interest. The benefit of this model is that it can detect whether a pre-treatment trend exists (i.e., "*a significant difference between treated and untreated markets before treatment*" (Greenwood and Wattal 2017)) to identify whether the untreated market is an acceptable control group.[9]

Figure 10 Relative time



Error bars show 99% confidence intervals.

Figure 10 shows the results. First, we observe that for all the dependent variables, there is no pretreatment trend (before the introduction of OPV, all coefficients are zero (baseline) or not statistically significantly different from zero). Second, once purchase verification is introduced, most

[9] We consider the period between 2.5 years before the introduction of OPV in the U.S., and 2.5 years after, i.e., before the introduction of OPV in the U.K. market, so that ζ_t can distinguish the two markets.

coefficients become statistically significant. Specifically, for assigned rating, the coefficients become negative, while for review length, reviewer experience, and review helpful votes, the coefficients become positive. These observations rule out several econometric concerns related to the assumptions of the difference in difference strategy, as there are no consistent detectable differences in the trends of the four dependent variables before the introduction of OPV (Angrist and Pischke 2008, Greenwood and Wattal 2017).

B.2 Exogeneity of the introduction of OPV

Amazon's official announcement explicitly states that Amazon's goal when introducing purchase verification was to improve the relevance of reviews. Yet, one could argue that Amazon's decision could be endogenous to one of the four dependent variables in this study. For instance, Amazon could have noticed that product ratings followed an increasing trend over time and decided to introduce optional purchase verification to deflate these ratings. In general, reputation inflation is a major problem that online markets are continuously trying to solve, as inflated ratings increase expectation disconfirmation and decrease trust in the whole reputation system (Akerlof 1978, Filippas et al. 2018, Kokkodis 2021). Column H1 in Table 13 shows that no pre-treatment trend in the U.S. market could have driven the decision to introduce optional purchase verification.

Similarly, Amazon could have noticed that the length and the helpfulness of reviews, or the reviewer experience were following a decreasing trend, and decided to stimulate these quantities by introducing OPV. Columns H2-H4 in Table 13 address these concerns by showing that none of these variables follows a statistically significant trend before introducing purchase verification. Hence, the decision to introduce purchase verification was likely not driven by a trend of any of the dependent variables.

Table 13 Time trends in the U.S. market before the introduction of OPV

	Assigned rating (H1)	Review length (H2)	Reviewer experience (H3)	Review helpful votes (H4)
Product FE	✓	✓	✓	✓
Product tenure	−2.40	135.15	379.43	33.27
	(1.32)	(170.67)	(391.93)	(18.70)
Accumulated product	-0.18^{**}	9.87	-41.03^{*}	0.23
rating (simple average)	(0.06)	(7.12)	(16.35)	(0.78)
Accumulated product reviews	-0.20^{***}	3.51	14.36	-2.84^{***}
	(0.05)	(5.87)	(13.48)	(0.64)
Ratings deviation	0.01	−2.84	−13.26	-2.08^{***}
	(0.03)	(4.25)	(9.75)	(0.47)
Verified reviews	1.28	−122.95	26.39	−17.62
	(1.42)	(184.07)	(422.69)	(20.17)
Likelihood of	0.14^{***}	95.61^{***}	31.27^{***}	1.10^{***}
top-ranked review	(0.01)	(0.98)	(2.25)	(0.11)
Accumulated product rating	−0.03	3.55	30.62^{**}	0.25
(reverse engineered)	(0.04)	(4.73)	(10.87)	(0.52)
Assigned rating				-2.16^{***}
				(0.12)
Time trend (bi-weekly)	0.05	−2.91	−8.75	−0.71
	(0.03)	(3.59)	(8.25)	(0.39)
Observations	20,607	20,607	20,607	20,607
R^2	0.65	0.36	0.35	0.59

Time-trend coefficients for one year before the introduction of OPV in the U.S. market. Clustered standard errors in parentheses. The constant term is estimated but omitted from the table. (*** $p<0.001$, ** $p<0.01$, * $p<0.05$).

C Heterogeneous effects

An alternative explanation suggests that the observed effects might be driven by a specific group of products (e.g., very high-quality products) and do not generalize to all products in the marketplace. To test this possibility, we cluster products into three quartiles depending on their average reputation before the introduction of the OPV: Low quality (bottom 33% quartile), Medium quality (between 33% and 66%), and High quality (upper 33% quartile). Table 14 shows the results. Even though there is some product heterogeneity (e.g., stronger effect on product ratings of high-quality products), the main effects of the introduction of OPV remain qualitatively similar to our main results presented in Table 2. Hence, we can conclude that product heterogeneity does not drive the observed results.

Table 14 Product quality does not drive the observed effects

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.05^{*}		1.57		3.26		-0.86^{*}
		(0.03)		(2.26)		(8.33)		(0.36)
Accumulated product rating (simple average)		0.17^{***}		-4.44^{***}		-3.06		0.15
		(0.01)		(0.83)		(3.19)		(0.11)
Ratings deviation		0.01		-2.63^{***}		-5.50^{*}		-0.46^{***}
		(0.01)		(0.60)		(2.33)		(0.09)
Accumulated product reviews		-0.00		-0.63		-4.32^{**}		-0.33^{***}
		(0.01)		(0.70)		(1.55)		(0.10)
Verified reviews		0.14^{***}		-6.21^{***}		-23.50^{***}		0.22^{***}
		(0.01)		(0.43)		(1.27)		(0.04)
Likelihood of top-ranked review		0.16^{***}		83.89^{***}		36.23^{***}		-0.12
		(0.01)		(0.46)		(1.25)		(0.09)
Accumulated product rating (reverse engineered)		-0.05^{***}		2.55^{*}		0.08		-0.07
		(0.01)		(1.09)		(2.89)		(0.15)
Review length								2.06^{***}
								(0.15)
Assigned rating								-1.31^{***}
								(0.06)
OPV introduction	-0.05^{**}	-0.18^{***}	9.35^{***}	3.85^{**}	20.48^{***}	37.30^{***}	2.02^{***}	1.73^{***}
	(0.02)	(0.02)	(1.77)	(1.45)	(4.74)	(4.96)	(0.18)	(0.19)
OPV × Medium quality	-0.03	-0.08^{***}	0.83	3.47^{*}	-8.88^{*}	-2.80	-0.42	-0.58^{**}
	(0.02)	(0.02)	(2.14)	(1.64)	(4.50)	(4.42)	(0.21)	(0.22)
OPV × High quality	-0.02	-0.12^{***}	10.83^{***}	9.59^{***}	-3.93	4.54	-0.48^{*}	-0.90^{***}
	(0.02)	(0.02)	(2.02)	(1.58)	(4.17)	(4.18)	(0.22)	(0.22)
Observations	180359	180359	180359	180359	180359	180359	180359	180359
R^2	0.01	0.04	0.11	0.43	0.01	0.02	0.02	0.06

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

D Dropping the post-treatment period

Table 15 shows the results of our analysis if we do not include observations after the introduction of OPV in the U.K. market. The results corroborate our main results presented in Table 2.

Table 15 **Effects of the introduction of OPV on product ratings and review characteristics (excluding observations after 03/2012)**

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.08^{**}		5.47^{*}		4.81		-1.33^{**}
		(0.03)		(2.74)		(10.12)		(0.46)
Accumulated product rating (simple average)		0.16^{***}		-4.31^{***}		-2.94		0.21
		(0.01)		(0.86)		(3.48)		(0.12)
Ratings deviation		0.01		-2.75^{***}		-5.14^{*}		-0.34^{***}
		(0.01)		(0.63)		(2.54)		(0.09)
Accumulated product reviews		-0.01		-1.94^{*}		-7.08^{**}		-0.62^{***}
		(0.01)		(0.98)		(2.26)		(0.14)
Verified reviews		0.14^{***}		-9.41^{***}		-19.32^{***}		0.22^{***}
		(0.01)		(0.60)		(1.60)		(0.06)
Likelihood of top-ranked review		0.16^{***}		87.31^{***}		37.24^{***}		-0.06
		(0.01)		(0.51)		(1.35)		(0.10)
Accumulated product rating (reverse engineered)		-0.08^{***}		3.21^{**}		1.85		-0.33
		(0.01)		(1.13)		(3.32)		(0.18)
Review length								2.06^{***}
								(0.15)
Assigned rating								-1.43^{***}
								(0.07)
OPV introduction	-0.06^{***}	-0.25^{***}	16.38^{***}	8.66^{***}	22.88^{***}	38.19^{***}	2.64^{***}	2.24^{***}
	(0.01)	(0.02)	(1.70)	(1.61)	(4.64)	(5.30)	(0.19)	(0.21)
Observations	143406	143406	143406	143406	143406	143406	143406	143406
R^2	0.00	0.04	0.06	0.39	0.01	0.02	0.02	0.05

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

E Prediction of product rating

In Section 3.1, we used the variable "Accumulated product rating (reverse engineered)" that proxies Amazon's algorithm for estimating a product's rating. To derive these product-rating estimates, we built predictive models that reverse-engineer Amazon's algorithm. Specifically, we used all the available information for each product, including:

- The simple average of ratings that the product has received.
- The median length of product reviews.
- The number of product reviews.
- The median helpful votes that the product reviews have received.
- The standard deviation of the received product ratings.
- The average probability of a review to be fake, as estimated by our predictive model described in E-Companion G.
- The percentage of verified reviews.

We used as our target variable the product rating assigned by Amazon on Amazon's product page.

For our predictive models, we used TPOT (Le et al. 2020), a "Python Automated Machine Learning tool that optimizes machine learning pipelines using genetic programming." Specifically, we ran the following code:

```
tpot = TPOTRegressor(generations=5, population_size=200, verbosity=2, random_state=1234,
                     scoring='neg_mean_absolute_error',cv=10)
tpot.fit(transformed,tt.product_rating)
```

The resulting best predictive pipeline was:

```
exported_pipeline = ExtraTreesRegressor(bootstrap=False, max_features=0.9,
                    min_samples_leaf=3, min_samples_split=9, n_estimators=100)
```

The average 10-fold cross-validation mean absolute error of this approach was 0.177. The trained model allowed us to provide the estimates of variable "Accumulated product rating (reverse engineered)."

F Prediction of review position

In Section 3.1, we also used the variable "Likelihood of top-ranked review" that captures the likelihood of a given review to be displayed on the front page of a product. To estimate this probability, we built predictive models that estimate this likelihood based on observed top-ranked reviews in Amazon products at the time of our data collection.

As predictive features, we used extracted information from the raw review text. Specifically, advances in deep learning allowed us to map review text into vectors of real numbers (Distributed Memory Model (DMM); see Le and Mikolov 2014). The primary parameter of this process is the dimensionality of the embedding space. We set this value to 100, as higher values did not improve the fit of our models. As a result, each review was mapped into a 100-dimensional vector of real numbers. We used these 100 dimensions as predictors. In addition, we considered the assigned product rating of a review. Then, we used the TPOT (Le et al. 2020) package as follows:[10]

```
tpot = TPOTClassifier(generations=5, population_size=100, verbosity=2, random_state=1234,
                      scoring='roc_auc',cv=3)
tpot.fit(np.array(X_toPredict),y)
```

This process yielded the following pipeline:

```
exported_pipeline = make_pipeline(
    OneHotEncoder(minimum_fraction=0.05, sparse=False, threshold=10),
    StandardScaler(),
    XGBClassifier(learning_rate=0.1, max_depth=1, min_child_weight=9, n_estimators=100, n_jobs=1,
    subsample=0.75, verbosity=0)
)
```

The average 3-fold cross-validated performance of this approach was as follows:

- Accuracy: 97.8%
- Precision: 95.6%
- Recall: 97.8%
- F-score: 96.7%

[10] Note that we used three folds as this process was computationally challenging (the feature space explodes through the numerical representations of text). The presented TPOT job was running for four days.

G Prediction of fake reviews

To build predictive models that estimate the likelihood of a review to be fake, we used a 100-dimensional review representation DMM model similar to the one described in E-Companion F. In particular, we train our DMM model on the external labeled Amazon reviews dataset described in Section 4.2.2 (Saxena 2018). Then, we used the TPOT (Le et al. 2020) package as follows:

```
tpot = TPOTClassifier(generations=5, population_size=100, verbosity=2, random_state=1234,
                      scoring='roc_auc',cv=10)
tpot.fit(np.array(X),y)
```

This process yielded the following pipeline:

```
exported_pipeline = make_pipeline(
    StackingEstimator(estimator=BernoulliNB(alpha=0.001, fit_prior=True)),
    XGBClassifier(learning_rate=0.001, max_depth=4, min_child_weight=14,
    n_estimators=100, n_jobs=1, subsample=0.1, verbosity=0)
)
```

The trained pipeline (on the external labeled Amazon dataset) can predict which reviews from the focal dataset are likely to be fake. For each focal review, the pipeline estimates its probability to be fake. To minimize noise, we label the top 10% ranked reviews as fake and reviews that rank in the bottom 10% as not fake. The average 10-fold cross-validated performance of this approach in the top-10% ranked instances was:

- Accuracy: 72.5%
- Precision: 72.8%
- Recall: 99.1%
- F-score: 83.9%

The average 10-fold cross-validated performance of this approach in the bottom-10% ranked instances was:

- Accuracy: 76.1%
- Precision: 76.3%
- Recall: 99.6%
- F-score: 86.3%

H Sensitivity analysis

The observed results in Table 2 could be incidental to the bi-weekly aggregations. To test their sensitivity to various aggregation periods, Tables 16, 17, and 18 show the results for monthly, weekly, and non-aggregated data. The results remain robust: introducing optional purchase verification yields lower product ratings (H1), longer reviews (H2), reviews posted by more experienced reviewers (H3), and more helpful reviews (H4).

Furthermore, our results might be affected by the potentially overdispersed distributions of some of our dependent variables. To test for this possibility, Table 19 shows the results for log-transformed versions of our dependent variables, while Table 20 shows negative binomial models for our three count variables. In both cases, our results remain robust and qualitatively the same as in our main Table 2.

Table 16 Results for monthly aggregations

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.05		3.67		3.47		-1.34^{***}
		(0.03)		(2.42)		(8.11)		(0.33)
Accumulated product rating (simple average)		0.17^{***}		-4.16^{***}		-4.11		0.07
		(0.01)		(0.85)		(3.40)		(0.11)
Ratings deviation		0.01		-2.28^{***}		-5.79^{*}		-0.37^{***}
		(0.01)		(0.64)		(2.52)		(0.09)
Accumulated product reviews		0.02^{*}		-0.87		-7.17^{***}		-0.26^{**}
		(0.01)		(0.65)		(1.52)		(0.09)
Verified reviews		0.12^{***}		-5.22^{***}		-21.42^{***}		0.15^{***}
		(0.01)		(0.45)		(1.41)		(0.04)
Likelihood of top-ranked review		0.15^{***}		81.00^{***}		35.91^{***}		0.00
		(0.01)		(0.49)		(1.35)		(0.08)
Accumulated product rating (reverse engineered)		-0.05^{***}		2.85^{*}		-0.75		-0.15
		(0.01)		(1.14)		(2.96)		(0.16)
Review length								1.72^{***}
								(0.12)
Assigned rating								-1.16^{***}
								(0.06)
OPV introduction	-0.07^{***}	-0.24^{***}	13.61^{***}	7.42^{***}	19.18^{***}	39.77^{***}	1.87^{***}	1.45^{***}
	(0.01)	(0.01)	(1.36)	(1.27)	(3.83)	(4.38)	(0.14)	(0.14)
Observations	131790	131790	131790	131790	131790	131790	131790	131790
R^2	0.01	0.04	0.12	0.44	0.01	0.02	0.03	0.06

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < .001$, ** $p < .01$, * $p < .05$).

Table 17 Results for weekly aggregations

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.06^{*}		2.09		-3.12		-0.65
		(0.03)		(2.24)		(8.93)		(0.37)
Accumulated product rating (simple average)		0.16^{***}		-4.20^{***}		-2.28		0.20
		(0.01)		(0.84)		(2.96)		(0.11)
Ratings deviation		0.00		-3.08^{***}		-4.71^{*}		-0.57^{***}
		(0.01)		(0.59)		(2.19)		(0.09)
Accumulated product reviews		-0.01		-0.89		-1.61		-0.40^{***}
		(0.01)		(0.77)		(1.46)		(0.11)
Verified reviews		0.14^{***}		-6.05^{***}		-23.91^{***}		0.24^{***}
		(0.01)		(0.41)		(1.21)		(0.04)
Likelihood of top-ranked review		0.17^{***}		85.87^{***}		37.52^{***}		-0.20^{*}
		(0.01)		(0.45)		(1.18)		(0.09)
Accumulated product rating (reverse engineered)		-0.05^{***}		2.67^{*}		2.09		-0.10
		(0.01)		(1.07)		(2.60)		(0.16)
Review length								2.31^{***}
								(0.15)
Assigned rating								-1.44^{***}
								(0.06)
OPV introduction	-0.06^{***}	-0.23^{***}	12.43^{***}	7.89^{***}	12.01^{***}	33.80^{***}	1.43^{***}	0.93^{***}
	(0.01)	(0.01)	(1.28)	(1.15)	(3.35)	(3.77)	(0.13)	(0.14)
Observations	221310	221310	221310	221310	221310	221310	221310	221310
R^2	0.01	0.04	0.10	0.42	0.01	0.02	0.02	0.05

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Table 18 Results for non-aggregated observations

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.07^{*}		0.80		-9.81		-0.14
		(0.03)		(2.59)		(10.01)		(0.41)
Accumulated product rating (simple average)		0.16^{***}		-4.35^{***}		-0.72		0.25
		(0.01)		(1.03)		(3.31)		(0.14)
Ratings deviation		-0.01		-3.84^{***}		-4.28		-0.94^{***}
		(0.01)		(0.63)		(2.22)		(0.12)
Accumulated product reviews		-0.01		-2.74^{**}		-1.03		-0.96^{***}
		(0.01)		(0.99)		(1.53)		(0.21)
Verified reviews		0.15^{***}		-6.34^{***}		-23.20^{***}		0.36^{***}
		(0.01)		(0.53)		(1.15)		(0.04)
Likelihood of top-ranked review		0.19^{***}		90.46^{***}		41.29^{***}		-0.47^{***}
		(0.01)		(0.54)		(1.27)		(0.12)
Accumulated product rating (reverse engineered)		-0.04^{*}		0.98		4.77		0.05
		(0.01)		(1.39)		(2.51)		(0.18)
Review length								3.05^{***}
								(0.21)
Assigned rating								-1.64^{***}
								(0.08)
OPV introduction	-0.05^{***}	-0.23^{***}	10.28^{***}	7.93^{***}	8.66^{*}	29.27^{***}	0.78^{***}	0.28^{*}
	(0.01)	(0.01)	(1.38)	(1.12)	(3.49)	(3.79)	(0.13)	(0.14)
Observations	336043	336043	336043	336043	336043	336043	336043	336043
R^2	0.01	0.04	0.09	0.41	0.00	0.02	0.02	0.06

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Table 19 Results of log-transformed dependent variables

	Assigned rating (H1)		Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓	✓	✓
Product tenure		-0.03^{**}		0.04^{*}		0.00		-0.14^{**}
		(0.01)		(0.01)		(0.04)		(0.04)
Accumulated product rating (simple average)		0.06^{***}		-0.03^{***}		-0.01		0.03^{**}
		(0.00)		(0.00)		(0.01)		(0.01)
Ratings deviation		0.01^{**}		-0.01^{***}		-0.02^{*}		-0.04^{***}
		(0.00)		(0.00)		(0.01)		(0.01)
Accumulated product reviews		0.00		0.01^{*}		0.01		-0.03^{**}
		(0.00)		(0.00)		(0.01)		(0.01)
Verified reviews		0.05^{***}		-0.04^{***}		-0.16^{***}		0.04^{***}
		(0.00)		(0.00)		(0.01)		(0.00)
Likelihood of top-ranked review		0.08^{***}		0.71^{***}		0.27^{***}		0.05^{***}
		(0.00)		(0.00)		(0.01)		(0.00)
Accumulated product rating (reverse engineered)		-0.02^{***}		0.02^{***}		-0.01		0.03^{**}
		(0.00)		(0.01)		(0.01)		(0.01)
Review length								0.12^{***}
								(0.01)
Assigned rating								-0.24^{***}
								(0.00)
OPV introduction	-0.03^{***}	-0.09^{***}	0.12^{***}	0.06^{***}	0.03^{*}	0.17^{***}	0.21^{***}	0.13^{***}
	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	180359	180359	180359	180359	180359	180359	180359	180359
R^2	0.01	0.06	0.16	0.66	0.01	0.04	0.10	0.21

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Table 20 Results for negative binomial specifications

	Review length (H2)		Reviewer experience (H3)		Review helpful votes (H4)	
	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)
Time (bi-weekly) FE	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓
Product tenure		-0.05^{***}		0.01		-0.03^{**}
		(0.01)		(0.01)		(0.01)
Accumulated product rating (simple average)		-0.01^{***}		-0.00		0.06^{***}
		(0.00)		(0.00)		(0.01)
Ratings deviation		-0.02^{***}		-0.01^{**}		-0.00
		(0.00)		(0.00)		(0.00)
Accumulated product reviews		0.02^{***}		0.05^{***}		-0.04^{***}
		(0.00)		(0.00)		(0.00)
Verified reviews		-0.04^{***}		-0.05^{***}		0.08^{***}
		(0.00)		(0.00)		(0.00)
Likelihood of top-ranked review		0.57^{***}		0.11^{***}		0.11^{***}
		(0.00)		(0.00)		(0.00)
Accumulated product rating (reverse engineered)		0.05^{***}		0.04^{***}		-0.07^{***}
		(0.00)		(0.00)		(0.01)
Review length						0.10^{***}
						(0.00)
Assigned rating						-0.34^{***}
						(0.00)
OPV introduction	0.10^{***}	0.05^{***}	0.01	0.05^{***}	0.36^{***}	0.22^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	180359	180359	180359	180359	179735	179735

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

I Analysis on search products

Our main analysis focuses on books; to showcase that our results likely hold across various types of products, we perform a similar analysis on a small sample of electronics that we were able to collect that appeared in both the U.S. and the U.K. markets. Table 21 shows that our results hold for these products as well. (Note that we could not retrieve the total number of reviews for each reviewer in this subsample so we only estimate specifications for Assigned rating, Review length, and Review helpful votes. In addition, we did not have enough information for these products to run the complete specification of Equation 1, and as a result we only present fixed-effect specifications.)

Table 21 Effects of OPV on electronics

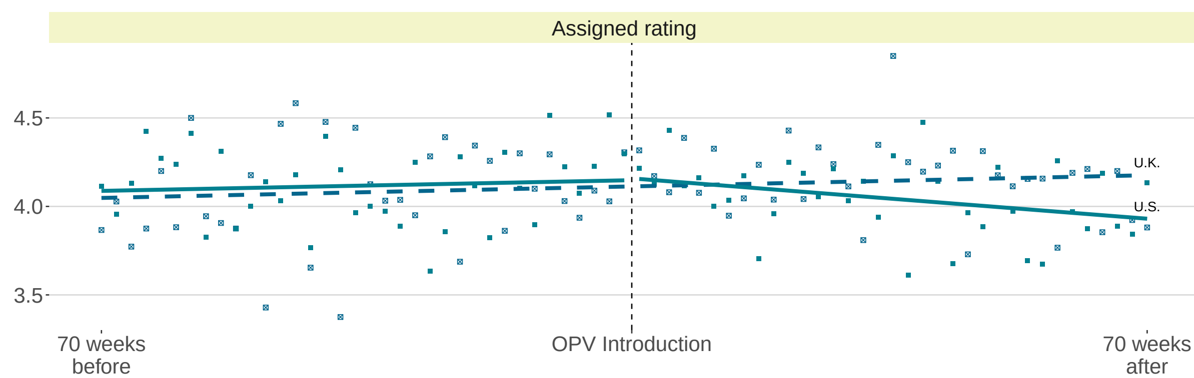
	Assigned rating (H2)	Review length (H3)	Review helpful votes (H4)
	(A1)	(B1)	(C1)
Time (bi-weekly) FE	✓	✓	✓
Product FE	✓	✓	✓
Market FE	✓	✓	✓
OPV introduction	-0.12^{***}	8.35^{***}	7.90^{***}
	(0.03)	(2.26)	(0.64)
Observations	76275	76275	76275
R^2	0.01	0.03	0.08

Clustered standard errors (on products) in parentheses. The constant term is estimated but omitted from the table. FE: Fixed Effects. (*** $p<0.001$, ** $p<0.01$, * $p<0.05$).

J Additional Figures

Figure 11 shows the visualization of the parallel trend as derived from our data on the two markets. To avoid any product-specific trending effects, the Figure shows the visualization of products with similar number (±3) of bi-weekly reviews in both markets; that way, differences in product popularity that a simple visualization cannot control for do not affect the graph (these differences are captured in the product fixed effects in our empirical specifications). The Figure shows a clear drop in product ratings in the treated platform right after the introduction of OPV, as conceptualized in Figures 1 and 4.

Figure 11 Visualization of the parallel trend



The figure shows the actual parallel trend for the average assigned ratings across the two platforms as captured in our data.

Figure 12 shows the correlations of our focal and control variables. As expected, some variables are relatively correlated as they capture similar quantities (e.g., “Accumulated product rating (simple average)” with “Accumulated product rating (reverse engineered)”). As shown in Table 2, our results hold for different specifications and subsets of these variables.

Figure 12 Correlations of the focal and control variables

	Review_length	OPV_introduction	Product_tenure	Accumulated_product_rating_.simple_average.	Ratings_deviation	Number_of_reviews	Verified_reviews	Linkelihood_of_top.ranked_review	Accumulated_product_rating_.reversed_engineered.
Accumulated_product_rating_.reversed_engineered.									
Linkelihood_of_top.ranked_review									0
Verified_reviews								−0.3	0.1
Number_of_reviews							0.3	−0.2	0
Ratings_deviation						0.2	0.1	−0.1	−0.5
Accumulated_product_rating_.simple_average.					−0.7	−0.1	0.1	0	0.8
Product_tenure				0	0.2	0.3	0.6	−0.3	0
OPV_introduction			0.7	0	0.2	0.4	0.7	−0.2	0
Review_length		−0.2	−0.3	0	−0.1	−0.1	−0.2	0.7	0
Assigned_rating	−0.1	0	0.1	0.4	−0.3	0	0.1	0.1	0.4

Shaded circles show correlations greater than 0.5.