



Department for  
Business & Trade

# FAKE ONLINE REVIEWS RESEARCH

Estimating the prevalence and impact of  
fake online reviews

April 2023

---

## About the authors



Alma Economics combines unparalleled analytical expertise with the ability to communicate complex ideas clearly.

[www.almaeconomics.com](http://www.almaeconomics.com)



© Crown copyright 2023

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://nationalarchives.gov.uk/doc/open-government-licence/version/3) or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: [psi@nationalarchives.gov.uk](mailto:psi@nationalarchives.gov.uk)

---

Any enquiries regarding this publication should be sent to us at:  
[enquiries@trade.gov.uk](mailto:enquiries@trade.gov.uk)

---

# Contents

List of Tables	5
List of Figures	6
Executive summary	7
Context and objectives	7
Methodology	8
Key findings	9
Limitations	10
Policy implications	11
Introduction	12
Estimating the prevalence of fake reviews	15
Building the network and other features	16
Evaluating feature importance	18
Selecting a model	20
Training the model	22
Amazon reviews data	23
Predictive model findings	23
Other e-commerce platforms	26
Results and implications	26
Estimating the impact of fake reviews on consumers	28
Hypotheses	29
Participants and sampling strategy	29
Experimental design	30
Materials	31
Online retail platform and product categories	31
Product reviews	32
Post-experiment questionnaire	34
Willingness to pay	34
Experiment follow-up	34
Procedure	35
	35
Descriptive statistics	35
Participant engagement	38

---

Findings	40
Consumer assessment of fake reviews	40
Impact on purchasing decisions	41
Supplementary regression analysis: Impact on consumer trust and future behaviour	45
Impact on consumer welfare and broader implications	48
An indicative model of consumer harm	48
Limitations	50
External validity	52
Conclusion	54
Bibliography	56
Appendix 1: Approach to predictive modelling	59
Appendix 2: Post-experiment questions	63
Appendix 3: Products and product reviews displayed to experiment participants on the online shopping platform	70
Appendix 4: Supplementary analyses	72
Appendix 5: Regression details	98

---

## List of Tables

Table 1: Product-level features generated from review content and product network.....	17
Table 2: Out-of-sample prediction performance, random forest classifier (simple train-test) ...	21
Table 3: Descriptive statistics of reviews for selected sectors.....	23
Table 4: Proportion of Amazon reviews predicted as fake by the RF algorithm, by sector.....	24
Table 5: Proportion of reviews predicted as fake (other popular e-commerce platforms) .....	26
Table 6: List of experiment groups .....	30
Table 7: Demographic breakdown of sample .....	36
Table 8: Participant evaluation of reviews .....	40
Table 9: Impact of fake reviews and informed silence on the probability of choosing a product (across all product categories) in the online shopping task. ....	41
Table 10: Impact of fake reviews and informed silence on the probability of choosing a product (by product category). ....	43
Table 11: Impact of fake reviews and informed silence on participant behaviour in online shopping task (products priced higher than £80 in the online shopping task) .....	44
Table 12: Impact of fake reviews and informed silence on consumer trust and future behaviour. .....	46
Table 13: Consumer harm from subtle fake reviews – purchasing product with fake reviews.	49

## List of Figures

Figure 1: Individual feature importance for the random forest classifier .....	19
Figure 2: Comparison of model partition of the decision space for sample of 150 reviews .....	22
Figure 3: Prevalence of fake reviews across time for different product categories.....	25
Figure 4: Screenshot of the product overview page .....	31
Figure 5: Screenshot of informed silence text box .....	33
Figure 6: Diagram of main experimental procedure .....	35
Figure 7: Factors driving purchasing decisions .....	39

# Executive summary

## Context and objectives

Online consumer reviews play an important role in the purchasing decisions other consumers make online. These reviews serve as an important source of information to mitigate uncertainty around product quality, particularly when consumers have not seen the products themselves beforehand<sup>1</sup> (Manes and Tchetichik 2018). Consumers generally perceive the information posted in online reviews as unbiased, and reviews can often “make or break” the success of a product or service<sup>2</sup> (de Langhe, Fernbach, and Lichtenstein 2016). This provides an incentive for product sellers to manipulate their online reviews by purchasing or anonymously posting fake reviews intended to deceive consumers and increase sales.

Fake reviews can be favourable towards the seller’s product or unfavourable towards the products sold by competing businesses. Both strategies are intended to make consumers purchase products that they might not have in the absence of fake reviews. For the purposes of this research, we define a fake review as a review of a product or service which does not reflect a genuine experience of that product or service and has been designed to mislead consumers.

There is no consensus on how consumer choice, trust and future behaviour are impacted by fake reviews. Consumers might be aware that manipulation is taking place through fake reviews and adjust their interpretations of online opinions accordingly<sup>3</sup> (Zhuang, Cui and Peng 2018). Alternatively, consumers might not be able to correct for the bias in evaluating product quality introduced by fake reviews if they cannot distinguish between fake and genuine reviews<sup>4</sup> (Hu, Liu and Sambamurthy 2011). In addition, few papers have distinguished between different types of fake reviews (such as fake reviews that are more or less obviously written), and most previous research focuses on consumers in the US rather than the UK.

Alma Economics (the authors of this study) was commissioned by the Department for Business and Trade (DBT) to answer the following research questions:

- What is the prevalence of online fake reviews on popular third-party UK e-commerce websites?<sup>5</sup>

---

<sup>1</sup> Manes, Eran, and Anat Tchetichik. 2018. ‘The Role of Electronic Word of Mouth in Reducing Information Asymmetry: An Empirical Investigation of Online Hotel Booking’. *Journal of Business Research* 85 (April): 185–96. <https://doi.org/10.1016/j.jbusres.2017.12.019>.

<sup>2</sup> Langhe, Bart de, Philip M. Fernbach, and Donald R. Lichtenstein. 2016. ‘Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings’. *Journal of Consumer Research* 42 (6): 817–33. <https://doi.org/10.1093/jcr/ucv047>

<sup>3</sup> Zhuang, Mengzhou, Geng Cui, and Ling Peng. 2018. ‘Manufactured Opinions: The Effect of Manipulating Online Product Reviews’. *Journal of Business Research* 87 (June): 24–35. <https://doi.org/10.1016/j.jbusres.2018.02.016>.

<sup>4</sup> Hu, Nan, Ling Liu, and Vallabh Sambamurthy. 2011. ‘Fraud Detection in Online Consumer Reviews’. *Decision Support Systems, On quantitative methods for detection of financial fraud*, 50 (3): 614–26. <https://doi.org/10.1016/j.dss.2010.08.012>

<sup>5</sup> Third-party e-commerce websites are online marketplaces that manage and host sales for other businesses.



- How do online fake reviews influence consumer choice when making online purchases?
- What is the harm to consumers caused by fake reviews?
- How effective are potential non-regulatory interventions in avoiding consumers being misled by fake reviews?

## Methodology

As part of this study, we combined two separate approaches to understand the impact of online fake reviews on UK consumers.

First, we built a machine learning model to predict whether reviews were genuine or fake.<sup>6</sup> Predictions were based on characteristics of reviews used in previous detection models (such as similarity with other reviews, review posting history and average review length), but also included network features that took into account whether products shared reviewers with other products. As fake reviews have become increasingly similar to genuine reviews over time as people who write fake reviews try to avoid detection, the content of the review itself has become less helpful in distinguishing genuine and fake reviews. As a result, characteristics of reviews not related to content, such as network features, are key in providing additional predictive power to the model.

Once the model was built, we then trained it on a dataset of known fake reviews collected from private Facebook groups where sellers buy reviews<sup>7</sup> (He et al. 2022b). Previous models were trained on datasets of AI-generated fake reviews (using language models such as GPT-2) or platform-filtered reviews, which are not necessarily similar to the fake reviews seen by UK consumers when they shop online. Because we know for certain which reviews are fake and which are genuine, our model can provide more accurate predictions when deployed on the reviews of e-commerce platforms. This trained model was then applied to a dataset of 2.1 million product reviews across 9 popular UK e-commerce platforms (this larger dataset was unlabelled, which means we do not know for certain which reviews are genuine or fake). The outputs from this model allowed us to estimate the percentage of product reviews on these 9 platforms predicted to be fake.

Second, we carried out an online experiment with 4,900 participants in the UK who had previously shopped online. In this experiment, participants were asked to complete an online shopping task and purchase one of three similar products.<sup>8</sup> The online shopping task was fully interactive and was designed to be as realistic as possible to the practice of shopping and reading reviews on a popular e-commerce site.

---

<sup>6</sup> Machine learning models are based on the concept of learning algorithms: by “training” a model on a dataset of product reviews which have been labelled as fake or genuine, it then becomes possible for the model to classify unlabelled reviews (i.e. reviews for which we do not know whether they are fake or genuine) without being explicitly programmed with the steps required to complete the task.

<sup>7</sup> He, Sherry, Brett Hollenbeck, and Davide Proserpio. 2022. ‘The Market for Fake Reviews’. SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.3664992>.

<sup>8</sup> There were 11 product types considered in this study: Bluetooth headphones, irons, kettles, desk chairs, smart speakers, keyboards, power banks, re-usable water bottles, yoga mats, sunscreen and vacuums.

As part of the online shopping task, some participants only viewed genuine reviews when they clicked on a specific product page, while other participants viewed a mix of genuine and fake reviews<sup>9</sup>. In addition, some participants saw a text box (displayed above all product reviews) stating that steps had been taken to moderate misleading content, including misleading reviews, on the platform. Following the online shopping task, participants completed a follow-up questionnaire covering their choices in the online shopping task, general shopping behaviour/preferences and demographics questions.

Based on the experiment described above, we compared the purchasing decisions made by participants who only viewed genuine product reviews with those who viewed both genuine and fake product reviews and those who saw the warning text box. We then assessed whether fake reviews (or warning consumers about fake reviews) changed the probability that a product with fake reviews was purchased. We also explored whether the impact of fake reviews differed across product type, product price and participant demographic characteristics, and how exposure to fake reviews changed consumer trust in platforms and future shopping behaviour.

To supplement these findings, we built a simple indicative model quantifying the annual harm to UK consumers caused by fake reviews on third-party platforms. This model was based on the idea that consumers misled by fake reviews make suboptimal choices, purchasing products that are lower in quality or do not align with their individual preferences (compared to the product they would have purchased in the absence of fake reviews).

### Key findings

Our results present a nuanced picture of how consumers are impacted by fake online product reviews:

- 1. The prevalence of fake reviews differs across product categories and platforms.** For e-commerce platforms widely used by UK consumers, we estimate that **11% to 15%** of all reviews for three common product categories (consumer electronics, home and kitchen, sports and outdoors) are fake.
- 2. Network features (whether a product had a reviewer in common with another product) are stronger predictors of fake reviews than the content of reviews.** Products with fake reviews have more reviewers in common than products that only have genuine reviews. This aligns with empirical evidence that most fake reviews are written by a small pool of individuals who participate in incentivised review service marketplaces (compared to the millions of users that buy products online and write genuine reviews).
- 3. Consumers are 5.3% less likely to purchase a product with poorly written (“strong”) fake reviews and 3.1% more likely to purchase a product with well-written (“subtle”) fake reviews.** However, the size of this impact depends on the price and category of the product. Fake reviews had a greater impact on consumer behaviour for consumer

---

<sup>9</sup> The fake reviews were written by the research team and we received feedback on the fake reviews from a representative at Which?, the UK consumer advocacy group.

electronics and higher-priced products, and in particular consumers were 9.2% more likely to purchase a product with subtle fake reviews if the product price was greater than £80.

4. **Informing consumers that steps have been taken to moderate misleading content on the platform does not impact consumer purchasing behaviour.** There was not a statistically significant difference in the likelihood of choosing a product with fake reviews when participants saw a banner with this additional information. However, other non-regulatory interventions may be effective in counteracting the influence of fake reviews on consumers and should be tested in future research.
5. **Exposure to fake reviews generally does not impact consumer trust and future behaviour.** Despite being exposed to fake reviews on the online platform, we did not observe consumers adapting their purchasing behaviour, leaving them potentially susceptible to being affected by further misinformation in the future.
6. **The impact of fake reviews on consumers does not vary depending on their demographic characteristics.** We did not find any differences in the effect of fake reviews on characteristics such as age, sex and ethnicity. This suggests that fake reviews have a similar impact on different groups of UK consumers.
7. **Fake review text on products alone causes an estimated £50 million to £312 million in total annual harm to UK consumers.** However, this estimate does not include the impact of fake reviews on consumers who purchase services, on future consumer behaviour or the separate impact of inflated star ratings (which often accompany fake reviews). As a result, this is a conservative estimate and the true consumer detriment arising from fake reviews is likely to be higher.

## Limitations

There are some important limitations to the approaches used for this study.

1. While products for the experiment shopping platform were chosen to be similar in star rating, price and other characteristics, there were still differences in visual appearance and key characteristics that may have influenced consumers' decision (in addition to variation in the content of reviews).
2. Participants were only asked to make a purchasing decision as part of the online shopping task that aligned with how they would act in the real world (instead of actually spending their own money, which means they may have been less motivated to find the "best" or "highest quality" product).
3. We only examined the prevalence and impact of fake review text on consumer products. As such, we cannot determine whether these findings also extend to services purchased online or other misleading review practices. Additionally, the estimated total harm caused to consumers does not include the impact of inflated star ratings.

However, the size of the experiment (based on the total number of participants), our integration of a bespoke online shopping platform that closely resembled an actual shopping experience

and our use of a broad range of fake review types means we can be confident that our study robustly estimates the impact of fake reviews on UK consumers.

### Policy implications

Our key findings, in particular (i) at least 10% of all product reviews on third-party e-commerce platforms are likely to be fake, and (ii) the presence of well-written “subtle” fake reviews leads to a statistically significant increase in the proportion of consumers buying the product with these fake reviews, highlight the importance of taking steps to reduce the prevalence of online fake product reviews. Our findings suggest that consumers are more susceptible to being misled by well-written fake reviews when purchasing products where reviews play a more prominent role in consumer decisions (such as consumer electronics or higher-priced products). If consumers are generally not able to distinguish genuine and well-written “subtle” fake reviews, and fake reviews are becoming more sophisticated and difficult to detect over time, the negative impacts of fake reviews on consumers are likely to increase over time as well. This suggests three main areas for future policy to consider:

Automated means of review moderation should focus on the characteristics of reviewers at least as much as the characteristics of reviews, given that network features (based on which products had reviewers in common with other products) are stronger predictors of whether reviews are fake than the content of reviews themselves.

The high levels of data and computational power required to generate product-reviewer networks for popular e-commerce platforms suggest that e-commerce platforms are better positioned to spot fake reviews compared to consumers. This aligns with previous research finding that even trained researchers cannot consistently and accurately distinguish between genuine and fake reviews<sup>10</sup> (Plotkina, Munzel and Pallud 2020).

Our research found that informing consumers that steps have been taken to moderate misleading content (such as misleading customer reviews) does not seem to counteract the influence of fake reviews. This provides evidence that consumer trust in product reviews tend to be grounded in prior online shopping experiences and cannot easily be altered. It is possible that interventions that use stronger language or are more salient to consumers could increase awareness of fake reviews and encourage consumers to be more cautious. Future research should test whether these types of non-regulatory interventions can be effective despite the evidence indicating that consumers struggle to spot well written fake reviews, as these can be straightforward and cost-effective for platforms to implement.

---

<sup>10</sup> Plotkina, Daria, Andreas Munzel, and Jessie Pallud. 2020. ‘Illusions of Truth—Experimental Insights into Human and Algorithmic Detections of Fake Online Reviews’. *Journal of Business Research* 109 (March): 511–23

## Introduction

With e-commerce in the UK now accounting for a third of the total retail market (Trade, 2022), shopping online has increasingly become the preferred method of consumption among individuals. The shift away from brick-and-mortar has further been accelerated by the Covid-19 pandemic, which saw consumers limited to shopping online to comply with government restrictions.<sup>11</sup> While such restrictions have ended and shopping in store is resuming to an extent, online shopping remains popular as consumers appreciate its convenience and efficiency, and in 2022 27% of all retail sales made by UK consumers took place online (Shaw et al. 2022)<sup>12, 13</sup>

Integral to online shopping are user-generated ratings and reviews which are not only valuable for consumers when deciding which products or services to buy, but equally essential for companies to ensure that they stand out among a sea of close competitors<sup>14</sup> (Chang et al. 2015). Reviews are widely available for both specific products, on websites or third-party review sites, as well as retailers themselves, through sites such as Trustpilot. Previous research has found that online reviews also have an important role in consumer decision-making for experience goods i.e., products that cannot be easily tried or evaluated before purchase, leading to information asymmetry between consumers and sellers<sup>15</sup> (Park and Lee 2009, Manes and Tchetichik 2018). However, this study focuses specifically on goods that people purchase and can get information on online prior to purchase.

Some reviews are submitted by individuals and organisations that do not reflect an actual consumer's genuine experience of a good or service. These "fake" reviews, while resembling the reviews of genuine consumers, are instead designed to influence consumers' purchasing decisions or target a particular business. Negative fake reviews are critical of a good or service and may be left by a business to harm their competitor. Most fake reviews however, are positive in nature and designed to encourage consumer purchases. These reviews often overstate a products' qualities and are more prevalent among low-quality products<sup>16</sup> (Akesson et al. 2022; He et al. 2022a). When there is a discrepancy between a product's reviews and its

---

<sup>11</sup><https://www.ons.gov.uk/businessindustryandtrade/retailindustry/articles/howourspendinghaschangedsincetheendofcoronaviruscovid19restrictions/2022-07-11>

<sup>12</sup> Shaw, Norman, Brenda Eschenbrenner, and Daniel Baier. 2022. 'Online Shopping Continuance after COVID-19: A Comparison of Canada, Germany and the United States'. *Journal of Retailing and Consumer Services* 69 (November): 103100. <https://doi.org/10.1016/j.jretconser.2022.103100>.

<sup>13</sup> <https://www.ons.gov.uk/businessindustryandtrade/retailindustry/timeseries/j4mc/drsi>

<sup>14</sup> Chang, Hsin Hsin, Po Wen Fang, and Chien Hao Huang. 2015. 'The Impact of On-Line Consumer Reviews on Value Perception: The Dual-Process Theory and Uncertainty Reduction'. *Journal of Organizational and End User Computing* 27 (2): 32–57. <https://doi.org/10.4018/joeuc.2015040102>.

<sup>15</sup> Manes, Eran, and Anat Tchetichik. 2018. 'The Role of Electronic Word of Mouth in Reducing Information Asymmetry: An Empirical Investigation of Online Hotel Booking'. *Journal of Business Research* 85 (April): 185–96. <https://doi.org/10.1016/j.jbusres.2017.12.019>

<sup>16</sup> Akesson, Jesper., Robert W. Hahn, Robert D. Metcalfe, and Manuel Monti-Nussbaum. 2022. "The Impact of Fake Reviews on Demand and Welfare". Unpublished manuscript, July 20 2022, typescript.

actual features and quality, fake reviews can negatively impact consumer welfare<sup>17</sup> (Mayzlin et al. 2014). While the importance of reviews for consumer decision-making and the increasing prevalence of fake reviews has been acknowledged, less is known about the extent of their prevalence across UK websites and the specific impact they have on consumers.

One recent report estimated that \$23 billion of UK consumer spending was potentially influenced by fake reviews.<sup>18</sup> Furthermore, in 2020, the UK consumer advocacy group Which? conducted an experimental study on the impact of fake reviews using an online shopping task. The findings from this study showed that fake reviews make consumers more likely to purchase lower quality products and have a larger influence on those that shop online frequently.<sup>19</sup> Some studies have also examined the impact of different non-regulatory interventions on consumer purchasing decisions. These interventions are designed to prevent the impact of fake reviews by providing consumers with additional information about the quality of the reviews that they see. This is important as consumer decision-making can be influenced by when and how different information is displayed on the platform<sup>20</sup> (Floyd et al. 2014). A study by Ananthakrishnan et al. (2020)<sup>21</sup>, which looked at tagging and displaying suspicious reviews, found that consumers tended to trust reviews more when platforms displayed both (clearly identified) fake and genuine reviews. However, this study focused specifically on restaurant reviews, and very little evidence to date has looked at the impact of non-regulatory interventions on consumers specifically within the online retail sector.

To gain further insight into the prevalence of fake reviews we built a network-based<sup>22</sup> machine learning model to detect fake reviews, then applied this model to 2.1 million product reviews across 9 popular UK e-commerce platforms. Subsequently, to determine the impact of fake reviews and the effectiveness of a non-regulatory intervention, we conducted an experiment that builds on the findings by Akesson et al<sup>23</sup>. (2022). Specifically, we designed an online shopping task, which mimicked the experience of purchasing products on a popular UK e-commerce platform. Participants were assigned to a product type and asked to select one of three products that they wished to purchase. They also took part in a willingness to pay scenario and were asked follow-up questions regarding their decision-making. Through this experiment we were able to assess:

---

<sup>17</sup> Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. 2014. 'Promotional Reviews: An Empirical Investigation of Online Review Manipulation'. *American Economic Review* 104 (8): 2421–55. <https://doi.org/10.1257/aer.104.8.2421>.

<sup>18</sup> [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/436238/Online\\_reviews\\_and\\_endorsements.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/436238/Online_reviews_and_endorsements.pdf)

<sup>19</sup> <https://www.which.co.uk/policy/consumers/5860/realfakereviews>

<sup>20</sup> Floyd, Kristopher, Ryan Freling, Saad Alhoqail, Hyun Young Cho, and Traci Freling. 2014. 'How Online Product Reviews Affect Retail Sales: A Meta-Analysis'. *Journal of Retailing, Empirical Generalizations in Retailing*, 90 (2): 217–32. <https://doi.org/10.1016/j.jretai.2014.04.004>.

<sup>21</sup> Ananthakrishnan, Uttara M., Beibei Li, and Michael D. Smith. 2020. 'A Tangled Web: Should Online Review Portals Display Fraudulent Reviews?' SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.3297363>.

<sup>22</sup> Network-based refers to how a network of reviewers are connected across products.

<sup>23</sup> Akesson, Jesper., Robert W. Hahn, Robert D. Metcalfe, and Manuel Monti-Nussbaum. 2022. "The Impact of Fake Reviews on Demand and Welfare". Unpublished manuscript, July 20 2022, typescript.

- the impact of fake reviews and their strength (i.e. how well-written these reviews are) on decision-making,
- the effectiveness of a non-regulatory intervention,
- the relationship between fake reviews and consumer trust and future purchasing behaviour and,
- the relationship between these outcomes with demographic factors such as age, ethnicity, gender and online shopping habits.

## Estimating the prevalence of fake reviews

Machine learning models employ various statistical techniques to learn patterns and relationships in data, and then make predictions or decisions based on that learning. The model typically involves a set of parameters that are adjusted during training to optimize the model's performance on a specific task or problem. Automatic detection that applies machine learning models to classify reviews as fake or genuine offer a more robust, scalable alternative to manually detecting fake reviews. Some models focus on text-based features (such as keywords, punctuation or similarity with other reviews): for example, if a review has copied multiple sentences word-for-word from another review, it is highly unlikely that the review posted second reflects a consumer's genuine experiences. Other models are based on behavioural features that capture a platform user's data and past history of reviews (user location or IP address, average review length, percentage of positive reviews): for example, if a single reviewer consistently leaves short reviews such as "Good product", this may suggest that the reviewer has not put any effort into writing the review or even used the product at all.

These predictive models work as follows:

1. A dataset of reviews (some of which are labelled as fake and the remainder are labelled as genuine) is split into a test dataset and a training dataset.
2. Using the training dataset, the model looks for relationships between the label (fake or genuine) and the other features of reviews in the dataset (such as review length, sentiment, whether text is repeated, etc.).
3. The model quantifies the relationship between the label and other features of reviews by assigning weights to different features (for example, review length is weighed more heavily if it is closely correlated with whether reviews are fake or genuine). The more optimised these weights are, the more accurately the model can predict whether a review is genuine or fake.
4. To evaluate its accuracy, the model is checked with the test dataset: the model's prediction (i.e. whether the review is fake or genuine) is compared with the actual labels in the dataset.

The model we developed was used to estimate the prevalence of fake reviews on popular UK third-party e-commerce platforms (platforms operated by independent sellers that do not focus on products they have manufactured themselves).<sup>24</sup> The automated model we developed to detect fake reviews is distinct in two ways:

---

<sup>24</sup> Siering et al. (2018) found that consumers perceive reviews for search and experience goods differently; subjectivity tends to play a more important role in reviews for experience goods due to underlying differences in individual perceptions, which impact an assessment of a good's quality. This suggests that a predictive model trained specifically for reviews of search goods may not be as effective as predicting fake reviews for experience goods, and a robust model would need to be trained separately on a dataset of reviews for only experience goods.



1. Our model is a supervised machine learning model that has been trained on data collected from private Facebook groups in which sellers buy reviews, which means we know for certain which products are buying fake reviews (He et al. 2022a). This means we can be confident in the accuracy of the “fake” or “genuine” labels in the training dataset, which cannot be assumed for AI-generated fake reviews or platform-filtered reviews.
2. Our model is based on network footprints (the relationships between products and reviewers), which overcomes the challenges associated with relying on observable behavioural or text features alone. For example, people who write fake reviews will try to avoid detection by changing how they write reviews or using language they would use for genuine reviews, while network features cannot be easily manipulated in this way. Alternatively, some genuine product reviews may have text features which make them look fake (such as strongly positive sentiment). In addition, this type of model takes advantage of the fact that sellers who pay for fake reviews rely on a much smaller pool of potential reviewers (compared to the general population of consumers who use the platform).

Our proposed model is therefore more robust to the specific ways in which fake reviews are written (which are likely to have changed over time) and instead is based on qualitative research investigating how sellers purchase fake reviews for their products and the process through which fake reviews are commissioned and written.

### Building the network and other features

We define our network based on a set of products (“nodes”) and common reviewers (“edges”). More specifically, two products are connected by an edge if the products share at least one reviewer in common. We do not specifically record the identity of a reviewer, but only whether or not products had reviewers in common.<sup>25</sup> Based on this network, we can estimate several different metrics such as the number of connections (i.e., common reviewers) a product has to other products and how well the product is connected to the neighbourhood of neighbouring products.

We take this approach because evidence suggests most fake reviews are written by a relatively small pool of individuals who participate in incentivised review services on social media platforms. As this activity is often an important source of income, individuals will typically be responsible for writing multiple fake reviews spanning a range of products rather than just one or two fake reviews written as a one-off event (Oak and Shafiq 2021). Intuitively, the connections between reviewers and products provides another source of information to the model (in addition to the text and metadata features) to help the model better predict which reviews are fake and which reviews are genuine.

As a comparison, we also calculate a broad range of common model features observed in previous research based on user behavioural features (metadata), text features, sentiment and

<sup>25</sup> In other words, this is a binary variable with value of 1 if two products had at least one reviewer in common and 0 otherwise.

synthetic features (which combine two or more features). The full list of features considered for inclusion in our model is outlined in Table 1 below.

**Table 1: Product-level features generated from review content and product network**

Type	Feature	Description
Network	Degree	Number of reviewers in common with other goods
	Clustering coefficient	Measure of how much the “neighbours” of each product (i.e., connections) are also connected amongst themselves
	Eigenvector centrality	Measure of connectedness of a product to other highly connected products
	PageRank	Measures the importance of a product in the network by how much highly connected products are also connected to it, adjusting for the total volume of connections
	Centrality	Measures the degree to which a product is influential in the network, i.e., how easy it is to get from it to any other good. Intuitively, this is a measure of common reviewer convergence on a given product.
Metadata	TF-IDF mean similarity	Degree of similarity between reviews of a product as measured by co-occurrence and importance of words
	Number of reviews	How many times the particular good has been reviewed by different reviewers
	Average review rating	Average score from the star rating for each product
	Time between reviews	Mean, minimum, maximum, and standard deviation of the number of days between reviews for a particular good
	Share of helpful votes	How many potential purchasers found the review useful
	Share of 1 star reviews	Number of 1 star quantitative reviews attributed to the product as a proportion of all reviews given

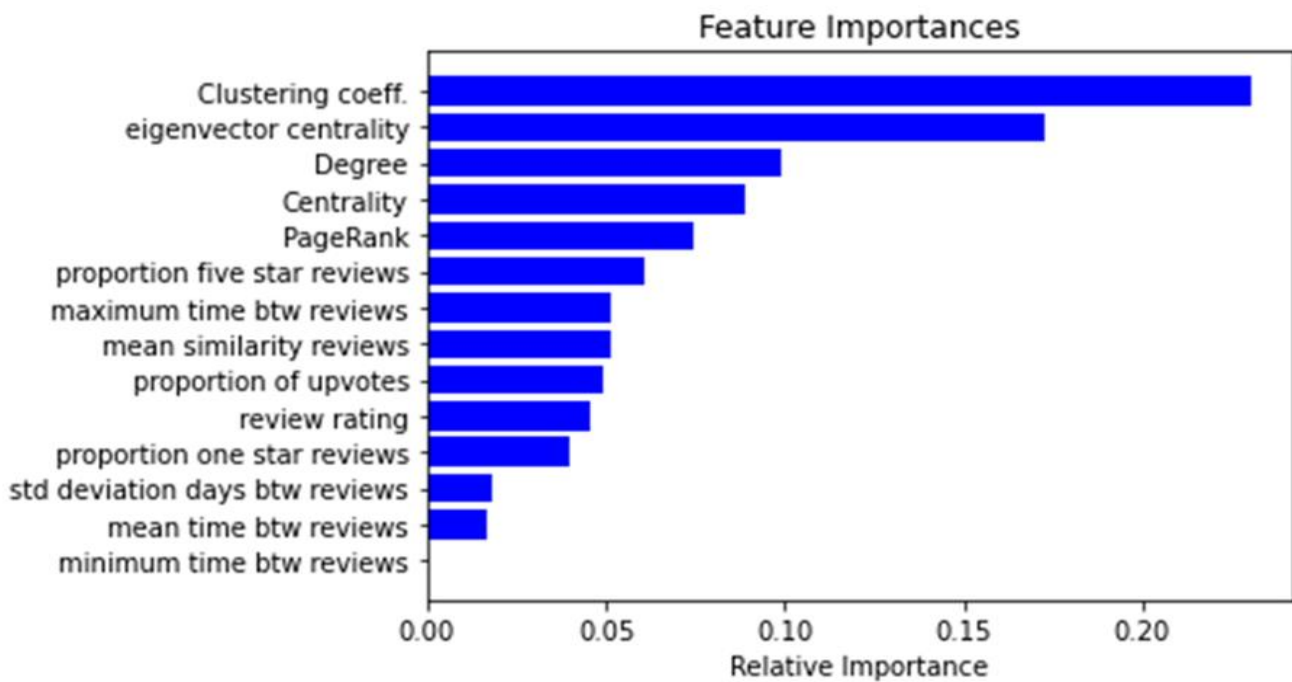
	Share of 5 star reviews	Number of 5 star quantitative reviews attributed to the product as a proportion of all reviews given
Text	TF-IDF features	Importance of words in a review, measured by their importance (in the totality of the reviews).
	Length of reviews	Length in characters of each review
	Parts of speech tagging	Proportion of words in a review that are adjectives, adverbs, verbs, pronouns, interjections, and nouns
Sentiment	Sentiment	Subjective state expressed by the review as measured by the negative or positive charge of the words contained therein.
Combined	Synthetic variables	Combinations of the variables above through mathematical operations that yield nonlinear insight into how these interact, and help boost accuracy

### Evaluating feature importance

In general, we aimed to maximise the predictive power of our model (measured by the five metrics of accuracy, recall, precision, F1 score and Area Under Curve discussed in the “Selecting a model” section below) by optimising which features the model uses as decision-making criteria. Models with high predictive power are likely to correctly identify whether a review is genuine or fake, while models with low predictive power are more likely to make inaccurate or unreliable estimates of review authenticity. Therefore, to develop a model with the “optimal” combination of features, we first consider each feature’s predictive power individually, then evaluate predictive power for combinations of different features.

As an example, Figure 1 presents estimates for the importance of individual features from Table 1 for the Random Forest classifier.

Figure 1: Individual feature importance for the random forest classifier



If a feature's relative importance is higher, then the feature is more useful in helping the model accurately predict fake or genuine label for reviews. A value of 1 means an individual feature perfectly predicts whether a review is fake or genuine, while a value of 0 means an individual feature has no impact on the "success" of a prediction.

Figure 1 demonstrates that the two most important features are the clustering coefficient and eigenvector centrality, both network features. This aligns with our hypothesis that products reliant on a smaller pool of fake reviewers will be more closely clustered in the network of products compared to products with genuine reviews. In general, when each feature is assessed individually, network features outperform behavioural features based on review metadata.

Even if we make sure to include features with higher relative importance in our model, this may not guarantee the random forest model can accurately predict whether reviews are fake or genuine, as we have not considered how variables interact with one another. Using 80% of the Amazon reviews from He et al. (2022a) for training and the remaining 20% for testing out-of-sample performance (referred to "simple test-train"), we calculated the relative importance and improvements in model performance for different combinations of features listed in Table 1.

We found that including sentiment, parts of speech tagging or TF/IDF (see Table 1 above) features did not increase the random forest model's predictive power, and these features were

dropped as a result.<sup>26</sup> In the remainder of this section, the predictive model only includes network and metadata features ("All features" refers to these two groups taken together).

### Selecting a model

We tested three different types of machine learning models: Random Forest (RF), Support Vector Classifier (SVC) and Logistic Regression (further technical details are provided in Appendix 1). These models were measured against the following performance metrics:

- Accuracy: Percentage of correctly classified reviews
- Recall: Percentage of correctly classified fake reviews among all reviews classified as fake
- Precision: Percentage of correctly classified fake reviews among all fake ones
- F1 score: Harmonic mean between recall and precision (these two metrics are inversely related: increasing recall will reduce precision, and vice versa).
- Area Under Curve: Plotting the false positive rate (FPR) against the true positive rate (TPR), and taking area enclosed

These goodness-of-fit metrics are measures of discriminatory power (i.e. how well our model's predictions matches the fake and genuine labels in our dataset of reviews) that take values between 0 and 1. For example, if a model's recall is 0.9, this means the model produces more accurate predictions of whether a review is genuine or fake (compared to a different model whose recall is 0.3). A value close to 0.5 indicates a model is not, in a statistical sense, very different from a coin toss in terms of deciding whether a review is fake or not.

Table 3 lists the performance metrics for the random forest classifier (our preferred model) for each group of variables. This table suggests that the random forest classifier achieves a high goodness-of-fit across all variable groupings. Detailed tables for the other classifiers tested on the simple test-train split as well as more robust cross-validation strategies are included in Appendix 1.<sup>27</sup> The random forest classifier also performed at a high level using these strategies, scoring higher than all other classifiers tested.

---

<sup>26</sup> This finding supports the idea that fake reviews have become more sophisticated over time (as more obvious features of reviews such sentiment and text are no longer good predictors of whether reviews are genuine or fake).

<sup>27</sup> More specifically, we use a five-fold cross-validation strategy. This involves dividing the data into five folds, training the model on four of these folds, and testing the quality of adjustment on the remaining fifth fold, all the while varying the training and testing folds in each iteration. Performance metrics are then averaged across all five folds.

**Table 2: Out-of-sample prediction performance, random forest classifier (simple train-test)**

Features	AUC	Accuracy	Recall	Precision	F1 score
Network	0.99953	0.99968	0.99948	0.99973	0.99965
Metadata	0.99998	0.99998	0.99998	0.99997	0.99998
All features	0.99998	0.99998	0.99998	0.99970	0.99997

**Table A: Out-of-sample prediction performance, support vector classifier (simple train-test)**

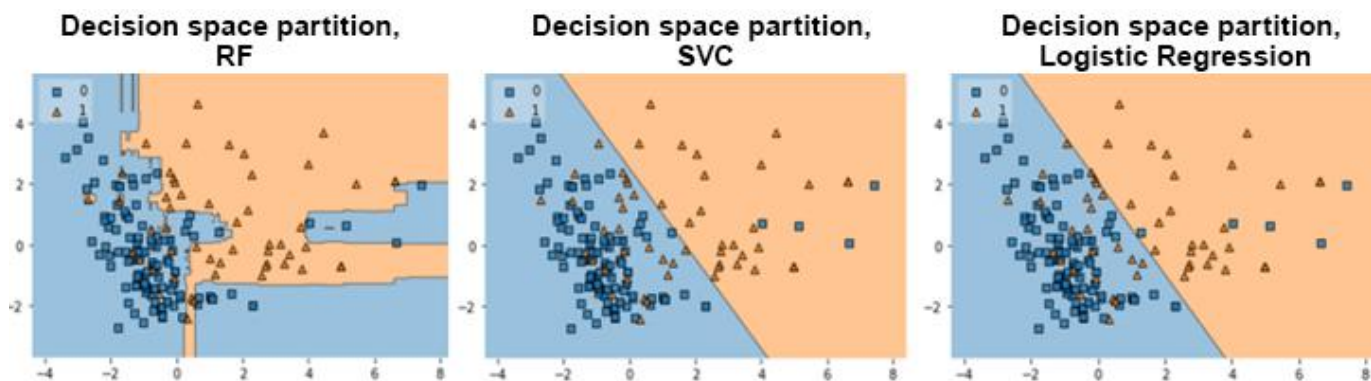
Features	AUC	Accuracy	Recall	Precision	F1 score
Network	0.76855	0.81078	0.76873	0.79661	0.77942
Metadata	0.69707	0.75470	0.69655	0.73274	0.70718
All features	0.79386	0.82845	0.79386	0.81579	0.71181

**Table A1.3: Out-of-sample prediction performance, logistic regression classifier (simple train-test)**

Features	AUC	Accuracy	Recall	Precision	F1 score
Network	0.77033	0.81192	0.77049	0.79761	0.78102
Metadata	0.70204	0.75703	0.70217	0.73522	0.71181
All features	0.80191	0.83410	0.80191	0.82139	0.80999

One way of visualising the performance of different model classifiers is to transform each review into a single number based on the review's specific characteristics, then plot a sample of review datapoints with the decision "rule" (shown as a line) used by the model to separate genuine and fake reviews. This is depicted in Figure 2 below:<sup>28</sup>

**Figure 2: Comparison of model partition of the decision space for sample of 150 reviews**



As Figure 2 shows, a Random Forest classifier can implement a more "convoluted" decision rule, and as a result this classifier can more accurately identify fake reviews that are similar to genuine reviews, and vice versa, compared to the other two classifiers.<sup>29</sup>

## Training the model

To train the aforementioned predictive model how to identify fake reviews which would allow us to estimate prevalence across popular UK e-commerce platforms, we used data from He et al. (2022a).<sup>30</sup> This data was taken from private Facebook groups in which sellers buy reviews. This means we know for certain which products are buying fake reviews. Reviews (including those later removed by Amazon) were collected for around 1,500 unique products across 26 different Facebook groups as well as 2,714 competitor products between October 2019 and November 2020. 34% of the reviews in this dataset belong to products from sellers that had

<sup>28</sup> In Figure 2, triangles are reviews classified as fake and squares are reviews classified as genuine. To create this figure, we first need to "reduce" the number of features from 12 to 2 using a technique called Principle Component Analysis.

<sup>29</sup> Random forests can present a more advanced decision rule than SVC and logistic regression because they are capable of modelling complex, nonlinear relationships between the input features and the output variable. This helps the model to capture a wide range of patterns in the data. In contrast, SVC and logistic regression assumes a linear relationship between the input features and the output variable, and can only model simple, linear decision boundaries. Therefore, random forests are often more powerful and flexible than logistic regression, especially when dealing with complex datasets that contain nonlinear relationships.

<sup>30</sup> This is the same dataset used to compare different model classifiers in the previous section.

purchased fake reviews for these products, and 9% of these reviews were removed by the platform at some point in the time period for which reviews were collected.

## Amazon reviews data

After the predictive model was trained, we then applied the model to the dataset of online product reviews and metadata from Ni, Li and McAuley (2019)<sup>31</sup>, which includes 233.1 million reviews from Amazon spanning May 1996 – October 2018. This was done to estimate the prevalence of fake reviews. Our analysis focused on three sectors of products that are widely purchased online: home and kitchen, electronics and sports and outdoors.<sup>32</sup> Table 2 below provides the descriptive statistics for each of these categories.

**Table 3: Descriptive statistics of reviews for selected sectors**

Product category	Number of unique products	Number of product reviews	Average star rating	Percentage of 1-star reviews (%)	Percentage of 5-star reviews (%)
Home & kitchen	189,172	777,242	4.33	21.8	1.8
Electronics	160,052	728,719	4.17	23.1	3.0
Sports & outdoors	104,687	332,447	4.37	21.4	2.0

## Predictive model findings

With the models trained and compared, we then created all the variables described in Table 1 for each of the sectors, then deploy the optimal specification (Random Forest) on the unlabelled Amazon dataset from Ni, Li and McAuley (2019)<sup>33</sup>. We did this to predict whether reviews are genuine or fake (details of parameter configuration are provided in Appendix 1).

<sup>31</sup> Ni, Jianmo, Jiacheng Li, and Julian McAuley. 2019. ‘Justifying Recommendations Using Distantly-Labeled Reviews and Fine-Grained Aspects’. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 188–97. Hong Kong, China: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1018>.

<sup>32</sup> These align with the product categories later used in our experiment.

<sup>33</sup> Ni, Jianmo, Jiacheng Li, and Julian McAuley. 2019. ‘Justifying Recommendations Using Distantly-Labeled Reviews and Fine-Grained Aspects’. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 188–97. Hong Kong, China: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1018>.



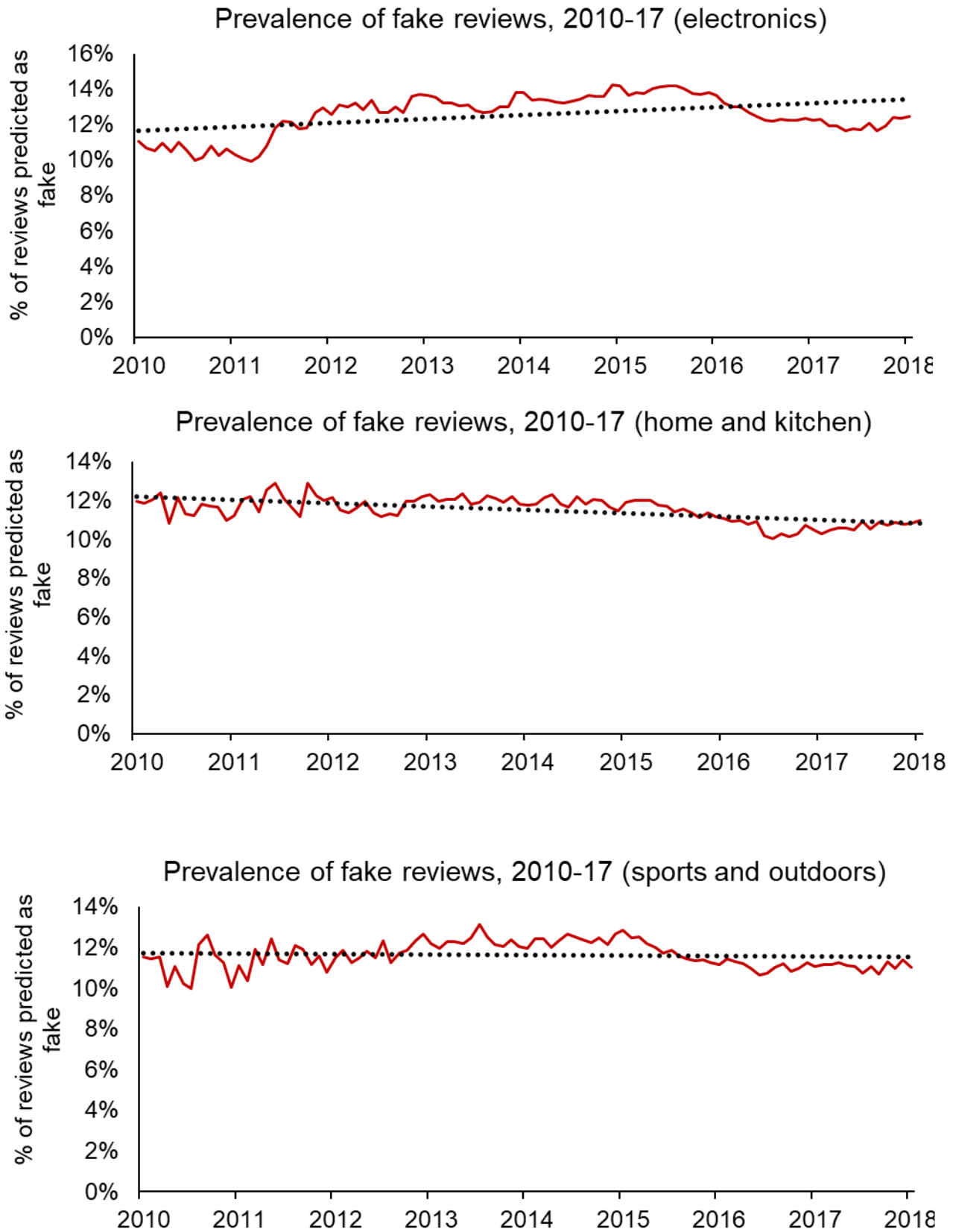
Table 4 below presents the results in terms of predicted proportion of fake reviews of the deployment of the RF on the Amazon dataset considering network features.

**Table 4: Proportion of Amazon reviews predicted as fake by the RF algorithm, by sector**

Category	Percentage of reviews predicted as fake (%)
Home & kitchen	11.1%
Electronics	12.9%
Sports & outdoors	11.5%

The dataset includes reviews collected over time which allows us to carry out time series analysis to see how prevalence has changed over time for reviews in the three product categories specified in Table 4. Graphs for each category are presented in Figure 3. There does not appear to be a clear trend in the prevalence of fake reviews over time: the proportion of fake reviews posted for consumer electronics has slightly increased, the proportion for home and kitchen products has slightly decreased and the proportion for sports and outdoors goods has remained relatively constant.

Figure 3: Prevalence of fake reviews across time for different product categories



## Other e-commerce platforms

Finally, we estimated the prevalence of fake reviews for eight other popular UK e-commerce platforms based on a dataset of around 300,000 reviews, with the results presented in Table 5. In general, we found that the prevalence of fake reviews was higher for these platforms than for Amazon (between 25-35% of all reviews). However, because reviews are much less common on non-Amazon platforms, the smaller number of users reviewing multiple products may lead to more biased estimates, and thus our estimates for these platforms are not directly comparable with our main estimate.

**Table 5: Proportion of reviews predicted as fake (other popular e-commerce platforms)**

Platform	# Products	# Reviews	% estimated fake reviews (network)	% estimated fakes (text alone)
Platform 1	979	112,135	34.6%	31.5%
Platform 2	1,115	85,216	38.3%	36.4%
Platform 3	1,552	14,323	33.8%	27.9%
Platform 4	916	7,718	46.1%	26.4%
Platform 5	895	5,742	42.7%	29.5%
Platform 6	151	2,134	32.5%	17.6%
Platform 7	671	3,471	37.2%	31.7%
Platform 8	881	4,559	36.4%	21.2%

## Results and implications

Our predictive model uncovered several key findings:

1. There is an estimated prevalence of fake reviews of approximately 11% to 15% on popular UK e-commerce platforms, but there are slight differences based on the category of product purchased.<sup>34</sup>

<sup>34</sup> The upper boundary of this range (15%) was calculated by weighing the proportion of fake reviews across all nine platforms by the number of reviews collected for each platform.

2. Network features are the strongest predictor of fake reviews, suggesting that automated means of review moderation should place at least as much emphasis on examining the characteristics of reviewers as the content of the review itself in isolation.
3. Review metadata such as language and sentiment have some predictive power in identifying fake reviews, but these features taken alone are generally not good predictors of a review's authenticity.

However, there are several important limitations of these estimates: our test dataset only includes product reviews up to 2018 (and we make the assumption that the approach consumers take to writing genuine reviews has not changed over the past five years), and Amazon itself has also taken steps to remove a number of reviews flagged by automated software as fake. Finally, building a robust network requires collecting data on a significant proportion of reviews posted on a platform, which means our estimates for the other eight e-commerce platforms in Table 5 are suggestive and should be treated with significantly more caution.

# Estimating the impact of fake reviews on consumers

The predictive model has outlined that the number of fake reviews has increased in line with overall review activity with 11% to 15% of reviews being fake. To understand whether these misleading reviews impact how consumers make purchases online, we designed and implemented an experimental online shopping task and follow-up questionnaire.

Our primary investigation tested the impact of fake reviews by assessing whether fake reviews increased the likelihood that a product was purchased and the amount of money that someone was willing to spend on it. We also explored whether there was a difference in impact when the fake reviews were strong (i.e., less well written and therefore more obviously fake), compared to subtle (i.e., more well written and therefore less obviously fake).

Furthermore, a non-regulatory intervention, referred to as “informed silence” (please see a screenshot of the informed silence text on p.29), was introduced to examine whether the impact of the fake reviews can be prevented by alerting participants to the possible existence of such reviews on the platform. Our supplementary investigation then explored differences in results depending on the product type participants were exposed to, their demographic background, as well as the influence that the fake reviews and intervention had on additional variables such as the confidence that consumers had in the platform.

The experiment offered an opportunity to understand the impact of fake reviews on UK consumers by:

- Isolating the impact of fake review text by controlling for the influence of price and star-rating, which ensures that any effect found is due to the fake review text and not external factors.
- Testing the impact of fake reviews based on how well-written they were (strong vs. subtle) and a non-regulatory intervention (informed silence) allows us to assess if the content of fake reviews matter, as well as whether a simple measure can be introduced to counteract any impact of fake reviews on consumers.
- Creating an online shopping platform that mimics the practice of shopping and reading reviews on a popular e-commerce site ensures a realistic experience.
- Focusing on participants that shop online and are resident in the UK means that conclusions drawn reflect the impact of reviews on UK consumers.

### Hypotheses

Based on findings from previous research outlined in the introduction, we predicted that exposure to fake reviews and the non-regulatory intervention would impact participants' purchasing decisions. It was hypothesised that:

- H1: participants' purchasing decisions will be impacted when a product has subtle fake reviews.
- H2: participants' purchasing decisions will be impacted when a product has strong fake reviews.
- H3a: participants' purchasing decisions will be impacted when an informed silence textbox is displayed.
- H3b: the effect of the informed silence intervention will be larger for the strong, compared to subtle, fake review condition.

### Participants and sampling strategy

A total of 4822 participants were recruited through the online recruitment platform Prolific (a breakdown of participants demographic characteristics is provided in Table 11). A convenience sampling technique was utilised whereby participants took part in the experiment on a first come, first serve basis and the only restriction placed on participation was the requirement to be a UK resident. Our sample was representative of the UK adult population by ethnicity and gender, but not on other characteristics such as age.<sup>35</sup>

Participants were able to complete the experiment using mobile devices, tablets, or desktops and they were paid £2.25 in return for their participation. Participants did not receive the product that they chose in the experiment or any payment that depended on the decisions that they made in the experiment.<sup>36</sup>

Individuals that did not pay attention, as determined by failing two attention checks within the experiment (see Materials for more details), were excluded and did not receive their payment. Four participants were excluded for this reason, leaving a final sample of 4818. A small pilot study with 50 participants was launched prior to the full experiment which confirmed that there were no technical difficulties that needed to be resolved.

---

<sup>35</sup> Our sample had a greater distribution of young and middle-aged adults than the UK population as a whole. However, evidence suggests that young and middle-aged people are disproportionately more likely to shop online than other age groups, and thus are more appropriate as the target audience for the experiment.

<sup>36</sup> We do not believe that this impacted the external validity of our study, please see p. 44 for further details.

## Experimental design

The experiment had a between-subjects design, which means that each participant was only exposed to one condition. There were six different conditions that participants could be allocated to<sup>37</sup>, and they varied depending on two factors:

- The type of reviews present (genuine, subtle fake, or strong fake)
- The non-regulatory intervention (no intervention or informed silence)

The first group (Group 1) was the control group while the remaining groups made up the treatment groups. The purchasing decisions for each treatment group were compared to that of the control group to isolate the impacts of the different treatments (the fake reviews and the intervention). The six different groups are outlined in Table 6 below.

**Table 6: List of experiment groups**

Group		Type of reviews	Non-regulatory intervention
Control	Group 1	Genuine	No intervention
Treatment	Group 2	Subtle	No intervention
	Group 3	Strong	No intervention
	Group 4	Genuine	Informed silence
	Group 5	Subtle	Informed silence
	Group 6	Strong	Informed silence

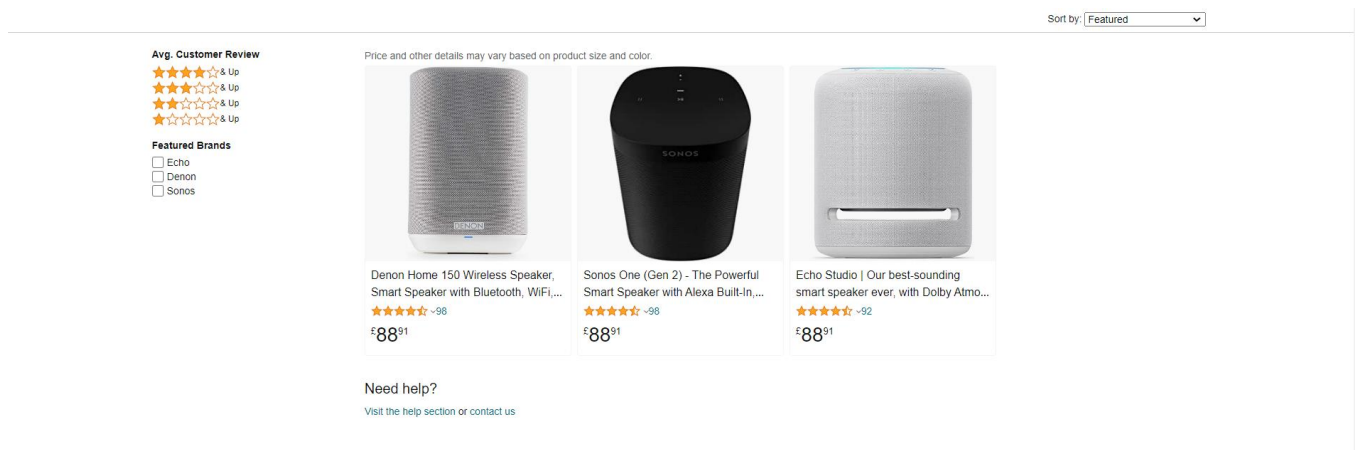
Participants were randomly assigned to one of 11 product types (see Materials for more details). Each participant saw three different products from their assigned product type (for example, three different keyboards) displayed on the online retail platform and the ordering of these were randomised (see Figure 4 below). Within the fake review conditions, one of the three products had fake reviews while the remaining two had genuine reviews. The specific product with fake reviews displayed was also randomly assigned. Within the informed silence condition, all three products displayed the banner regardless of whether they had been assigned genuine or fake reviews.

---

<sup>37</sup> Participants were assigned to each group using the least-fill method, which is when participants are assigned randomly and equally to each group.

The probability that a product was purchased was the variable of interest to be observed (i.e., the dependent variable). Further variables explored included participants' WTP and consumer trust and future behaviour, as well as how these varied depending on demographic characteristics. We also collected information regarding the number of products participants clicked on, the time that they spent reading reviews, how many pages of reviews that they read, as well as how they filtered the reviews.

**Figure 4: Screenshot of the product overview page**



## Materials

### Online retail platform and product categories

**Online retail platform:** The online retail platform was developed as a standalone React web app and closely resembled the layout, features and user experience of a popular UK e-commerce site (so participants would feel more comfortable navigating within a familiar environment). Three key pages were created, a product overview page, a product-specific page, and a shopping cart confirmation page. Screenshots of the platform are included in Appendix 3.

The first page that participants saw was the product overview page which displayed the three different products from the same product type. Participants were able to see the name, brand, and price of the product, the product's star rating, and the total number of reviews that it had. The price for all three products within each type were fixed to an estimated average cost of a product within that category (determined by examining the price range for that product on Amazon). The star rating for all three products were also fixed with each product having a star rating of 4.5 stars. Price and star rating are typically the two most visible characteristics of a product when a consumer searches for a product, and controlling for these characteristics was intended to encourage participants to review product-specific information on each individual product page.<sup>38</sup> However, in the real world, consumers rarely choose between products with

<sup>38</sup> More specifically, on most e-commerce platforms, when a consumer types in a specific search keyword, a list of products matching the keyword is returned, along with the price and star rating for each product. Thus, the consumer looks at the product price and star rating as two factors in deciding which product page to visit (and eventually which product to purchase).



identical characteristics, so our experimental design served to capture how consumers weighed the content of reviews as one factor in their broader decision-making framework.

Participants were able to click on each of these products and navigate back to the list of products as needed. Once participants clicked on any of the products, they were directed to the product-specific page. The product-specific page contained a detailed description of the product (including key features, technical details, and frequently asked questions) as well as a set of 25 user reviews.

When participants decided to purchase a product by adding it to their shopping cart, they were directed to the shopping cart confirmation page. Once they confirmed their purchase on this page, they were directed to the next stage of the experiment and were not able to navigate back to any of the previous pages.

**Products:** Participants were randomly assigned to one of 11 different product types. The following product types were selected: kettle, iron, vacuum cleaner, desk chair, Bluetooth headphones, keyboard, mobile charger, smart speaker, skincare product, yoga mat, and reusable water bottle. These categories were chosen because they satisfied a range of criteria to ensure that they would be desirable to purchase for a wide range of participants. The criteria used was that the products are commonly purchased online, commonly sold on e-commerce websites, physical and non-perishable, gender balanced, intended for use by adults, and distinguishable to the extent that perceived quality would play a role in decision-making. By including these different product types, we were able to ensure that any significant findings were due to the manipulations of the experiment and any conclusions drawn will be applicable to a wide range of commonly purchased goods.

Within each product type, three different products (for example, three keyboards) were selected and displayed to participants. These were chosen from real products sold on the Amazon platform. The products chosen were similar in price and star rating to ensure they were comparable and that participants would be likely to consult products reviews to determine which one to purchase.

### Product reviews

**Type and strength of reviews:** Participants were able to read 25 product reviews for each of the three products they were able to purchase. The 25 reviews were spread across five pages containing five reviews each. There were two types of reviews, genuine and fake. Participants in the genuine review conditions were exclusively shown genuine reviews for all products while participants in the fake review conditions saw both genuine and fake reviews. The fake reviews that they were exposed to were present for one of the three products and made up 20% of the reviews for that product.<sup>39</sup> The genuine reviews were real reviews taken from the product's page on Amazon. To ensure that the reviews were genuine, we did not select any that met the criteria that we used when writing our own fake reviews. Furthermore, we passed the reviews

---

<sup>39</sup> To reach this 20% estimate, we averaged our estimated prevalence of reviews predicted as fake (11-15%) with a commonly-cited estimate from Fakespot of 31% (<https://risnews.com/report-30-online-customer-reviews-deemed-fake>). We then adjusted the average to ensure fake reviews were evenly distributed across all review pages for each product.

that we did select through our detection model and replaced any that the model flagged as fake.

The fake reviews were manually written by the research team building on the criteria used by Akesson et al<sup>40</sup>. (2022). The fake reviews were positive and promoted the product that they were reviewing. Participants assigned to the fake review treatment either saw subtle or strong versions of these. Prior to launching the experiment, we received feedback on the fake reviews from a representative at Which?, the UK consumer advocacy group. We used the following criteria to write our fake reviews:

**Subtle fake review text:** Deliberate inclusion of one of the following elements that might raise suspicion: overly vague or generic language, exaggerated language, several reviews left on the same date or the same reviewer leaving two reviews. The following is an example of a subtle fake review:


“Beautiful, sturdy, and meaningful! MAJOR transformation and definitely worth the money. [reviewer name is "ProductReviewer"]

**Strong fake review text:** Deliberate inclusion of at least two of the following elements: excessive capitalisation or punctuation, repetitive phrases and formatting, reviews covering completely different products than the product listed on the page, acknowledgement of financial compensation for positive reviews. In addition, the reviews were written by the open-source text generation model GPT-2 (to mimic fake reviews written by bots). The following is an example of a strong fake review:

“A must have machine for your everyday use... very easy to use and great quality output... I was hesitant but it's simple and easy and value for money!!! HOW do I GET MY £20 VOUCHER?”

**Non-regulatory intervention.** The non-regulatory intervention that we selected to examine was informed silence. Within this intervention condition, participants were able to see a text box displayed prominently on all product-specific pages. The text stated that steps had been taken to moderate misleading content (such as misleading customer reviews) on the platform. Please note that no specific reviews were flagged in this condition. The text is displayed in Figure 5 below.

**Figure 5: Screenshot of informed silence text box**



⚠ **Steps have been taken to moderate misleading and harmful content on this website.**  
Thank you for shopping with us. We'd like to let you know that we've taken steps to moderate misleading and harmful content on this website. This includes misleading customer reviews, misleading advertisements and counterfeit goods.

---

<sup>40</sup> Akesson, Jesper., Robert W. Hahn, Robert D. Metcalfe, and Manuel Monti-Nussbaum. 2022. “The Impact of Fake Reviews on Demand and Welfare”. Unpublished manuscript, July 20 2022, typescript.

### Post-experiment questionnaire

#### Willingness to pay

Willingness to pay (WTP) refers to the maximum price a specific consumer would pay for a product or service<sup>41</sup> (Gall-Ely 2009). We created a task to assess participants' stated WTP for the product that they chose to purchase.<sup>42</sup> Participants were provided with information regarding the price of the product type that they had been allocated to. Specifically, they were told an estimate of what the lowest, average, and highest prices were for that product. These figures were derived from looking at the first five pages of search results for that product type on Amazon. Participants were then directly asked how much they would be willing to purchase the product for and were given several price intervals to choose from. The price intervals were equally large and covered the entire range of prices between the lowest and highest price described to participants.

#### Experiment follow-up

A survey followed the main experiment and the willingness to pay task. It had three sections.

The first section contained questions regarding the specific purchase that participants made in the online shopping task. The section was designed to gain further insight into why participants chose to purchase the product that they did and how likely they would be to make that purchase in real life. They were also asked to score four product reviews in terms of their helpfulness, credibility, and relevance.

The second section concerned participants' general shopping behaviour and preferences. Participants were asked to rank different factors (such as star rating and number of reviews) in order of importance when they shop online. They were also asked to estimate the number of product reviews (in real life) that are fake as well as rate what their response would be to different scenarios involving online shopping and fake reviews.

The final section included demographic questions, which were designed to gain further insight into the make-up and background of the participants.

**Comprehension and attention checks:** Two comprehension and two attention checks were created. The comprehension check was designed to ensure that participants understood what we were asking of them. They were asked two simple questions, one regarding the experiment instructions and one regarding the product type they had chosen to purchase. For the former, participants had to correctly answer this question before they could proceed with the experiment while for the latter, we only checked whether a participant's response lined up with the product type they had been allocated to if their remaining behaviour in the experiment was unusual (such as completing the experiment unusually quickly). The answers to the comprehension questions were multiple choice and participants were not automatically excluded for not answering correctly. On the other hand, the two attention checks were

---

<sup>41</sup> Gall-Ely, Marine Le. 2009. 'Definition, Measurement and Determinants of the Consumer's Willingness to Pay: A Critical Synthesis and Directions for Further Research'. Post-Print, Post-Print, June. <https://ideas.repec.org/p/hal/journal/hal-00522828.html>

<sup>42</sup> Participants were asked to provide their WTP for a generic example of the product type they were assigned to, not the specific product they added to their shopping cart in the online platform.

designed to ensure that participants were paying attention during the experiment and not just clicking random answers. These checks were integrated in the post-task questionnaire and instructed participants to pick specific answers. If participants failed both attention checks, we excluded their data from the experiment and withheld payment.

### Procedure

Prior to joining the experiment, participants were presented with an information sheet and a privacy notice. These documents set out the aim of the experiment, their rights as a participant, and our obligations when handling their data. Upon reading these documents, participants were asked to fill out an online consent form. Once participants had given their informed consent, they were given the instructions for the first part of the experiment.

Participants were instructed that they would be presented with three different products of the same type. They were told that they could click on each of the three products and review a product's image, price, star rating and reviews. Participants were then instructed to consider each of the three products and subsequently, select one of the items to purchase. They were able to go back and forth between products and there was no time limit imposed on their shopping experience.

Participants then completed the experiment and answered the post-experiment questionnaire. Once completed, participants were thanked for their participation, provided with debriefing information, and redirected back to Prolific. The experiment was expected to take approximately 15 minutes to complete. Figure 6 below shows how participants proceeded through the experiment.

**Figure 6: Diagram of main experimental procedure**



### Descriptive statistics

In total, 4,818 participants took part in the experiment. Table 7 below presents the demographic characteristics of the sample. We also carried out checks to ensure that participants were randomly assigned to each of the six experiment groups (based on their demographic characteristics), with the results presented in Appendix 4.

Table 7: Demographic breakdown of sample

Variable	Number of participants	Percentage of total participants (%)
Income		
Up to £9,999	636	13.2
£10,000 - £24,999	1332	27.6
£25,000 - £49,999	1908	39.6
£50,000 - £74,999	408	8.5
£75,000 - £99,999	112	2.3
£100,000 or more	63	1.3
Prefer not to answer	359	7.5
Highest level of education completed		
Less than primary school / primary school not completed	3	0.1
Primary	8	0.2
Secondary	961	19.9
Vocational	812	16.9
Undergraduate	1960	40.7
Postgraduate	996	20.7
Prefer not to answer	78	1.6

Age		
18-24 years old	542	11.2
25-34 years old	1603	33.3
35-44 years old	1279	26.5
45-54 years old	717	14.9
55-64 years old	453	9.4
65 years or older	200	4.2
Prefer not to answer	24	0.5
Sex		
Female	2399	49.8
Male	2378	49.4
Intersex	5	0.1
Prefer not to answer	36	0.7
Ethnicity		
White	4119	85.5
Mixed/multiple ethnic groups	153	3.2
Asian/Asian British	323	6.7
Black / African / Caribbean / Black British	131	2.7
Other ethnic groups	41	0.9

Prefer not to answer	51	1.1
Frequency of online shopping		
more than once a week	890	18.5
about once a week	1294	26.9
several times a month	1545	32.1
about once a month	788	16.4
once in a few months or longer	301	6.2
Frequency of Amazon purchases		
more than once a week	473	9.8
about once a week	831	17.2
several times a month	1499	31.1
about once a month	1121	23.3
once in a few months or longer	805	16.7
never	89	1.8

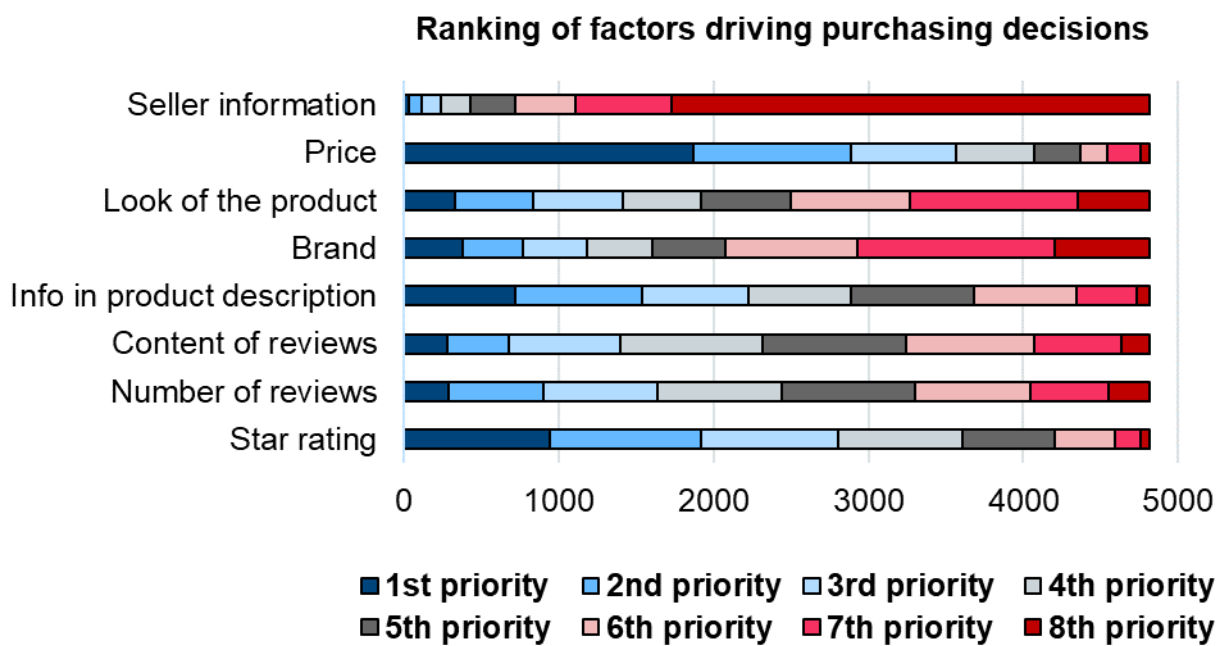
### Participant engagement

Respondents spent an average of 12.3 minutes completing the entire experiment, including the post-task questionnaire. Out of the 4,818 participants that completed the experiment, 3,253 (67.5%) read reviews for all three products they were presented with (there was a less than 1% difference for participants who viewed the informed silence intervention compared to participants who were not assigned to the intervention). This shows that participants generally engaged with the experiment and shopping experience as expected and that the product reviews played a part in their purchasing decisions. However, since not all participants viewed reviews for all three products, it is likely that some made their decisions based on variations in

brand, image and item description. This is in line with the findings presented in Figure 7 below, which show that content of reviews is not often the primary driver behind a purchase.

Figure 7 below illustrates the aspects that participants reported finding important when shopping online. Participants most commonly ranked price, star rating, and information in the product description as the top factors that affected their purchasing decisions, while seller information was reported by most as the least important. In relation to product reviews, participants most frequently reported the number and content of reviews to be in fourth, fifth, or sixth place of importance when shopping.

Figure 7: Factors driving purchasing decisions





## Findings

We present our findings for the impact of fake reviews and the non-regulatory intervention on consumers shopping online in the following sections. We first present summary statistics of how participants evaluated the types of reviews on different measures. Subsequently, we describe the impact of the fake reviews and intervention on the product participants chose to purchase in our general regression analysis. We then examine how these effects vary by product category in our product-specific regression analysis. Finally, we explore how fake reviews and the intervention impact on consumer trust and their future purchasing intentions in our supplementary regression analysis. For all regression analyses, we restricted our sample to individuals who viewed all three products they were shown (3,253 participants) as those who only viewed one or two products were unable to effectively compare reviews across products.

### Consumer assessment of fake reviews

**Table 8: Participant evaluation of reviews<sup>43</sup>**

Review displayed	Mean helpfulness rating (self)	Mean helpfulness rating (others)	Mean credibility rating	Mean relevance rating
Genuine	3.9 [0.9]	4 [0.9]	3.8 [0.9]	4 [1.0]
Strong fake	2.2 [1.1]	2.2 [1.1]	2.8 [1.2]	2.2 [1.1]
Subtle fake	3 [1.4]	2.8 [1.4]	2.4 [1.4]	3.1 [1.3]

Table 8 showcases the average ratings, on a 5-point Likert scale<sup>44</sup>, that participants gave on four measures for genuine, subtle fake and strong fake reviews in the experiment follow up survey. Participants were presented with four reviews (two genuine, one subtle fake, and one strong fake) that they had not seen before. For each review, they were then asked to rate the extent that they themselves found the review helpful, how helpful other consumers wishing to purchase the product would find it, how credible they found the review, as well as how relevant the review was to their purchasing decision. Across measures, the average ratings for the genuine reviews were higher than for the fake reviews, indicating that participants could discern some differences between the review types when they were presented in isolation.<sup>45</sup>

<sup>43</sup> Figures in brackets show the standard deviation.

<sup>44</sup> For more details on the Likert scale please refer to Questions 11-14 in Appendix 2.

<sup>45</sup> It may be easier for consumers to detect fake reviews when a review is seen in isolation, compared to fake reviews presented in combination with other product reviews.

However, it is important to note that participants were only asked to rate one or two reviews of each type.

### Impact on purchasing decisions

Regression analysis<sup>46</sup> was performed to create a statistical model of consumer decision-making based on the data gathered from the experiment. Here, our primary variable of interest was whether a consumer purchased a product or not. The degree to which changes in this variable observed during the experiment are explained by the independent variables (also known as explanatory variables) was then fitted based on the aggregated outputs of the experiment. In this case, the independent variables included the type of the review a participant was exposed (genuine, subtle fake or strong fake) and whether they were exposed to the informed silence textbox. The coefficients assigned to each explanatory variable by the model during this process represent the observed impact that variable had on decision-making relative to the control group. For example, for the consumer purchase variable, a coefficient of 0.03 on the subtle fake review explanatory variable means that this treatment made participants 3% more likely to purchase a product with fake reviews relative to the control group. More detail on the final model specifications can be found in Appendix 6.

#### General regression analysis

We first ran a general regression analysis to estimate the impact of reviews on consumer decision-making and the results are presented in Table 9 below. Consumer decision-making was measured by the dependent variable product chosen, i.e., which product an individual chose to purchase on the online shopping platform. We also carried out two different robustness checks. We tested a range of different controls for individual-level characteristics, and our regression estimates were consistent in sign and magnitude across all specifications. The tables for both sets of robustness checks are listed in Appendix 4.

**Table 9: Impact of fake reviews and informed silence on the probability of choosing a product (across all product categories) in the online shopping task.**

Treatment	Impact on consumer likelihood to purchase product
Group 2 (Subtle fake reviews + no intervention)	0.0314**
Group 3 (Strong fake reviews + no intervention)	-0.0534***
Group 4 (Genuine reviews + informed silence)	-0.0009

---

<sup>46</sup> A regression analysis is a statistical method that allows you to examine the relationship between two or more variables of interest.

Group 5 (Subtle fake reviews + informed silence)	-0.0105
Group 6 (Strong fake reviews + informed silence)	0.0178
Observations	3,255
R <sup>2</sup> <sup>47</sup>	0.0015

Note: \* =  $p < 0.1$ , \*\* =  $p < 0.05$ , \*\*\* =  $p < 0.01$

In our baseline specification, the results for Group 2 show that people who were exposed to subtle fake reviews were 3.1% more likely to purchase the product with the fake reviews, compared to individuals who saw only genuine reviews for that product. The results for Group 3 show that people who were exposed to strong fake reviews were instead 5.3% less likely to purchase the product with the fake review, compared to individuals who saw only genuine reviews for that product. As such, these findings are in line with our initial hypotheses H1 and H2, which stated that the proportion of participants that purchase a product will be different in the subtle fake, and strong fake, compared to the genuine review condition.

The findings for Group 4, 5, and 6, demonstrate the impact on consumer behaviour when the non-regulatory intervention informed silence is introduced. Regardless of the type of reviews present, there was no significant effect of the intervention on the product purchased by participants. As such, these results are not in line with our initial hypothesis H3a, which stated that the proportion of participants that purchase a product when the informed silence textbox was displayed will be different in the intervention, compared to no intervention, condition. Furthermore, they are also not in line with our hypothesis H3b, which stated that the effect of the informed silence intervention would be larger for the strong, compared to subtle, fake review condition.

Overall, these findings indicate that fake reviews influence consumer decision-making by making consumers more or less likely to purchase a certain product, depending on the strength of the fake review. The reduction in the likelihood of purchasing a product with strong fake reviews, indicates that consumers are not only able to recognise more extreme forms of misleading content, but they are also pushed towards purchasing a different product when they have identified it (in particular, when asked to explain their decision to purchase a specific product, around 10% of participants specifically mentioned the strong fake review referencing the wrong product or sounding fake). This interpretation is supported by the ratings presented in Table 8 whereby the strong fake review was rated as less helpful, credible, and relevant in almost all cases compared to the other two types of reviews.

---

<sup>47</sup> R-squared (R<sup>2</sup>) is a statistical measure that represents the proportion of the variance for a dependent variable that is explained by an independent variable or variables in a regression model. The closer R<sup>2</sup> is to 1 the more variation is explained by the variables within the model. For behavioural studies low R<sup>2</sup> is to be expected given the wide range of unobservable cognitive and human experience factors which feed into an individual's decision making.

On the other hand, the increased likelihood of purchasing a product when it had subtle fake reviews suggests instead that consumers are not only unable to identify subtler forms of misleading content, but that these reviews are also more persuasive than genuine ones. This could be because genuine reviews are i) not always positive in nature and may be more likely to reflect a nuanced consumer experience of a product and ii) not guaranteed to be written well (for example, contains grammatical and spelling errors).

The findings in relation to the non-regulatory intervention informed silence suggest that alerting consumers that measures have been taken to deal with misleading content does not counteract the influence of fake reviews on the products that individuals choose to purchase. This means that heightened awareness about the presence and prevalence of misleading content on e-commerce sites does not seem to aid a consumer in their ability to detect fake reviews. This result may be driven in part by the relatively neutral language of the informed silence warning: “moderate” does not provide a clear course of action taken by the platform, and “misleading customer reviews” may not be perceived by consumers as equivalent to “fake reviews”. However, informed silence is only one type of non-regulatory intervention and there may be other interventions, which are easy and cost-effective for platforms to implement, that counteract the impact of fake reviews.

**Demographic characteristics.** As part of the main regression analysis, we also tested whether there were any differences in the impact of the fake reviews on different demographic variables. These were: (i) sex, (ii) age, (iii) income, (iv) education, (v) ethnicity, (vi) disability, and (vii) online shopping frequency and the results of the analysis are presented in Appendix 4. No significant effects were found, which suggests that there are generally no differences between UK consumers in how they are impacted by fake reviews. However, our demographic analysis was based on the people that participated in our survey and therefore some groups may have been under or oversampled.

### Product-specific regression analysis

Secondly, we ran a product-specific regression analysis, and the results are presented in Table 10 below. This model was identical to the general regression analysis outlined above except that it assessed the impact of fake reviews based on the product category that participants were exposed to.

**Table 10: Impact of fake reviews and informed silence on the probability of choosing a product (by product category).**

Product category	Electronics	Household	Health and beauty
Group 2 (Subtle fake reviews + no intervention)	0.0780***	0.0014	0.0113
Group 3 (Strong fake reviews + no intervention)	-0.0822***	-0.0238	-0.0570*

Group 4 (Genuine reviews + informed silence)	0.0014	-0.0019	-0.0036
Group 5 (Subtle fake reviews + informed silence)	-0.0260	-0.0005	-0.0038
Group 6 (Strong fake reviews + informed silence)	0.0274	0.0079	0.0191
Observations	1192	1210	851
R <sup>2</sup>	0.0051	0.0002	0.0012

Note: \* =  $p < 0.1$ , \*\* =  $p < 0.05$ , \*\*\* =  $p < 0.01$

The table above shows that people who were exposed to subtle or strong fake reviews for an electronic product were 7.8% more likely and 8.2% less likely to purchase that product respectively. A similar negative relationship was found between exposure to strong fake reviews and the product chosen for health and beauty items, however a significant effect was not found for subtle fake reviews. The significant findings did not extend to any products within the household category.

It is possible that this effect may in part be driven by price, as electronics products tend to be more expensive than products in the other two categories. To test for this effect, we estimated our main specification for the subsample of four products in the online shopping task priced above £80 (vacuums, desk chairs, Bluetooth headphones and smart speakers).

**Table 11: Impact of fake reviews and informed silence on participant behaviour in online shopping task (products priced higher than £80 in the online shopping task)**

Treatment	Product chosen
Group 2 (Subtle fake reviews + no intervention)	0.0920***
Group 3 (Strong fake reviews + no intervention)	-0.0420
Group 4 (Genuine reviews + informed silence)	0.0078

Group 5 (Subtle fake reviews + informed silence)	-0.0309
Group 6 (Strong fake reviews + informed silence)	0.0136
Observations	1,230
R <sup>2</sup>	0.0044

Note: \* =  $p < 0.1$ , \*\* =  $p < 0.05$ , \*\*\* =  $p < 0.01$

The table above shows that people who were exposed to subtle fake reviews for products priced higher than £80 were 9.2% more likely to purchase that product, though no significant effect was observed for strong fake reviews.

Taken together, Tables 10 and 11 reveal that fake reviews do not have a uniform impact on products of all types: consumers are more susceptible to subtle fake reviews for electronic products (compared to household/health and beauty products) or higher-priced products. In particular, consumers were nearly three times more likely to purchase a product with subtle fake reviews if the product was more expensive than £80 compared to all products in the online shopping task more generally. Because the subtle fake reviews displayed to experiment participants were identical except the name of the product, it is likely this result is driven by consumers spending more time reading reviews when purchasing a more expensive product (to build a better understanding of product quality).

### Supplementary regression analysis: Impact on consumer trust and future behaviour

Lastly, we ran a supplementary regression analysis to assess the impact of the fake reviews and the intervention on four additional dependent variables<sup>48</sup> derived from participants' responses in the post-experiment questionnaire. These variables were designed to assess the level of trust participants had in making purchases from the online shopping platform, as well as other platforms, and whether their future behaviour would change following a bad experience with purchasing a product that did not work as well as its reviews suggested. The results of the supplementary regression analysis are presented in Table 11 below.

---

<sup>48</sup> Please see Appendix 2 for more details on how each variable was measured.

Table 12: Impact of fake reviews and informed silence on consumer trust and future behaviour.

Group	Confidence in platform	Estimated percentage of fake reviews (%)	Pay more attention to reviews (experiment site)	Pay more attention to reviews (other sites)
Group 2 (Subtle fake reviews + no intervention)	0.0036	-0.0028	-0.0117	-0.0075
Group 3 (Strong fake reviews + no intervention)	-0.0096	-0.0012	0.0026	-0.0019
Group 4 (Genuine reviews + informed silence)	0.0134	-0.0076*	0.0175	0.0052
Group 5 (Subtle fake reviews + informed silence)	-0.0247*	-0.0054	-0.0265*	-0.0150
Group 6 (Strong fake reviews + informed silence)	-0.0109	-0.0059	-0.0164	-0.0096
Observations	3,255	3,255	3,255	3,255
R <sup>2</sup>	0.0175	0.0125	0.0311	0.0269

Note: \* =  $p < 0.1$ , \*\* =  $p < 0.05$ , \*\*\* =  $p < 0.01$

The first dependent variable was the confidence that participants felt in making future purchases on the online retail platform. There was only a significant effect for participants in Group 5, who were 2.4% less likely to report that they have confidence in purchasing from the platform in the future. The second dependent variable was the estimate that participants provided in terms of the percentage of online reviews that they thought were fake. A very small

effect was found for participants in Group 4, whose estimate was 0.7% smaller than participants in the control group. No further significant effects were found in relation to this estimate.

The third dependent variable was the attention that participants reported that they would pay to reviews on the retail website in the future, if they had just purchased a product that did not work as well as the reviews had suggested. As before, only participants in Group 5 were 2.6% less likely to report that they would spend more time reading reviews on the retail platform in the future. The fourth dependent variable was identical to the third except it concerned paying attention to reviews on sites other than the retail website, however the analysis revealed that there was no significant difference between the groups on this measure.

Overall, the results in Table 11 suggest that consumers are strongly anchored in their prior beliefs about the trustworthiness of online platforms and product reviews, and exposure to fake reviews as part of an individual shopping experience is not sufficient to shift their pre-existing beliefs about product reviews more generally. The results also show that consumers are generally not likely to change their future purchasing behaviour despite being exposed to fake reviews. However, as the informed silence intervention had a limited impact on certain treatment groups, for example, participants exposed to subtle fake reviews reported a reduction in confidence in the platform, there can be some shift in participants' prior beliefs depending on the types of reviews present.<sup>49</sup>

The lack of consistency across findings in relation to the informed silence intervention highlights the importance of rigorously testing the impact of different non-regulatory intervention types prior to their implementation. For instance, it is important that an intervention that increases consumers' trust is not implemented on a platform where fake reviews are still present as this could mean consumers become more susceptible to their impact.

---

<sup>49</sup> The post-experiment survey included a question about participant health conditions and illnesses. When we interacted all treatment variables with a dummy variable if participants that had selected any one (or more) of the nine answer choices (excluding "None of the above"), we did not find that these individuals were disproportionate impacted in their shopping behaviour due to fake reviews. The full table can be found in appendix 4.



# Impact on consumer welfare and broader implications

## An indicative model of consumer harm

UK consumers spent a total of £106 billion on online retail platforms in 2022, an increase of over 40% since 2019.<sup>50</sup> Around one-third of this spending took place on third-party e-commerce platforms such as Amazon. With our research finding that fake reviews make up 11-15% of all product reviews posted on these platforms, it is possible that consumers who are exposed to fake reviews when shopping online are negatively impacted in two ways. First, consumers may make suboptimal choices (i.e. purchasing a lower quality product) if they are misled by fake reviews. Second, consumers may no longer trust reviews in general if they spot fake reviews, leading them to make less-informed decisions if they disregard helpful reviews that reflect other consumers' genuine experiences. A model of consumer harm tries to capture these two impacts as changes in consumer welfare, so a specific monetary estimate can be used to quantify the welfare loss from fake reviews. While there are a number of different approaches that can be used to quantify welfare loss, such as eliciting WTP for products of varying quality as in Akesson et al. (2022)<sup>51</sup>, our indicative model below is based on the proportion of consumers which purchased products with fake reviews (using findings from our experiment).

In this model, because there was a statistically significant difference in the proportion of consumers purchasing a product depending on whether the consumer had seen fake reviews, we could interpret this as consumers making a suboptimal choice (purchasing a product they would not have otherwise in the absence of fake reviews). Thus, our model proceeds in four steps:

1. We start with total online spending by UK consumers in 2022, adjusted for the proportion of spending which takes place on third-party platforms.
2. We combine total online spending on third-party platforms with the proportion of product reviews which are subtle fake reviews. This yields an estimate of the total online spending potentially influenced by (subtle) fake reviews. In this step, we assume that fake reviews are distributed evenly across products (this means that if 20% of all product reviews are subtle fake reviews, then our calculation assumes that 20% of each product's reviews are subtle fake reviews).
3. We multiply total online spending potentially influenced by fake reviews with our experimental estimates of the proportion of consumers that purchase different products due

---

<sup>50</sup> <https://www.statista.com/statistics/315506/online-retail-sales-in-the-united-kingdom/>

<sup>51</sup> Two key findings of this paper was that exposure to both inflated star ratings and fake product reviews led to 1) a welfare loss of \$0.12 for every \$1 spent by consumers on online shopping platforms, and 2) an increase in the probability of purchasing a low-quality product by 12.6 percentage points.

to exposure to fake reviews. This yields an estimate of total “misinformed” spending due to fake reviews.

- Finally, we assume that consumers are negatively impacted (i.e. lose utility) from “misinformed” spending due to misalignment with consumer preferences, shorter lifespan of the product and potential physical harm if the product purchased poses safety risks. This yields an estimate of annual harm to UK consumers caused by fake reviews on third-party platforms.

Note that the estimates presented below do not include services, the purchase of which are often also influenced by consumer reviews shared online.<sup>52</sup> An alternative model using the point estimates from regressions on WTP is presented in Appendix 4, though these estimates were not statistically significant and we could not reject the null hypothesis that there was no change in WTP for a product if the consumer had seen fake reviews.

**Table 13: Consumer harm from subtle fake reviews – purchasing product with fake reviews.**

Product categories impacted by subtle fake reviews	Fake review type	Assumed % of fake reviews classified as subtle <sup>53</sup>	% of fake reviews out of all reviews	% change in probability of purchase	Assumed loss of utility due to misalignment with consumer preferences	UK online retail spend via third-party platforms (£b) <sup>54</sup>	Total annual harm (£m)
All	Subtle	90%	20%	3.1%	90%	38	191
All	Subtle	90%	20%	3.1%	50%		106
All	Subtle	90%	10%	3.1%	90%		95
All	Subtle	90%	10%	3.1%	50%		50

<sup>52</sup> This includes both online services (such as cloud photo storage or music streaming) and offline services (such as restaurants or recreational activities).

<sup>53</sup> The assumption that 90% of fake reviews are well-written ‘subtle’ fake reviews is aligned with previous literature which suggest fake online reviews have become more advanced.

<sup>54</sup> <https://www.ons.gov.uk/businessindustryandtrade/retailindustry/datasets/retailsalesindexinternetsales>. Note this figure has been adjusted by a factor of 0.35, which represents an estimate of total UK e-commerce sales conducted on third-party platforms (as it is unlikely that platforms that sell their own products use fake review campaigns). The factor of 0.35 has been taken from <https://www.cityam.com/amazon-accounts-for-a-quarter-of-all-uk-online-spending/>, which states that 27% of online sales in the UK take place on Amazon UK, adjusted upward to account for other e-commerce platforms.

All	Subtle	90%	42% <sup>55</sup>	3.1%	75%		312
All (blended) <sup>56</sup>	Subtle	90%	20%	2.9%	75%		149

Our estimates listed in Table 13 suggest that the annual harm to UK consumers caused by fake reviews on third-party platforms ranges from £50 million (if we use a conservative estimate for the proportion of fake reviews and assume consumers are relatively less impacted by purchasing a suboptimal product) to £312 million (if we use an upper-bound estimate for the proportion of fake reviews and assume that consumers derive limited utility from purchasing the product with fake reviews), with the midpoint estimate being around £149 million. This is similar in magnitude to an estimate of consumer harm based on a comparable method using findings from Akesson et al<sup>57</sup>. (2022):

1. As before, we first multiply total online spending on third-party platforms with the proportion of product reviews which are fake reviews.
2. Next, we combine the output from step 1 with the behavioural impact of fake reviews (12.6 percentage point increase in the proportion of consumers buying low-quality products) resulting estimate of the total online spending directly influenced by fake reviews.
3. Finally, we combine this estimate with the welfare loss due to fake reviews (12% of total consumer spending) to get an alternate estimate of consumer harm of £115 million.

## Limitations

These estimates are very sensitive to assumptions around two parameters (the proportion of all product reviews which are fake and the extent to which consumers are negatively impacted by suboptimal purchases). For example, the annual harm to UK consumers would be reduced if consumers are still relatively happy with the products they purchased even after fake reviews caused them to switch products. In addition, it is likely that Table 13 underestimates the total harm to UK consumers caused by fake reviews for a number of reasons:

1. Our model only captures the impact of fake reviews on purchasing physical goods and does not take into account spending on services (such as hotels or restaurants). More generally, the total harm to consumers from fake reviews is likely to continue increasing in line with the growth in consumer spending on e-commerce platforms.

<sup>55</sup> <https://www.chicagotribune.com/business/ct-biz-amazon-fake-reviews-unreliable-20201020-lfbjdq25azfdpa3iz6hn6zvtwq-story.html>

<sup>56</sup> For this row, we have adjusted the % change in probability of buying a product by “splitting the difference” between the estimates for household goods and consumer electronics. More specifically, our calculation was (% change for electronics) \* (% retail spend on electronics) + (% change for household goods) \* (% retail spend on household goods) + (% change for all products) \* (% retail spend on other products).

<sup>57</sup> Akesson, Jesper., Robert W. Hahn, Robert D. Metcalfe, and Manuel Monti-Nussbaum. 2022. “The Impact of Fake Reviews on Demand and Welfare”. Unpublished manuscript, July 20 2022, typescript.

2. Our model does not consider how consumers may change their purchasing behaviour in the future due to loss of trust in product reviews. For example, consumers may incur greater search costs if they now seek out other sources of product information beyond product reviews posted on the shopping platform itself.
3. Most importantly, our model focuses on the impact of fake reviews alone, although in the real world fake reviews are often accompanied by inflated star ratings (if fake reviews are intended to raise a consumer's evaluation of product quality, it would make sense that sellers would post highly positive fake reviews). Our research findings suggest that consumers are not able to detect subtle fake reviews, which means they are also unlikely to detect which reviews have inflated star ratings. Therefore, the combined effect of fake review text and inflated star ratings on consumer purchasing behaviour is likely to be greater than the effect of fake reviews alone. As a result, the true consumer harm caused is likely higher than our estimate.

In general, not all consumers pay attention to reviews when deciding which product to purchase (our survey found that the number and content of reviews tended to rank as less important compared to price, star rating and information in the product description). Some consumers might still choose to buy a product with fake reviews if it has other attractive qualities (such as product colour or size), and other consumers might be willing to pay more for a product if they see fake reviews and no longer trust reviews for products within their initial price range.<sup>58</sup> Both of these behavioural responses would cause our model to overestimate the annual harm to UK consumers caused by fake reviews on third-party platforms. On the other hand, there is potential for significant harm to consumers if fake reviews cause consumers to purchase products that are dangerous or safety hazards. In addition, there may be significant negative impacts on mental health/well-being. For example, consumers could end up spending a significant proportion of their savings on a poor-quality product if they were influenced by fake reviews they thought were genuine. These would cause our model to underestimate the annual harm to UK consumers caused by fake reviews on third-party platforms. Because we cannot more precisely estimate the relative magnitude of these effects, the results presented are meant to be indicative (rather than definitive).

In short, by combining the total consumer spending influenced by well-written fake reviews with assumptions around the loss in utility caused by consumers purchasing goods with fake reviews, we estimate fake reviews cause £50 million to £312 million of consumer detriment per year. However, this is driven by the prevalence of fake reviews, the impacts they have on consumer behaviour and the degree to which any deception results in a loss of consumer utility as result of unrealised consumer expectations (which will be particularly large for dangerous or faulty products). Given that this estimate does not cover the impact of fake reviews in the services sector or on future consumer behaviour as well as the separate impact of inflated star ratings, we believe this is a conservative estimate and that the true consumer detriment arising from fake reviews is likely to be higher.

---

<sup>58</sup> Our experiment could not capture these behavioural responses as we did not control for all possible product characteristics and did not allow consumers to choose between products at different price ranges. As a result, these responses are not reflected in our estimated impact of fake reviews on purchase probability.

## External validity

We recognise that there are some important limitations with our study, in particular with regards to experiment design:

- Since participants were not spending their own money, they may have been less motivated to find the “best” or “highest quality” product.
- Due to the nature of Prolific as an online platform (and its popularity on forums such as Reddit), our sample skewed younger than the UK population as a whole.
- The online shopping task only included products and we are therefore not able to determine how fake reviews impact purchasing decisions for services.
- While the price and star rating for the three products viewed by consumers in the online shopping task were equivalent, some differences remained between the products (in particular the product’s image and set of features) and a number of participants based their choice on these characteristics rather than the content of reviews.

However, despite these limitations, we argue that our results can be seen as externally valid for the following reasons:

1. Our experimental design differs from previous research in that instead of using screenshots of product pages, we built a fully functioning, interactive online shopping platform that closely resembled real-world shopping experiences (with the same information presented to users). This meant that participants were more engaged in the experiment and more likely to behave as they would if they were actually buying a product online in the real world. We selected products that were closely similar in price, star rating, product characteristics and total number of reviews, which would encourage participants to read specific reviews to help them differentiate between the products. In the real world, consumers do not choose between products with identical characteristics whose only difference is their reviews. Instead, reviews are simply one factor they consider. By controlling the two most salient criteria used to evaluate products that are present on the initial product selection page (price and star rating), we encouraged participants to carefully review the information provided about each product. As a result, our experimental results capture how much weight consumers place on the content of reviews within their broader decision-making framework.
2. While many online experiments use financial incentives when investigating consumer behaviour with explicit extrinsic motivations, their use in our experiment design might encourage participants to “guess” which product had the highest quality rather than making a choice that accurately reflected their personal preferences, moving the experiment further away from a real-world shopping experience<sup>59</sup> (Eckerd et al. 2020).

---

<sup>59</sup> Eckerd, Stephanie, Scott DuHadway, Elliot Bendoly, Craig R. Carter, and Lutz Kaufmann. 2021. ‘On Making Experimental Design Choices: Discussions on the Use and Challenges of Demand Effects, Incentives, Deception, Samples, and Vignettes’. *Journal of Operations Management* 67 (2): 261–75. <https://doi.org/10.1002/joom.1128>.

3. Participants could not just click on a specific product to proceed with the experiment; instead they had to navigate to a product-specific page, then add the product to the shopping cart, then click through the cart. This encouraged participants to take their time with the experiment, as the instructions for how to proceed with the experiment were embedded within the shopping platform itself.
4. Our analyses are conducted on a large sample of UK adults that are representative by ethnicity and gender. While our sample has a greater distribution of young and middle-aged adults than the UK population as a whole, evidence suggests that a much greater proportion of young and middle-aged adults shop online compared to older adults.<sup>60</sup> The attrition rate for the experiment was less than 1% (almost entirely due to one-off technical difficulties).
5. Our survey included a wide range of questions about trust, expectations/beliefs and future behaviour, which means our results are not dependent on participants misunderstanding specific questions. In addition, we only included participants who had viewed all three products in our analyses, as these participants were the most likely to have read through different sets of product reviews and make an informed decision (instead of clicking on the first product they saw).

---

<sup>60</sup> [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=E-commerce\\_statistics\\_for\\_individuals](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=E-commerce_statistics_for_individuals)

## Conclusion

Our study consisted of two parts: (i) estimating the prevalence of fake reviews on popular UK e-commerce platforms, building on a network model similar to He et al. (2022a), and (ii) estimating the impact of fake reviews on UK consumer online shopping behaviour and perceptions, building on Akesson et al. (2022). Across the two parts of our study, our research produced six main findings:

1. Roughly 11% to 15% of product reviews on popular UK e-commerce platforms are predicted to be fake, but this proportion varies across product categories.
2. Network features are the strongest predictor of fake reviews, suggesting that automated means of review moderation should place at least much emphasis on examining the characteristics of reviewers as opposed to the content of the review itself in isolation. Review metadata such as syntax and sentiment generally only have limited predictive power in identifying fake reviews. As a result, e-commerce platforms, which have access to the data and computational power required to calculate network features, are better positioned to spot fake reviews compared to consumers (who can manually investigate only a limited number of reviews and users in isolation).
3. Consumers can generally differentiate between genuine and strong, more obviously written, fake reviews, with the presence of strong fake reviews on a specific product pushing them to purchase other products of the same type instead.
4. Consumers cannot differentiate between genuine and subtle fake reviews; subtle fake reviews tend to be conceived as “genuine” reviews and increase the likelihood that a product is purchased. Fake reviews have become increasingly more sophisticated and difficult to detect over time (moving from automated spam bots to reviews authored by skilled individuals compensated by sellers or AI-powered natural language models). As such, this finding suggests that the loss in consumer welfare caused by fake reviews will only increase over time unless platforms take greater steps to monitor fake reviews and take enforcement actions when necessary.
5. Fake reviews do not uniformly impact consumer decision-making, with purchases of certain product categories being more influenced by such reviews than others. Electronics are by far the largest e-commerce category (representing 19% of all online purchases) and our findings suggest that consumers purchasing high-value electronics goods are most susceptible to subtle fake reviews.<sup>61</sup> Future research should explore the importance of reviews in purchasing decisions across product and price categories.
6. Informing consumers that steps have been taken to moderate misleading content (such as misleading customer reviews) on the platform does not counteract the influence of fake reviews. Future research should test additional types of non-regulatory interventions (as

---

<sup>61</sup> <https://news.adobe.com/news/news-details/2022/Adobe-U.S.-Consumers-Spent-1.7-Trillion-Online-During-the-Pandemic-Rapidly-Expanding-the-Digital-Economy/default.aspx>. Note this data is from the United States but it is plausible that consumption patterns for US and UK consumers are similar.

these tend to be low-cost and straightforward for businesses to implement on platforms) to see how consumers respond, as little evidence exists on what features of interventions (such as framing of language or position on webpage) determine the salience of the intervention. This is particularly important as our experiment findings also suggested that neither fake reviews nor text-based interventions altered future purchasing intentions or consumers' perceptions or confidence in the e-commerce platform. This suggests that consumer beliefs are grounded in prior online shopping experiences and a more salient or strongly worded intervention could increase awareness of fake reviews and encourage consumers to be more cautious.



## Bibliography

- Akesson, Jesper., Robert W. Hahn, Robert D. Metcalfe, and Manuel Monti-Nussbaum. 2022. "The Impact of Fake Reviews on Demand and Welfare". Unpublished manuscript, July 20 2022, typescript.
- Ananthakrishnan, Uttara M., Beibei Li, and Michael D. Smith. 2020. 'A Tangled Web: Should Online Review Portals Display Fraudulent Reviews?' SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.3297363>.
- Chang, Hsin Hsin, Po Wen Fang, and Chien Hao Huang. 2015. 'The Impact of On-Line Consumer Reviews on Value Perception: The Dual-Process Theory and Uncertainty Reduction'. *Journal of Organizational and End User Computing* 27 (2): 32–57. <https://doi.org/10.4018/joeuc.2015040102>.
- Costa, Ana, João Guerreiro, Sérgio Moro, and Roberto Henriques. 2019. 'Unfolding the Characteristics of Incentivized Online Reviews'. *Journal of Retailing and Consumer Services* 47 (March): 272–81. <https://doi.org/10.1016/j.jretconser.2018.12.006>.
- Eckerd, Stephanie, Scott DuHadway, Elliot Bendoly, Craig R. Carter, and Lutz Kaufmann. 2021. 'On Making Experimental Design Choices: Discussions on the Use and Challenges of Demand Effects, Incentives, Deception, Samples, and Vignettes'. *Journal of Operations Management* 67 (2): 261–75. <https://doi.org/10.1002/joom.1128>.
- Floyd, Kristopher, Ryan Freling, Saad Alhoqail, Hyun Young Cho, and Traci Freling. 2014. 'How Online Product Reviews Affect Retail Sales: A Meta-Analysis'. *Journal of Retailing, Empirical Generalizations in Retailing*, 90 (2): 217–32. <https://doi.org/10.1016/j.jretai.2014.04.004>.
- Gall-Ely, Marine Le. 2009. 'Definition, Measurement and Determinants of the Consumer's Willingness to Pay: A Critical Synthesis and Directions for Further Research'. Post-Print, Post-Print, June. <https://ideas.repec.org/p/hal/journal/hal-00522828.html>.
- He, Sherry, Brett Hollenbeck, Gijs Overgoor, Davide Proserpio, and Ali Tosyali. 2022. 'Detecting Fake Review Buyers Using Network Structure: Direct Evidence from Amazon'. SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.4147920>.
- He, Sherry, Brett Hollenbeck, and Davide Proserpio. 2022. 'The Market for Fake Reviews'. SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.3664992>.
- Hu, Nan, Ling Liu, and Vallabh Sambamurthy. 2011. 'Fraud Detection in Online Consumer Reviews'. *Decision Support Systems, On quantitative methods for detection of financial fraud*, 50 (3): 614–26. <https://doi.org/10.1016/j.dss.2010.08.012>.

- Langhe, Bart de, Philip M. Fernbach, and Donald R. Lichtenstein. 2016. 'Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings'. *Journal of Consumer Research* 42 (6): 817–33. <https://doi.org/10.1093/jcr/ucv047>.
- Manes, Eran, and Anat Tchetchik. 2018. 'The Role of Electronic Word of Mouth in Reducing Information Asymmetry: An Empirical Investigation of Online Hotel Booking'. *Journal of Business Research* 85 (April): 185–96. <https://doi.org/10.1016/j.jbusres.2017.12.019>.
- Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. 2014. 'Promotional Reviews: An Empirical Investigation of Online Review Manipulation'. *American Economic Review* 104 (8): 2421–55. <https://doi.org/10.1257/aer.104.8.2421>.
- Moon, Sangkil, Moon-Yong Kim, and Paul K. Bergey. 2019. 'Estimating Deception in Consumer Reviews Based on Extreme Terms: Comparison Analysis of Open vs. Closed Hotel Reservation Platforms'. *Journal of Business Research* 102 (September): 83–96. <https://doi.org/10.1016/j.jbusres.2019.05.016>.
- Ni, Jianmo, Jiacheng Li, and Julian McAuley. 2019. 'Justifying Recommendations Using Distantly-Labeled Reviews and Fine-Grained Aspects'. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 188–97. Hong Kong, China: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1018>.
- Oak, Rajvardhan, and Zubair Shafiq. 2022. 'The Fault in the Stars: Understanding Underground Incentivized Review Services'. arXiv. <https://doi.org/10.48550/arXiv.2102.04217>.
- Park, Cheol, and Thae Min Lee. 2009. 'Information Direction, Website Reputation and EWOM Effect: A Moderating Role of Product Type'. *Journal of Business Research* 62 (1): 61–67. <https://doi.org/10.1016/j.jbusres.2007.11.017>.
- Plotkina, Daria, Andreas Munzel, and Jessie Pallud. 2020. 'Illusions of Truth—Experimental Insights into Human and Algorithmic Detections of Fake Online Reviews'. *Journal of Business Research* 109 (March): 511–23. <https://doi.org/10.1016/j.jbusres.2018.12.009>.
- Sandulescu, Vlad, and Martin Ester. 2015. 'Detecting Singleton Review Spammers Using Semantic Similarity'. In *Proceedings of the 24th International Conference on World Wide Web*, 971–76. <https://doi.org/10.1145/2740908.2742570>.
- Shaw, Norman, Brenda Eschenbrenner, and Daniel Baier. 2022. 'Online Shopping Continuance after COVID-19: A Comparison of Canada, Germany and the United States'. *Journal of Retailing and Consumer Services* 69 (November): 103100. <https://doi.org/10.1016/j.jretconser.2022.103100>.
- Siering, Michael, Jan Muntermann, and Balaji Rajagopalan. 2018. 'Explaining and Predicting Online Review Helpfulness: The Role of Content and Reviewer-Related Signals'. *Decision Support Systems* 108 (April): 1–12. <https://doi.org/10.1016/j.dss.2018.01.004>.

Zhuang, Mengzhou, Geng Cui, and Ling Peng. 2018. 'Manufactured Opinions: The Effect of Manipulating Online Product Reviews'. *Journal of Business Research* 87 (June): 24–35. <https://doi.org/10.1016/j.jbusres.2018.02.016>.

# Appendix 1: Approach to predictive modelling

## Approach to model selection

We tested three different supervised learning classifiers, including:

- Random forest (RF): a set of decision trees, an algorithm that, as its metaphorical namesake implies, splits datasets from a “root” according to the gain in discriminatory power gained by the split of a variable)
- Support vector classifier (SVC) (that fits a plane that optimally separates the classes in the feature space, i.e., real and fake reviews)
- Logistic regression (LogReg): (a baseline model that is often used as benchmark and works by mapping the predicted response (fake or real review) through a sigmoidal curve on the space of features).

## Network variables creation

The network indicators were computed from the dataset matching reviewers and products for each sector, first by listing, for each product, the set of reviewers, then by computing, for each pairwise combination of products, the set of common reviewers. If one or more common reviewers were shared across two products, a corresponding edge was added to the network. The cost in terms of time and computing power required for checking, by brute force, all possible combinations is generally prohibitive (for example, if our dataset has 100.000 unique products, this amounts to close to 5 billion combinations that need to be checked).<sup>62</sup> As a result, at this stage the network density was determined on 10% of the data of the Amazon dataset, and then the uncovered distribution was extended to the remaining 90% of the data. We then calculated the degree of each node, as well as the PageRank, (Betweenness) Centrality, Eigenvector centrality, and Clustering Coefficient.<sup>63</sup> The alpha parameter of the PageRank algorithm was set at 0.9.

## Creation of other variables

The similarity between reviews for each good was computed first by taking the term frequency-inverse document frequency (TF/IDF) representation of the set of reviews for the 1000 most frequent 1- or 2-grams (i.e. one- or two- word combinations), then sub-setting the set of vectors for each good and computing its mean cosine-similarity from the unique cosine distance matrix obtained per product. Products with many similar reviews in terms of content should have a

---

<sup>62</sup> The number of pairwise combinations  $C_r^n$  is given by the formula,  $C_r^n = \frac{n!}{(n-r)!r!}$ , with  $n$  the total number of products in the dataset and  $r = 2$  (the number of products in each combination).

<sup>63</sup> Definitions for these network features are provided in Table 1 in the main body of the report.

cosine similarity closer to 1 than reviews done by independent reviewers, which should have cosine similarities fluctuating around 0. The working hypothesis is that similar reviews may be an indicator of an underlying campaign or a relatively small set of reviewers using similar text.

The other metadata variables were created by simply grouping dates and reviews of the same product and computing the statistics of interest for each group.

The Parts of Speech Tagging was determined by, for each review, dividing the number of interjections, verbs, pronouns, proper nouns, punctuation marks, adjectives, and adverbs by the total length of the review.

The synthetic variables were created using a technique called Genetic Programming (specifically the Symbolic Transformer method). This technique involves the computer trying different combinations of variables to find the best solution for a problem, in this case identifying fake reviews. It helps to find new variables that are related to each other and can explain the data in a better way. For example, the computer might try to square a person's age to see if it helps explain their income. This iterative testing occurs in an unsupervised manner i.e. automatically without being told what to look for. The transformer was run for 50 generations, starting with 2000 options, keeping the five highest-performing features, and with otherwise the standard parameter set as in the SymbolicTransformer class of the GPLearn package.

The Sentiment was computed per review according to the results of the "compound" measure of the VADER sentiment analyser, expressing an intensity of a feeling according to the semantic features of a text, between -1 (very negative) and 1 (very positive feeling).

After the variables were created, they were transformed so that they could be more easily compared to each other. This was done using a method called z-score normalization, which subtracts the population mean from each value and then divides by the standard deviation. The result is a standardized score, or z-score, that represents how many standard deviations a particular value is from the population mean.<sup>64</sup>

## Feature importance

Our chosen metric of feature importance is Gini importance, which measures how much a feature contributes to the overall ability of the model to discriminate between different classes or categories. It is based on the idea that discriminating between real and fake reviews can be improved by identifying a threshold value of the Clustering Coefficient, which is a measure of how closely connected a product is to other products in a network. By selecting a threshold value and creating a split on the dataset based on this value, the Gini importance metric can be used to measure the gain in discriminating power that results from considering this feature. The split between values above and below the threshold that results in the most discrimination between real and fake reviews is considered to be the most informative. By repeating this

---

<sup>64</sup> The z-score can be calculated by  $z = \frac{x-\mu}{\sigma}$ , where  $x$  is the realization of the random variable/column of the dataset,  $\mu$  is the relevant population mean, and  $\sigma$  its standard deviation.

process iteratively, the most important features for discriminating between real and fake reviews can be identified.

### Predictive modelling

The models used for comparison were all created with the same set of rules. Specifically, the Random Forest model was created with a set of parameters that includes: using 100 trees, splitting the samples into at least 2 groups, allowing each leaf to have at least 1 sample, not weighting any particular sample more heavily than others, only considering a square root of all possible features, not restricting the maximum number of leaf nodes, not requiring a minimum decrease in impurity for a split, and using bootstrapping (randomly resampling the data with replacement data) during tree building.

The chosen algorithm (Random Forest) is by its nature robust to the problem of multicollinearity (correlation between explanatory variables) as it consists of an ensemble of decision trees and splits the data one parameter at a time.<sup>65</sup> As a further robustness check, the removal of the most correlated variables in the network partition does not affect the model outcome.

In each case, the model was trained on the training fold with labels, and then used to predict the test labels. The data was split into 5 equal parts, and the model was trained and tested on each part in turn, using the other parts for training. This process was repeated 10 times to ensure the results were reliable. The test sets were carefully chosen to have a similar proportion of real and fake reviews as the training sets, to avoid any bias in the results. The average results across all test sets were reported in tables 6 to 8.

The importance measure for the Random Forest model is the Gini (or Mean Decrease in Impurity Index) index, measuring the decrease in "contamination" that is achieved if a split among the trees composing the Forest is done by each of the variables.

The plotting of the decision regions of each model was done first by taking a sample of 150 fake and authentic reviews, then by compressing the feature space of the 12 original features into two through Principal Component Analysis (PCA), then by estimating each of the models in this new compressed space. PCA is a technique used to simplify complex models by reducing the number of variables while retaining as much of the original variation as possible. It does this by creating new variables, called principal components, which are a combination of the original variables. The new variables are chosen to explain as much of the variation in the data as possible and are orthogonal (independent) to each other.

---

<sup>65</sup> Multicollinearity is undesirable as it can cause problems in the estimation of regression coefficients, leading to unstable and unreliable results

## Alternate classifiers and robustness checks

**Table A1.1: Out-of-sample prediction performance of the RF classifier (5-fold cross-validation)**

Features	AUC	Accuracy	Recall	Precision	F1 score
Network	0.99999	0.99960	0.99947	0.99997	0.99960
Metadata	0.99999	0.99998	0.99998	0.99998	0.99998
All features	0.99999	0.99998	0.99998	0.99998	0.99998

**Table A1.2: Out-of-sample prediction performance of the SVC classifier (5-fold cross-validation)**

Features	AUC	Accuracy	Recall	Precision	F1 score
Network	0.88168	0.80996	0.76788	0.79756	0.77870
Metadata	0.79511	0.75442	0.69666	0.73287	0.69994
All features	0.88621	0.82845	0.79420	0.81605	0.80304

**Table A1.3: Out-of-sample prediction performance of the LogReg classifier (5-fold cross-validation)**

Features	AUC	Accuracy	Recall	Precision	F1 score
Network	0.87983	0.81146	0.77023	0.79893	0.78082
Metadata	0.79620	0.75738	0.70236	0.73552	0.70387
All features	0.88543	0.83437	0.80254	0.82194	0.81088

## Appendix 2: Post-experiment questions

1. You are shopping online for a new [reusable water bottle]<sup>66</sup>. When looking online, you notice the following information about prices:

- a. The average reusable water bottle price is around £23.
- b. However, some prices for a reusable water bottle are as low as £9.
- c. Other prices for a reusable water bottle are as high as £44.

On average, how much would you be willing to pay for a reusable water bottle? *Please select one of the intervals below.*

- a. £10.00 - £15.99
- b. £16.00 - £21.99
- c. £22.00 - £27.99
- d. £28.00 - £33.99
- e. £34.00 - £39.99

2. *Please tell us more specifically what you would be willing to pay for this item by selecting one of the intervals below.*

- a. £10.00 - £10.99
- b. £11.00 - £11.99
- c. £12.00 - £12.99
- d. £13.00 - £13.99
- e. £14.00 - £14.99
- f. £15.00 - £15.99

3. COMPREHENSION CHECK: What product type did you see on the online retail platform?

- a. [Reusable water bottle]
- b. Children's socks
- c. Washing machine

---

<sup>66</sup> The name of the product type and the specific prices in the question/answer choices will vary based on the product type shown to the participant on the online shopping platform.



- d. Cutlery set
  - e. Multi-vitamin tablets
4. FREE-TEXT: Think back to the product you selected to purchase just now. In as much detail as possible, why did you select this specific product?
  5. In the real world, how likely would you be to purchase the product type you selected?
    - a. Very likely
    - b. Somewhat likely
    - c. Neither likely nor unlikely
    - d. Somewhat unlikely
    - e. Very unlikely
  6. ATTENTION CHECK: To show that you are paying attention, please select 'very unlikely' from the options below.
    - a. Very likely
    - b. Somewhat likely
    - c. Neither likely nor unlikely
    - d. Somewhat unlikely
    - e. Very unlikely

We will now present four different reviews for the product you selected to purchase. Please rank each review on the following characteristics. [5-point slider]

7. I found this review [very unhelpful/neither helpful nor unhelpful/very helpful].
8. I think that other consumers who wish to purchase this product would find this review [very unhelpful/neither helpful nor unhelpful/very helpful].
9. I found this review [not credible at all/somewhat credible/extremely credible].
10. I found this review [not relevant/somewhat relevant/extremely relevant].
11. How often do you purchase items online?
  - a. More than once a week
  - b. About once per week
  - c. Several times a month
  - d. About once a month

e. Once in a few months or longer

12. How often do you purchase items on Amazon?

a. More than once a week

b. About once per week

c. Several times a month

d. About once a month

e. Once in a few months or longer

f. Never

13. When shopping online, what factors do you consider when deciding which product to purchase? Rank the following factors from the most to the least important to you.

a. Star rating

b. Number of reviews

c. Content of reviews

d. Information in the product description

e. Brand

f. Look of the product

g. Price

h. Seller information

14. In your opinion, what proportion of product reviews online are not genuine? [0-100% slider]

15. Imagine that the product you had just purchased did not work as well as the reviews suggested. If you saw a product that you wanted to purchase from the same supplier with a 5-star rating, how likely would you be to purchase the product?

a. Very likely

b. Somewhat likely

c. Neither likely nor unlikely

d. Somewhat unlikely

e. Very unlikely

16. To what extent do you agree with the following statement: I would feel confident making future purchases from this online retail platform. [strongly disagree/disagree/neither agree nor disagree/ agree/strongly agree].
17. ATTENTION CHECK: This is an attention check. Please select 'neither helpful nor unhelpful' from the options below.
- Very helpful
  - Helpful
  - Neither helpful nor unhelpful
  - Unhelpful
  - Very unhelpful
18. Imagine that the product you had just purchased did not work as well as the reviews suggested. How would this impact the time spent reading reviews **on the retail website** about the next product you purchase online?
- I would spend much more time reading reviews.
  - I would spend somewhat more time reading reviews.
  - I would not change the amount of time spent reading reviews.
  - I would spend somewhat less time reading reviews.
  - I would spend much less time reading reviews.
19. Imagine that the product you had just purchased did not work as well as the reviews suggested. How would this impact the time spent reading reviews **outside the retail website** about the next product you purchase online?
- I would spend much more time reading reviews
  - I would spend somewhat more time reading reviews
  - I would not change the amount of time spent reading reviews
  - I would spend somewhat less time reading reviews
  - I would spend much less time reading reviews
20. FREE-TEXT: Imagine that the product you had just purchased did not work as well as the reviews suggested. Beyond time spent reading reviews, how else would your behaviour change when deciding which product to purchase?

You are almost there! Just a few more questions about yourself. [Note: all multiple-choice questions will include "Prefer not to answer" as an option.]

21. Please enter the first half of your postcode (Type 0 if you do not want to answer this question).
22. Do you have any health conditions or illnesses which affect you in any of the following areas? Please select all options that apply to you.
- a. Learning or understanding or concentrating
  - b. Memory
  - c. Mental health
  - d. Socially or behaviourally (for example associated with autism spectrum disorder (ASD) which includes Asperger's, or attention deficit hyperactivity disorder (ADHD))
  - e. Vision (for example blindness or partial sight)
  - f. Hearing (for example deafness or partial hearing)
  - g. Mobility (for example walking short distances or climbing stairs)
  - h. Dexterity (for example lifting and carrying objects, using a keyboard)
  - i. Stamina or breathing or fatigue
  - j. Prefer not to say
  - k. None of the above
23. Which of the following best describes your personal income, before taxes, last year?
- a. Up to £9,999
  - b. £10,000 - £24,999
  - c. £25,000 - £49,999
  - d. £50,000 - £74,999
  - e. £75,000 - £99,999
  - f. £100,000 or more
24. What is the highest level of education you have completed?
- a. Less than primary school / primary school not completed
  - b. Primary
  - c. Secondary
  - d. Vocational

e. Undergraduate

f. Postgraduate

25. What is your age?

a. 18-24 years old

b. 25-34 years old

c. 35-44 years old

d. 45-54 years old

e. 55-64 years old

f. 65 years or older

26. What is your sex?

a. Female

b. Male

c. Intersex

27. Is the gender you identify with the same as your sex registered at birth?

a. Yes

b. No

28. If No, what is your gender identity?

a. Woman (incl. trans woman)

b. Man (incl. trans man)

c. Non-binary, gender fluid or gender queer

29. What is your ethnicity?

a. White (includes English/Welsh/Scottish/Northern Irish/British/Gypsy or Traveller/Any other White background)

b. Mixed/Multiple ethnic groups (includes White and Black Caribbean/White and Black African/White and Asian/Other Mixed)

c. Asian/Asian British (includes Asian British/Indian/Pakistani/Bangladeshi/Chinese/Other Asian)

d. Black/African/Caribbean/Black British (includes Black British/African/Caribbean/Other Black)

e. Other ethnic groups (includes Arab/Any other ethnic group)

30.FREE-TEXT: Did you encounter any technical difficulties while completing this experiment? (optional)

31.FREE-TEXT: Do you have any additional comments after completing this experiment? (optional)

# Appendix 3: Products and product reviews displayed to experiment participants on the online shopping platform

Figure A3.1: Screenshot of the product overview page

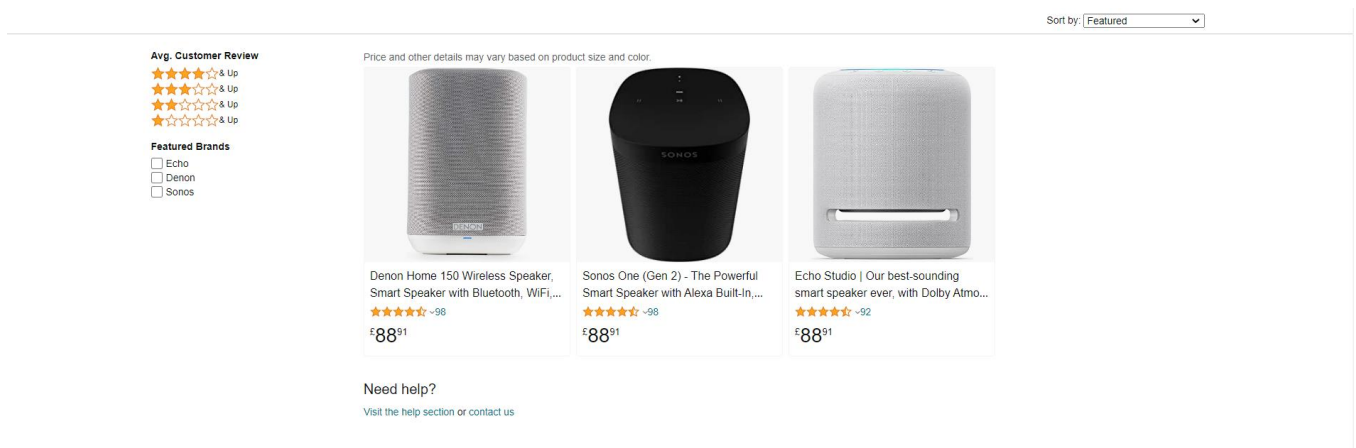
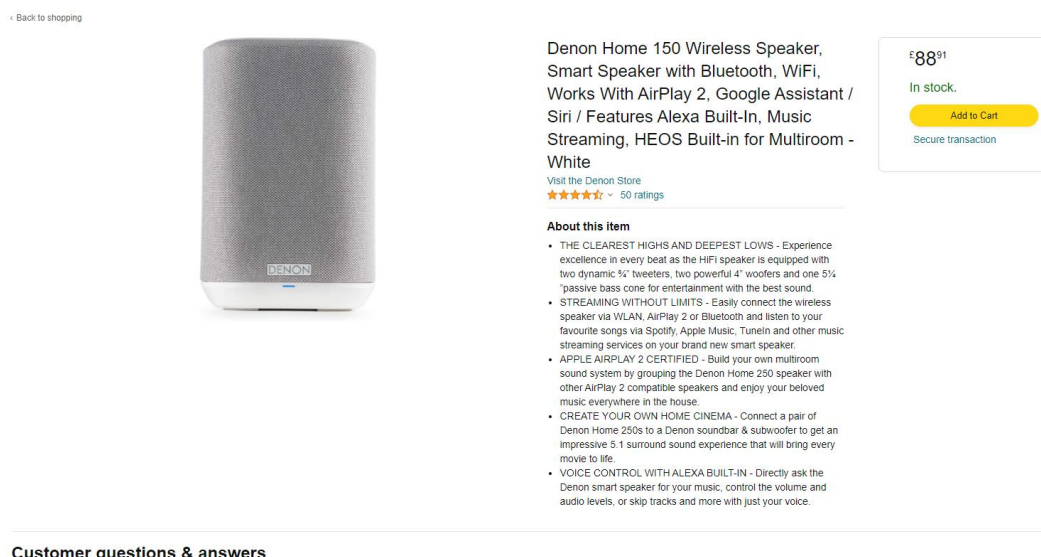
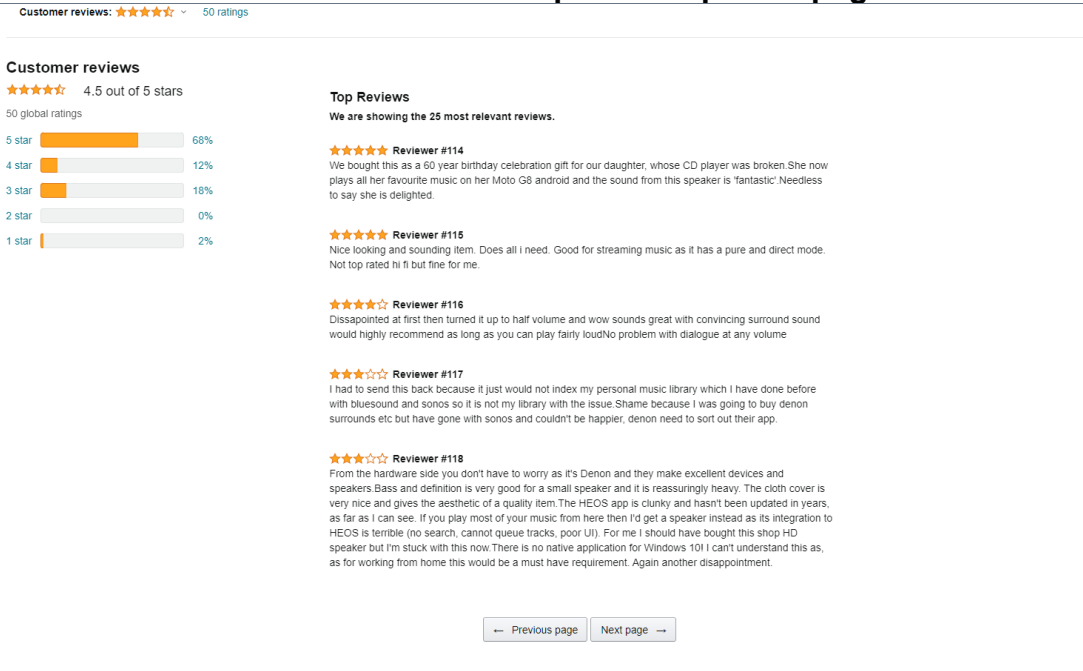


Figure A3.2: Screenshot of the product-specific page



## Figure A3.3: Screenshot of the reviews on the product-specific page





## Appendix 4: Supplementary analyses

**Table A4.1: Participant WTP for each product type**

Product type	Experimented listed price (£)	Mean WTP interval	Standard deviation	Number of participants
Kettle	52.13	42.1	12.0	445
Iron	44.5	39.4	10.1	441
Vacuum	94.12	109.0	34.6	431
Desk chair	90.32	88.1	17.5	444
Bluetooth headphones	86.58	77.7	36.3	431
Keyboard	38.46	32.9	10.2	434
Powerbank	31.98	29.4	5.5	430
Smart speaker	88.91	74.6	26.9	433
Sunscreen	21.18	14.1	4.97	445
Yoga mat	21.32	20.6	3.05	447
Re-usable water bottle	25.46	17.7	5.63	437

**Table A4.2: Impact of fake reviews and informed silence on participant WTP in online shopping task**

Treatment	Willingness to pay (WTP)
Group 2 (Subtle fake reviews + no intervention)	0.0049
Group 3 (Strong fake reviews + no intervention)	-0.0064
Group 4 (Genuine reviews + informed silence)	-0.0242***
Group 5 (Subtle fake reviews + informed silence)	0.0076
Group 6 (Strong fake reviews + informed silence)	0.0150
Observations	3,255
R <sup>2</sup>	0.8130

Note: \* =  $p < 0.1$ , \*\* =  $p < 0.05$ , \*\*\* =  $p < 0.01$

**Table A4.3: Impact of fake reviews and informed silence on participant WTP in online shopping task (by product category)**

<b>Product category</b>	<b>Electronics</b>	<b>Household</b>	<b>Health and beauty</b>
Group 2 (Subtle fake reviews + no intervention)	-0.0006	0.0016	0.0170
Group 3 (Strong fake reviews + no intervention)	-0.0261	0.0157	-0.0115
Group 4 (Genuine reviews + informed silence)	-0.0057	-0.0346***	-0.0230
Group 5 (Subtle fake reviews + informed silence)	-0.0260	-0.0005	-0.0038
Group 6 (Strong fake reviews + informed silence)	0.0274	0.0079	0.0191
Observations	1192	1210	851
R <sup>2</sup>	0.5849	0.7467	0.2906

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.4: Impact of fake reviews and informed silence on participant behaviour in online shopping task (products priced less than £40 in the online shopping task)**

Treatment	Product chosen	WTP
Subtle fake reviews	0.0145	0.0094
Strong fake reviews	-0.0858***	-0.0188
Intervention informed silence	-0.0061	-0.0185
Observations	1,421	1,421
R <sup>2</sup>	0.0026	0.5390

Note: \* =  $p < 0.1$ , \*\* =  $p < 0.05$ , \*\*\* =  $p < 0.01$

**Table A4.5: Impact of fake reviews and informed silence on participant behaviour in online shopping task (participants who ranked content of reviews as top 3 most important factors when deciding which product to buy)**

Treatment	Product chosen	WTP
Group 2 (subtle fake reviews + no intervention)	0.0751**	0.0010
Group 3 (strong fake reviews + no intervention)	-0.0137	0.0103
Group 4 (genuine reviews + informed silence)	0.0077	-0.0537***
Group 5 (subtle fake reviews + informed silence)	-0.0250	-0.0201
Group 6 (strong fake reviews + informed silence)	0.0045	0.0253
Observations	3,255	3,255
R <sup>2</sup>	0.0024	0.8427

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.6: Impact on consumer trust and future behaviour (participants who thought subtle fake reviews were as credible as genuine reviews)**

Treatment	Confidence in platform	Estimated % of fake reviews	Pay more attention to reviews (experiment site)	Pay more attention to reviews (other sites)
Group 2 (subtle fake reviews + no intervention)	0.0663	-0.0199	-0.0267	-0.0328
Group 3 (strong fake reviews + no intervention)	-0.0395	-0.0151	0.0241	0.0471
Group 4 (genuine reviews + informed silence)	0.1311***	-0.0419***	0.0018	-0.0062
Group 5 (subtle fake reviews + informed silence)	-0.1569*	0.0336	-0.0454	-0.0053
Group 6 (strong fake reviews + informed silence)	0.0431	0.0401	-0.0844	-0.0581
Observations	1287	1287	1287	1287
R <sup>2</sup>	0.0175	0.0125	0.0311	0.0269

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.7: Impact of fake reviews and informed silence on trust and future behaviour (controlling for online shopping frequency)**

Treatment	Confidence in platform	Estimated % of fake reviews	Pay more attention to reviews (experiment site)	Pay more attention to reviews (other sites)	Likely to trust 5-star rating if previous experience was bad
Group 2 (subtle fake reviews + no intervention)	-0.0008	-0.0036	-0.0072	-0.0096	-0.0040
Group 3 (strong fake reviews + no intervention)	-0.0035	-0.0053	0.0030	0.0002	-0.0045
Group 4 (genuine reviews + informed silence)	0.0139	-0.0035	0.0085	-0.0075	0.0154
Group 5 (subtle fake reviews + informed silence)	-0.0207	-0.0097	-0.0121	-0.0175	0.0203
Group 6 (strong fake reviews + informed silence)	-0.0240	0.0070	-0.0010	-0.0025	-0.0127
Shops online at least once weekly	0.0592***	0.0068*	0.0487***	0.0385***	0.0196**
Observations	3,255	3,255	3,255	3,255	3,255
R <sup>2</sup>	0.0255	0.0174	0.0337	0.0310	0.0133

 Note: \* =  $p < 0.1$ , \*\* =  $p < 0.05$ , \*\*\* =  $p < 0.01$

**Table A4.8: Impact of fake reviews and informed silence on participant behaviour in online shopping task (treatment variables interacted with income)**

Treatment	Product chosen	WTP
Group 2 (subtle fake reviews + no intervention)	0.0398**	-0.0063
Group 3 (strong fake reviews + no intervention)	- 0.0642***	0.0029
Group 4 (genuine reviews + informed silence)	-0.0006	-0.0237**
Group 5 (subtle fake reviews + informed silence)	-0.0133	0.0060
Group 6 (strong fake reviews + informed silence)	0.0214	-0.0386
Group 2 (subtle fake reviews + no intervention) x Income < £25k	-0.0139	-0.0509***
Group 3 (strong fake reviews + no intervention) x Income < £25k	0.0208	-0.0691***
Group 4 (genuine reviews + informed silence) x Income < £25k	-0.0000	-0.0264
Group 5 (subtle fake reviews + informed silence) x Income < £25k	0.0139	0.0317
Group 6 (strong fake reviews + informed silence) x Income < £25k	-0.0208	0.1105**
Group 2 (subtle fake reviews + no intervention) x Income £25-£50k	-0.0139	0.0381***
Group 3 (strong fake reviews + no intervention) x Income £25-£50k	0.0208	-0.0171
Group 4 (genuine reviews + informed silence) x Income £25-£50k	-0.0000	0.0304
Group 5 (subtle fake reviews + informed silence) x Income £25-£50k	0.0138	-0.0370
Group 6 (strong fake reviews + informed silence) x Income £25-£50k	-0.0208	0.0764



Group 2 (subtle fake reviews + no intervention) x Income > £50k	-0.0139	0.1956***
Group 3 (strong fake reviews + no intervention) x Income > £50k	0.0208	0.1651***
Group 4 (genuine reviews + informed silence) x Income > £50k	-0.0000	0.0119
Group 5 (subtle fake reviews + informed silence) x Income > £50k	0.0139	-0.1047*
Group 6 (strong fake reviews + informed silence) x Income > £50k	-0.0208	-0.0631
Observations	3,255	3,255
R <sup>2</sup>	0.0018	0.8126

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.9: Impact of fake reviews and informed silence on participant behaviour in online shopping task (treatment variables interacted with online shopping frequency)**

Treatment	Product chosen	WTP
Group 2 (subtle fake reviews + no intervention)	0.0351**	-0.0053
Group 3 (strong fake reviews + no intervention)	- 0.0579***	-0.0107
Group 4 (genuine reviews + informed silence)	-0.0010	-0.0408***
Group 5 (subtle fake reviews + informed silence)	0.0193	0.0340**
Group 6 (strong fake reviews + informed silence)	-0.0117	0.0182
Group 2 (subtle fake reviews + no intervention) x Shops online at least once weekly	-0.0127	0.0593***
Group 3 (strong fake reviews + no intervention) x Shops online at least once weekly	0.0183	0.0401***

Group 4 (genuine reviews + informed silence) x Shops online at least once weekly	0.0000	0.0680***
Group 5 (subtle fake reviews + informed silence) x Shops online at least once weekly	0.0127	-0.0764***
Group 6 (strong fake reviews + informed silence) x Shops online at least once weekly	-0.0184	-0.0635**
Observations	3,255	3,255
R <sup>2</sup>	0.0018	0.8126

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.10: Impact of fake reviews and informed silence on participant behaviour in online shopping task (treatment variables interacted with impairment)**

Treatment	Product chosen	Confidence in platform	Estimated % of fake reviews	Pay more attention to reviews (experiment site)	Pay more attention to reviews (other sites)	Likely to trust 5-star rating if previous experience was bad
Group 2 (subtle fake reviews + no intervention)	0.0329**	0.0014	-0.0032	-0.0152	-0.0143	-0.0104
Group 3 (strong fake reviews + no intervention)	-0.0555***	-0.0013	-0.0055	0.0108	0.0065	0.0044
Group 4 (genuine reviews + informed silence)	-0.0006	0.0015	-0.0180***	0.0182	-0.0073	0.0099

## Estimating the Impact and Prevalence of Fake Online Reviews

Group 5 (subtle fake reviews + informed silence)	-0.0109	-0.0155	0.0074	-0.0151	-0.0144	0.0260
Group 6 (strong fake reviews + informed silence)	0.0185	-0.0106	0.0197**	-0.0009	-0.0032	-0.0208
Group 2 (subtle fake reviews + no intervention) x Impairment	-0.0115	-0.0259	-0.0039	0.0806***	0.0410	0.0469**
Group 3 (strong fake reviews + no intervention) x Impairment	0.0179	-0.0326	-0.0010	-0.0397	-0.0407	-0.0584**
Group 4 (genuine reviews + informed silence) x Impairment	-0.0000	0.0251	0.0492***	-0.0254	-0.0022	0.0166
Group 5 (subtle fake reviews + informed silence) x Impairment	0.0115	0.0033	-0.0593***	-0.0644	-0.0501	-0.0594
Group 6 (strong fake reviews + informed	-0.0179	-0.0326	-0.0427**	0.0119	0.0285	0.0851*

silence) x Impairment						
Observations	3,255	3,255	3,255	3,255	3,255	3,255
R <sup>2</sup>	0.0015	0.0225	0.0198	0.0331	0.0302	0.0141

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.11: Impact of fake reviews and informed silence on participant behaviour in online shopping task (treatment variables interacted with highest level of education achieved)**

Treatment	Product chosen	Confidence in platform	Estimated % of fake reviews	Pay more attention to reviews (experiment site)	Pay more attention to reviews (other sites)	Likely to trust 5-star rating if previous experience was bad
Group 2 (subtle fake reviews + no intervention)	0.0368**	0.0013	-0.0022	-0.0011	0.0008	0.0082
Group 3 (strong fake reviews + no intervention)	-0.0605** *	-0.0064	-0.0025	0.0135	0.0054	-0.0005
Group 4 (genuine reviews + informed silence)	-0.0006	0.0216	-0.0103	0.0426*	0.0280	0.0308
Group 5 (subtle fake reviews + informed silence)	-0.0123	0.0028	-0.0064	-0.0520*	-0.0855** *	0.0085

## Estimating the Impact and Prevalence of Fake Online Reviews

Group 6 (strong fake reviews + informed silence)	0.0201	-0.0434	0.0194*	-0.0686**	-0.0373	-0.0114
Group 2 (subtle fake reviews + no intervention) x Undergraduate or higher	-0.0128	-0.0001	-0.0101*	-0.0391**	-0.0501** *	-0.0576***
Group 3 (strong fake reviews + no intervention) x Undergraduate or higher	0.0196	0.0081	-0.0142**	-0.0545***	-0.0413**	-0.0403**
Group 4 (genuine reviews + informed silence) x Undergraduate or higher	0.0000	-0.0121	0.0035	-0.0852***	-0.0876** *	-0.0568**
Group 5 (subtle fake reviews + informed silence) x Undergraduate or higher	0.0128	-0.0425	0.0037	0.0983**	0.1545** *	0.0673*
Group 6 (strong fake reviews + informed silence) x	-0.0195	0.0223	-0.0082	0.1555***	0.0912**	0.0322

Undergraduate or higher						
Observations	3,255	3,255	3,255	3,255	3,255	3,255
R <sup>2</sup>	0.0017	0.0226	0.0180	0.0337	0.0321	0.0160

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.12: Impact of fake reviews and informed silence on participant behaviour in online shopping task (treatment variables interacted with sex)**

Treatment	Product chosen	Confidence in platform	Estimated % of fake reviews	Pay more attention to reviews (experiment site)	Pay more attention to reviews (other sites)	Likely to trust 5-star rating if previous experience was bad
Group 2 (subtle fake reviews + no intervention)	0.0352**	-0.0032	-0.0015	-0.0029	-0.0065	-0.0070
Group 3 (strong fake reviews + no intervention)	-0.0579***	-0.0055	-0.0017	0.0018	0.0064	0.0004
Group 4 (genuine reviews + informed silence)	-0.0012	-0.0177	-0.0046	0.0036	-0.0208	-0.0004
Group 5 (subtle fake reviews + informed silence)	-0.0117	-0.0161	-0.0063	-0.0205	-0.0156	0.0276

## Estimating the Impact and Prevalence of Fake Online Reviews

Group 6 (strong fake reviews + informed silence)	0.0194	0.0002	0.0035	-0.0192	-0.0408	-0.0039
Group 2 (subtle fake reviews + no intervention) x Male	-0.0128	0.0253	-0.0186***	-0.0181	-0.0275	0.0094
Group 3 (strong fake reviews + no intervention) x Male	0.0182	0.0137	-0.0268***	-0.0078	-0.0478**	-0.0211
Group 4 (genuine reviews + informed silence) x Male	0.0001	0.0732***	-0.0107	0.0002	0.0053	0.0293
Group 5 (subtle fake reviews + informed silence) x Male	0.0128	-0.0326	0.0098	0.0320	0.0210	-0.0221
Group 6 (strong fake reviews + informed silence) x Male	-0.0183	-0.0651	0.0314**	0.0436	0.1215***	-0.0013
Observations	3,255	3,255	3,255	3,255	3,255	3,255

R <sup>2</sup>	0.0016	0.0224	0.0186	0.0309	0.0301	0.0131
----------------	--------	--------	--------	--------	--------	--------

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.13: Impact of fake reviews and informed silence on participant behaviour in online shopping task (treatment variables interacted with age)**

Treatment	Product chosen	Confidence in platform	Estimated % of fake reviews	Pay more attention to reviews (experiment site)	Pay more attention to reviews (other sites)	Likely to trust 5-star rating if previous experience was bad
Group 2 (subtle fake reviews + no intervention)	0.0405**	0.0003	0.0002	-0.0007	-0.0008	0.0008
Group 3 (strong fake reviews + no intervention)	-0.0656***	0.0000	-0.0001	-0.0011	-0.0010	0.0002
Group 4 (genuine reviews + informed silence)	-0.0011	-0.3334**	-0.0877*	-0.5148***	-0.5143***	-0.1393
Group 5 (subtle fake reviews + informed silence)	-0.0131	0.3130	0.0589	0.0450	0.0268	0.4728**
Group 6 (strong fake reviews + informed silence)	0.0229	0.5928*	-0.0902	0.0572	0.0476	-0.2339



## Estimating the Impact and Prevalence of Fake Online Reviews

Group 2 (subtle fake reviews + no intervention) x 18-34 years	-0.0135	0.0702***	-0.0184**	-0.0215	-0.0324	-0.0100
Group 3 (strong fake reviews + no intervention) x 18-34 years	0.0219	0.0007	-0.0076	0.0037	-0.0048	-0.0072
Group 4 (genuine reviews + informed silence) x 18-34 years	0.0011	0.3980***	0.0779	0.5112***	0.4685***	0.1507
Group 5 (subtle fake reviews + informed silence) x 18-34 years	0.0130	-0.4542**	-0.0384	-0.0418	0.0102	-0.4632**
Group 6 (strong fake reviews + informed silence) x 18-34 years	-0.0229	-0.6903**	0.1005	-0.0390	-0.0076	0.2832
Group 2 (subtle fake reviews + no intervention) x 35-54 years	-0.0136	-0.0564**	-0.0156*	0.0162	0.0032	-0.0012

## Estimating the Impact and Prevalence of Fake Online Reviews

Group 3 (strong fake reviews + no intervention) x 35-54 years	0.0217	-0.0022	-0.0270***	-0.0005	-0.0107	-0.0166
Group 4 (genuine reviews + informed silence) x 35-54 years	0.0009	0.3125**	0.0729	0.5723***	0.5618***	0.1330
Group 5 (subtle fake reviews + informed silence) x 35-54 years	0.0132	-0.2335	-0.0639	-0.0828	-0.0822	-0.4282**
Group 6 (strong fake reviews + informed silence) x 35-54 years	-0.0227	-0.5676*	0.1311	-0.1274	-0.1180	0.1897
Group 2 (subtle fake reviews + no intervention) x 55+ years	-0.0136	-0.1137***	-0.0179	-0.1183***	-0.1037***	-0.0965***
Group 3 (strong fake reviews + no intervention) x 55+ years	0.0217	-0.1215***	-0.0481***	-0.0400	-0.0542	-0.0767**
Group 4 (genuine	0.0010	0.2399*	0.0589	0.3574***	0.3628***	0.1521

reviews + informed silence) x 55+ years						
Group 5 (subtle fake reviews + informed silence) x 55+ years	0.0132	-0.2299	-0.0683	0.0592	0.0596	-0.3983**
Group 6 (strong fake reviews + informed silence) x 55+ years	-0.0227	-0.4592	0.1276	0.0956	0.1024	0.2275
Observations	3,255	3,255	3,255	3,255	3,255	3,255
R <sup>2</sup>	0.0018	0.0239	0.0171	0.0319	0.0302	0.0158

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.14: Impact of fake reviews and informed silence on participant behaviour in online shopping task (treatment variables interacted with ethnicity)**

Treatment	Product chosen	Confidence in platform	Estimated % of fake reviews	Pay more attention to reviews (experiment site)	Pay more attention to reviews (other sites)	Likely to trust 5-star rating if previous experience was bad
Group 2 (subtle fake reviews + no intervention)	0.0322**	0.0011	-0.0026	-0.0090	-0.0140	0.0026

## Estimating the Impact and Prevalence of Fake Online Reviews

Group 3 (strong fake reviews + no intervention)	-0.0542***	-0.0009	-0.0091	-0.0020	-0.0072	-0.0053
Group 4 (genuine reviews + informed silence)	-0.0010	0.0123	-0.0079	0.0132	0.0073	0.0063
Group 5 (subtle fake reviews + informed silence)	-0.0107	-0.0051	-0.0062	-0.0179	-0.0203	0.0281
Group 6 (strong fake reviews + informed silence)	0.0181	-0.0353*	0.0093	-0.0061	-0.0033	-0.0145
Group 2 (subtle fake reviews + no intervention) x Asian	-0.0118	-0.0643	0.0009	0.0067	0.1301***	-0.1413***
Group 3 (strong fake reviews + no intervention) x Asian	0.0170	-0.0920*	0.0871***	-0.0231	0.0831	-0.1257**
Group 4 (genuine reviews + informed silence) x Asian	-0.0001	0.0347	0.0346*	-0.0630	-0.0578	0.0665

## Estimating the Impact and Prevalence of Fake Online Reviews

Group 5 (subtle fake reviews + informed silence) x 18-Asian	0.0116	-0.1130	-0.0208	0.1861**	0.0663	0.0109
Group 6 (strong fake reviews + informed silence) x 18-Asian	-0.0171	-0.0471	-0.0566*	0.2140**	0.0256	0.1798**
Group 2 (subtle fake reviews + no intervention) x Black	-0.0118	0.1042*	-0.0025	0.0433	0.0414	-0.0890
Group 3 (strong fake reviews + no intervention) x Black	0.0168	-0.1995**	0.0856**	0.3765***	0.3711***	-0.0589
Group 4 (genuine reviews + informed silence) x Black	-0.0001	-0.1449*	0.1597***	-0.0904	-0.2496***	-0.0372
Group 5 (subtle fake reviews + informed silence) x Black	0.0118	-0.1909	-0.1336***	-0.2011	-0.0349	0.2006*
Group 6 (strong fake	-0.0168	0.3795**	-0.2237***	-0.3608**	-0.3002*	0.1556

## Estimating the Impact and Prevalence of Fake Online Reviews

reviews + informed silence) x Black						
Group 2 (subtle fake reviews + no intervention) x Mixed/Other	-0.0119	-0.2093***	0.0193	0.0801	0.1163**	-0.0586
Group 3 (strong fake reviews + no intervention) x Mixed/Other	0.0169	-0.0838	0.0064	0.0834	0.1626***	0.0292
Group 4 (genuine reviews + informed silence) x Mixed/Other	-0.0003	-0.3299***	0.0196	0.0688	0.0681	-0.0667
Group 5 (subtle fake reviews + informed silence) x Mixed/Other	0.0123	0.3715***	-0.0032	-0.0929	-0.2695**	-0.0163
Group 6 (strong fake reviews + informed silence) x Mixed/Other	-0.0168	0.6329***	0.0209	-0.3866***	-0.4729***	-0.3008**
Observations	3,255	3,255	3,255	3,255	3,255	3,255

R <sup>2</sup>	0.0015	0.0242	0.0203	0.0264	0.0292	0.0169
----------------	--------	--------	--------	--------	--------	--------

Note: \* = p<0.1, \*\* = p<0.05, \*\*\* = p < 0.01

**Table A4.15: Consumer harm from fake reviews – change in elicited WTP if fake reviews are viewed**

Product categories impacted by fake reviews	Type of fake review	% of fake reviews	% of fake reviews out of all reviews <sup>67</sup>	Impact on WTP	UK online retail spend via third-party platforms (2022) <sup>68</sup>	Total annual harm
All	Subtle	90%	20%	0.49%	£38b	£165m
	Strong	10%	20%	-0.64%		-£24m
All	Subtle	99%	20%	0.49%		£182m
	Strong	1%	20%	-0.64%		-£2m
All	Subtle	60%	20%	0.49%		£110m
	Strong	40%	20%	-0.64%		-£96m
All	Subtle	90%	10%	0.49%		£82m
	Strong	10%	10%	-0.64%		-£12m
All	Subtle	90%	42% <sup>69</sup>	0.49%		£346m
	Strong	10%	42%	-0.64%		-£50m

<sup>67</sup> 20% is the baseline estimate (with a multiplier of 1) as this was the proportion of fake reviews used in the experiment.

<sup>68</sup> The same adjustment has been made for spending on third-party platforms as in footnote 21.

<sup>69</sup> <https://www.chicagotribune.com/business/ct-biz-amazon-fake-reviews-unreliable-20201020-lfbjdq25azfdpa3iz6hn6zvtwq-story.html>

Table A4.16: Randomisation check

Variable	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Income						
Up to £9,999	16%	13%	13%	14%	12%	16%
£10,000 - £24,999	24%	29%	29%	29%	28%	27%
£25,000 - £49,999	42%	38%	41%	36%	38%	37%
£50,000 - £74,999	8%	8%	9%	8%	10%	9%
£75,000 - £99,999	3%	2%	2%	3%	2%	2%
£100,000 or more	1%	1%	1%	2%	2%	1%
Prefer not to answer	6%	9%	6%	8%	8%	8%
Highest level of education completed						
Less than primary school / primary school not completed	0%	0%	0%	0%	0%	0%
Primary	0%	0%	0%	0%	0%	0%
Secondary	19%	17%	16%	21%	20%	23%
Vocational	16%	16%	17%	15%	18%	17%
Undergraduate	40%	44%	44%	39%	40%	44%
Postgraduate	23%	22%	22%	23%	21%	15%
Prefer not to answer	1%	2%	1%	2%	2%	1%



## Estimating the Impact and Prevalence of Fake Online Reviews

Age						
18-24 years old	14%	12%	8%	10%	10%	11%
25-34 years old	35%	33%	34%	30%	36%	32%
35-44 years old	28%	25%	28%	29%	26%	27%
45-54 years old	13%	16%	16%	16%	12%	18%
55-64 years old	7%	8%	11%	10%	10%	10%
65 years or older	3%	5%	4%	5%	6%	2%
Prefer not to answer	0%	1%	0%	1%	0%	1%
Sex						
Female	48%	55%	49%	50%	51%	52%
Male	51%	44%	50%	49%	48%	47%
Intersex	0%	0%	0%	0%	0%	0%
Prefer not to answer	1%	0%	0%	1%	1%	1%
Ethnicity						
White	88%	86%	84%	88%	90%	87%
Mixed/multiple ethnic groups	4%	3%	3%	2%	2%	4%
Asian/Asian British	5%	6%	8%	6%	6%	6%
Black / African / Caribbean / Black British	2%	2%	4%	2%	2%	2%

## Estimating the Impact and Prevalence of Fake Online Reviews

Other ethnic groups	1%	2%	1%	0%	0%	0%
Prefer not to answer	1%	1%	1%	2%	1%	1%
Frequency of online shopping						
more than once a week	18%	18%	19%	20%	15%	14%
about once a week	26%	27%	31%	23%	26%	29%
several times a month	34%	33%	30%	32%	35%	34%
about once a month	16%	17%	15%	15%	17%	17%
once in a few months or longer	6%	6%	5%	9%	7%	6%
Frequency of Amazon purchases						
more than once a week	10%	10%	10%	11%	8%	7%
about once a week	14%	16%	19%	18%	17%	17%
several times a month	32%	30%	32%	28%	31%	33%
about once a month	26%	22%	21%	20%	23%	23%
once in a few months or longer	16%	20%	15%	19%	19%	18%
never	1%	1%	2%	3%	2%	2%

## Appendix 5: Regression details

In our main specifications, we used linear probability models (LPM) and ordinary least squares (OLS) to examine the effects of different types of fake reviews and an “informed silence” text banner on participants’ purchasing choices and willingness to pay. More specifically, for participant  $i$  assigned to purchase product type  $j$  :

$$Y_i = a_0 + \beta_1 SUBTLE_i + \beta_2 STRONG_i + \beta_3 INFORMED\_SILENCE_i + \beta_4 SUBTLE_i * INFORMED\_SILENCE_i + \beta_5 STRONG_i * INFORMED\_SILENCE_i + X' \lambda_i + \delta_j + \epsilon_i$$

Where:

$Y_i$  is a binary variable equalling 1 if the participant chose a product and 0 otherwise, or the participant’s stated willingness to pay for the product

$SUBTLE_i$  is a dummy variable equalling 1 if the participant saw subtle fake reviews and 0 otherwise

$STRONG_i$  is a dummy variable equalling 1 if the participant saw strong fake reviews and 0 otherwise

$INFORMED\_SILENCE_i$  is a dummy variable equalling 1 if the participant saw a text banner stating fake products/reviews had been removed by the platform and 0 otherwise

$X'$  is a matrix of individual characteristics, specifically age, income and gender (all transformed to dummy variables)

$\delta_j$  are product fixed effects

$\epsilon_i$  is the regression error term

$\beta_1$  and  $\beta_2$  reveal how participant choice of product and willingness to pay are impacted by the presence of subtle or strong fake reviews compared to participants who only viewed genuine product reviews. Positive and significant coefficients would suggest that fake reviews were effective in increasing participant perception of a product, while negative and significant coefficients would suggest that fake reviews were not effective and decreasing participant perception of a product.

Similarly,  $\beta_3$  indicates how participants are impacted by the presence of the informed silence informational text box in the absence of fake reviews.  $\beta_4$  indicates how participants are impacted by the presence of the informed silence text box while subtle fake reviews are also present, and  $\beta_5$  indicates how participants are impacted by the presence of the informed silence text box while strong fake reviews are also present.



© Crown copyright 2023

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://nationalarchives.gov.uk/doc/open-government-licence/version/3) or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: [psi@nationalarchives.gov.uk](mailto:psi@nationalarchives.gov.uk). Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication available from <https://www.gov.uk/government/organisations/department-for-business-and-trade>

Contact us if you have any enquiries about this publication, including requests for alternative formats, at:

Department for Business and Trade  
Old Admiralty Building  
London SW1A 2DY  
Tel: +44 (0) 20 4551 0011

Email: [enquiries@trade.gov.uk](mailto:enquiries@trade.gov.uk)