The real impact of fake reviews

a behavioural experiment on how fake reviews influence consumer choices

Contents

Introduction

One of the most striking ways in which the internet has transformed how consumers make purchases is the unprecedented access it has given them to the experiences and opinions of other customers. Consumers heavily rely on online customer reviews (OCRs) as a way to aid selection between a range of possible options and it has been widely found that reviews can have a strong effect on sales.

However, consumer trust in these reviews may often be misplaced as there is growing evidence that reviews are faked to manipulate consumer decisions. Which? has investigated the spread of false and misleading information through OCRs for several years. Our research and investigations have identified myriad instances of products and services with suspicious reviews posted on popular sites like Amazon, Tripadvisor and ebay. This included instances of reviews being traded on social media, sellers offering customers free items or cash incentives for positive reviews and the use of paid 'review farms' or bots to flood products with five star reviews.

The potential for these activities to create consumer harm is clear, with consumers exposed to misleading information being susceptible to making poor decisions. Which?'s investigative research has found that in many cases of fake reviews, the products have been of exceptionally poor quality and, in the worst circumstances, unsafe. Furthermore, some of the platforms hosting these reviews have even given their explicit endorsements to products harbouring fake review material.

While there is a rich literature on how reviews influence consumer choice, there has been little research specifically on the impact of fake reviews so the extent of the harm they cause is impossible to estimate. Our research addresses this by developing understanding of the extent to which consumer behaviour can be influenced by fake reviews.

Working with the research consultancy The Behaviouralist, we developed an experiment to test how consumer behaviour would change in the presence of fake reviews. The experiment required consumers to choose a product from five identically-priced items for one of three product types (headphones, dash cams and cordless vacuum cleaners) on screens designed to look like the Amazon website. It examined how adding fake reviews to the products influenced the consumers' choices and to what extent. Different types and levels of fakery were explored and the experiment included a treatment to test whether consumer education could help to mitigate the influence of fake or suspicious reviews.

The pages and content that were shown to participants in the experiment were not real Amazon pages or content. Amazon states that it has policies that prohibit abuse of its community features, including reviews, and that it suspends, bans and takes legal action against those who violate these policies. Amazon also states it has systems and tools in place aiming to stop abusive reviews before they are ever published, and allowing customers concerned about the authenticity of reviews left on a product to use the "report abuse" link, available on each review.

The experiment was run online with 9,988 consumers and gave us strong and robust evidence that fake reviews can be used to mislead consumers into making poor choices. **We found that adding fake**

reviews to poor quality (Which? Don't Buy) products substantially increased the number of people choosing them as their favoured option. This effect was present when just manipulating the star ratings of the products, but became even stronger when the text of the reviews was manipulated and was even the case when clear signs of that manipulation were present. Furthermore, **we found that review platform endorsements could exacerbate the effects of fake reviews**.

Finally, we tested the inclusion of a banner on the webpages warning people of the possible presence of fake reviews and some tips to help people spot them. This was effective in reducing the proportion of people picking the product with fake reviews but still left a significantly larger proportion choosing that product than when its reviews were not manipulated.

The results are consistent across demographic groups, while more frequent online shoppers were more likely to buy products with fake reviews. The results are also consistent across product types, indicating that consumers are likely to be deceived at a range of price points. Indeed, in our experiment consumers were most likely to buy as a result of fake reviews in the most expensive product category, where consumer harm is likely to be greatest. Finally, we found that fake reviews attracted choices away from all other products, but disproportionately so from the best (Which? Best Buy) alternative.

This research provides robust evidence of the harm from fake reviews and highlights the clear need for action to stop fake reviews. The results demonstrate that the harm arising from fake reviews could be vast, with many consumers at risk of being misled into spending their money on substandard products.

The report is structured as follows. First, we provide a brief review of the literature on online product reviews and fake online reviews and we set out our research objectives. Next, we explain in detail the design and structure of the experiment and finally we look at the results and their implications.

The potential impact of fake reviews

## The effects of online customer reviews generally

There is a comprehensive literature on the extent to which online customer reviews (OCRs) affect consumers' decision making. OCRs have been studied in a large variety of different product and service markets including 'experience goods' like books, music, TV shows, films and 'search goods' like digital cameras, consumer electronics and e-book readers (e.g. Floyd et al, 2014).

The majority of studies find that reviews can indeed influence consumer decision-making. In summary, this influence is due to different types of information that consumers receive from OCRs, which interact with each other and with product attributes, summarised in Figure 1 below. The key features of OCRs that are generally identified in the literature are valence, volume and variation. These can be defined as followed:

- Valence is the overall sentiment of the reviews (positive or negative). Most sites summarise the valence through a star rating or mark out of 5 or 10.
- Volume is the total number of reviews posted on a particular product or service.
- Variation is the spread of reviews across different valences, often presented as a distribution across the different star ratings given to the product or service.

Generally, consumers will take these into account when making their product choices, considering them alongside the attributes of the product like brand, price or technical specifications.

Figure 1 - How information from customer reviews influences consumer choice
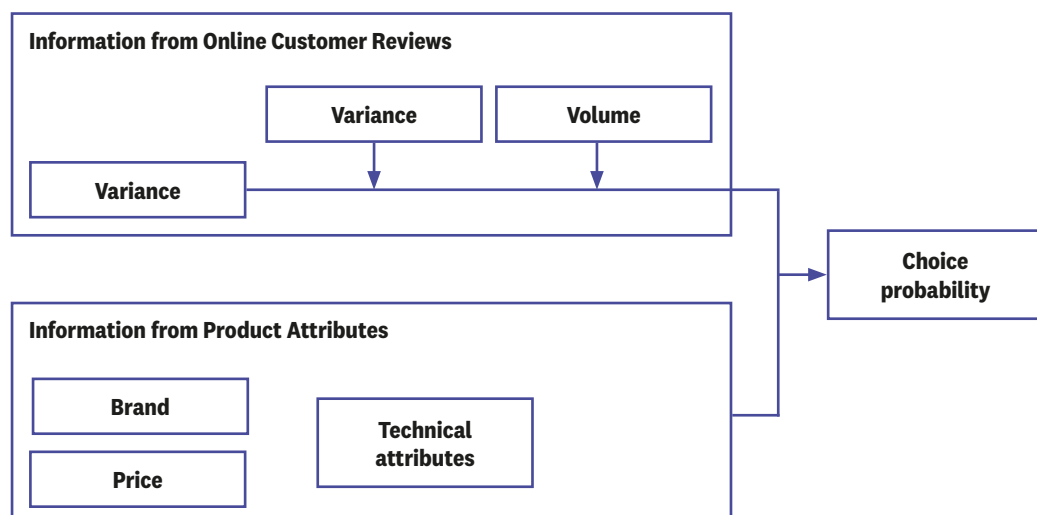
Source: Kostyra et al (2016), Decomposing the effects of online reviews on brand, price and product attributes

Studies find that consumers can use the valence of reviews as a heuristic to draw conclusions about product quality (Zhang, 2014 and Malthouse et al, 2017). These conclusions then support consumer choice, with positive valence associated with a lower risk of making a bad choice,

reducing consumers' price sensitivity and increasing their willingness to pay for products (Kostyra et al, 2016). Empirically, the impact of review valence on consumer choices is supported across a wide range of markets with valence having an effect on the sales of books (Chevalier and Mayzlin, 2006), movies (Chintaqunta et al, 2010), mobile phones (Gopinath et al, 2014), electronics (Archak et al, 2011) and beer (Clemons et al, 2006). The findings are not universal however, with some studies failing to find a link between valence and sales.

The number of reviews left by customers can signal the popularity of a brand, product or service and lead to higher sales. However, higher volume typically influences consumers through its interaction with valence, showing a level of consensus around a product or service's quality. For instance, a positive rating can increase sales even more if it is based on a higher number of reviews. In some cases however, the mere presence of reviews could lead to more sales, even without a positive valence.

Variation in reviews is also closely linked with volume and valence, with highly divergent views found to moderate the effects of valence. For instance, higher variation was associated with a higher purchase probability for poorly-rated books but a lower purchase probability for highly-rated books (Sun, 2012).

Customer reviews can be an important and influential part of a customer's shopping experience, affecting the likelihood of choosing items and how much they are willing to pay for goods. But factors like price, brand and technical features remain highly important in determining product or service choice. The effects of these factors and information from reviews are however interlinked, with each affecting the influence of the other in the purchase decision. For example, while a high number of positive reviews may allow a seller to charge a higher price, a higher price also magnifies the decision risk and can thus reduce the impact of the heuristic elements of reviews like star ratings. Similarly, higher priced items may need more reviews for the valence to affect the decision making.

Online reviews may also be more influential when left for products where the branding is weaker, and conversely less effective where strong branding or technical attributes are important. In the context of fake reviews, this may be important as our investigatory work has identified frequent instances of products from unknown brands using fake reviews to hastily build a reputation for their goods being sold on online platforms.

The influence of reviews on consumers may differ between different demographics. One study (von Helversen et al, 2018) found that younger consumers used more of the review information available to them (volume, valence, variation) but a single, emotionally positive or negative review could override their preference for a higher-rated product. Older consumers on the other hand were not influenced by ratings or individual positive reviews, but were strongly influenced by individual negative reviews.

Implications for fake reviews

Most of the literature on the effect of reviews does not explore the level of honesty included in opinions left by reviewers, but it does have important implications for how influential we believe that fake reviews could be. Most papers find a link between reviews and consumer purchase behaviour. Thus there is an incentive for sellers to manipulate reviews in order to increase sales, weaken incumbent brand power and increase consumer willingness-to-pay.

Those looking to manipulate reviews have several means of doing so, principally by increasing the overall rating of the products (valence). The effect of increased valence can also be amplified by increasing the volume of reviews and improving the distribution of reviews across ratings.

Furthermore, even single, high-quality fakes could be effective in driving consumer behaviour toward purchases and even override considerations of average ratings.

Other external factors are also important in their interaction with how reviews (and hence also fake reviews) influence decision making. The literature suggests that fake reviews are likely to be most effective where:

- **Technical characteristics of products are harder to establish.** Consumers rely more heavily on the experiences of others where hard facts aren't known
- **Weak branding is present.** Increases in willingness-to-pay associated with OCRs are higher for products with weaker brands. This also increases the incentives for firms with weak brands to engage in deception.
- **Cheap products have inflated ratings.** Consumers will use simple measures of valence and volume as heuristics with cheaper products, meaning that simple manipulation of star ratings could be effective without consumers reading the review content.
- **Expensive products have high quality review manipulation.** Consumers use more information to inform their purchases in order to reduce risk, engaging systematic thought processes rather than heuristics. This could lead to consumers more easily identifying weak deception but could also leave them more vulnerable to convincing deception.
- **High quality, affect-rich reviews are on display.** The content of reviews can be influential and even a single, randomly chosen, emotional review can override a lower average score.

Of course, if consumers are able to anticipate some level of deception then they may take this into account when making decisions and fake reviews will have less impact. The available evidence of whether this occurs in practice is scarce and inconclusive. One study concluded that, as not all firms deceive, consumers will not be able to fully protect themselves against manipulation (Hu et al, 2011). Whereas others have found that manipulation of OCRs of books sold through Amazon did not lead to increased sales, suggesting that consumers are able to correct for deception (Hu et al, 2012).

We know from the literature that fake reviews are present on online platforms, although estimates of the proportion of reviews that are manipulated vary significantly. Investigative research by Which? and others has found many different ways in which platforms have engaged in various types of review manipulation and fakery. These range from simple activities like soliciting reviews from friends and family to much more sophisticated methods where hundreds or thousands of reviews can be procured through the use of bots or paid-for 'reputation management' services and even bought or sold through popular online marketplaces. Other methods include review hijacking or merging, where positive reviews for similar or, in some cases, totally different products can be shared or appropriated.

The fact that firms are willing to incur costs to procure fake reviews does suggest that they hold the potential to influence sales. Key questions remain however on what this effect is, how large it might be or for which products it might be most effective.

Our research questions

There is a clear gap in the current research on how fake reviews in particular could influence consumer behaviour and how big the resulting harm could be. Thus we formed our research questions around this gap, taking into account what we know about how reviews are used and how they influence both from the academic research and our in-house expertise on the practices and incidence of reviews.

The key questions we designed the research to answer are:

1. Can fake reviews influence what products consumers buy and to what extent?
2. Can fake reviews make people more likely to choose poor-value products?
3. How sophisticated does review fakery need to be to influence purchase decisions?
4. Can platform endorsements whose award may be linked to fake reviews, lead to greater harm?
5. Can consumer education mitigate the influence of fake or suspicious reviews?

By answering these questions, we should gain insight into the types of harms that may be caused by the presence of fake reviews, and how significant those harms could be.

Experimental design

We worked with research consultancy The Behaviouralist to develop a behavioural experiment to answer our research questions on how consumer behaviour changes in the presence of fake reviews. Together we designed an online survey-based randomised control trial, where consumers were tasked with choosing a product from five identically-priced items (e.g. headphones) presented on screens designed to look like the Amazon website.

In order to isolate the effects of fake reviews on the choices, each respondent was randomly assigned to one of six groups. One of these was a control group, while the other treatment groups had varying degrees of fake reviews added to one of the products displayed. The fake review content on the pages participants saw were not real but incorporated features that Which? has seen in real life. They were not taken from actual Amazon pages and in real life these particular fake reviews may have been prevented from appearing on the site. In one treatment group the product with fake reviews also received a platform endorsement, while in another group the respondents saw a banner that warned that some reviews may be fake. The details of each of these treatment groups are described below, but first we describe the shopping task.

The shopping task

The survey environment presented to participants was kept as close as possible to a normal retail environment. This included offering both mobile and desktop versions of the survey depending on the device the participant chose to complete the survey on. Roughly half the sample used mobile and half used desktop, which appears to be a reasonable reflection of online shopping habits (IMRG, 2016).

Each participant made a single choice from a set of products in one of three product categories, which were headphones, dash cams and cordless vacuum cleaners. Which? has previously observed the use of fake reviews for poor quality products in each of these categories (see Which? Magazine, November 2019), while they also provide a range of price points, since it is known that consumers use reviews in different ways depending on the price of their purchase. For instance, Floyd et al (2014) suggest higher priced products are more likely to engage consumers' systematic information processing, paying closer attention to the deeper content of reviews rather than relying on ratings as heuristics. Similarly Malthouse et al (2017) found that when consumers shop for relatively inexpensive items, they may need a smaller number of reviews for the valence to influence their decision making.

Within each product category, the specific products chosen were all available at or near a single price and in the experiment were fixed to a single price so that the effects of reviews on the decision-making could be isolated. These were £25 for headphones, £50 for dash cams and £150 for cordless vacuum cleaners.

In order to incentivise respondents to honestly choose the product they liked best, they were provided with a chance of winning the product they chose in a prize draw. However, to further increase motivation, participants were allowed to choose a product category. This led to larger proportions of respondents choosing headphones and vacuum cleaners, so that some consumers were randomly allocated to dash cams to ensure a minimum sample size of 2,000 for that product.

In each of the product categories, we gave respondents a choice of five real products of varying quality, making use of Which?'s expert product testing. This included one Best Buy, one Don't Buy product, and three products that had not been reviewed by Which? or had been reviewed but not awarded either Best Buy or Don't Buy status. Participants in the experiment were not shown any information about the Which? reviews or scores for any of the products.
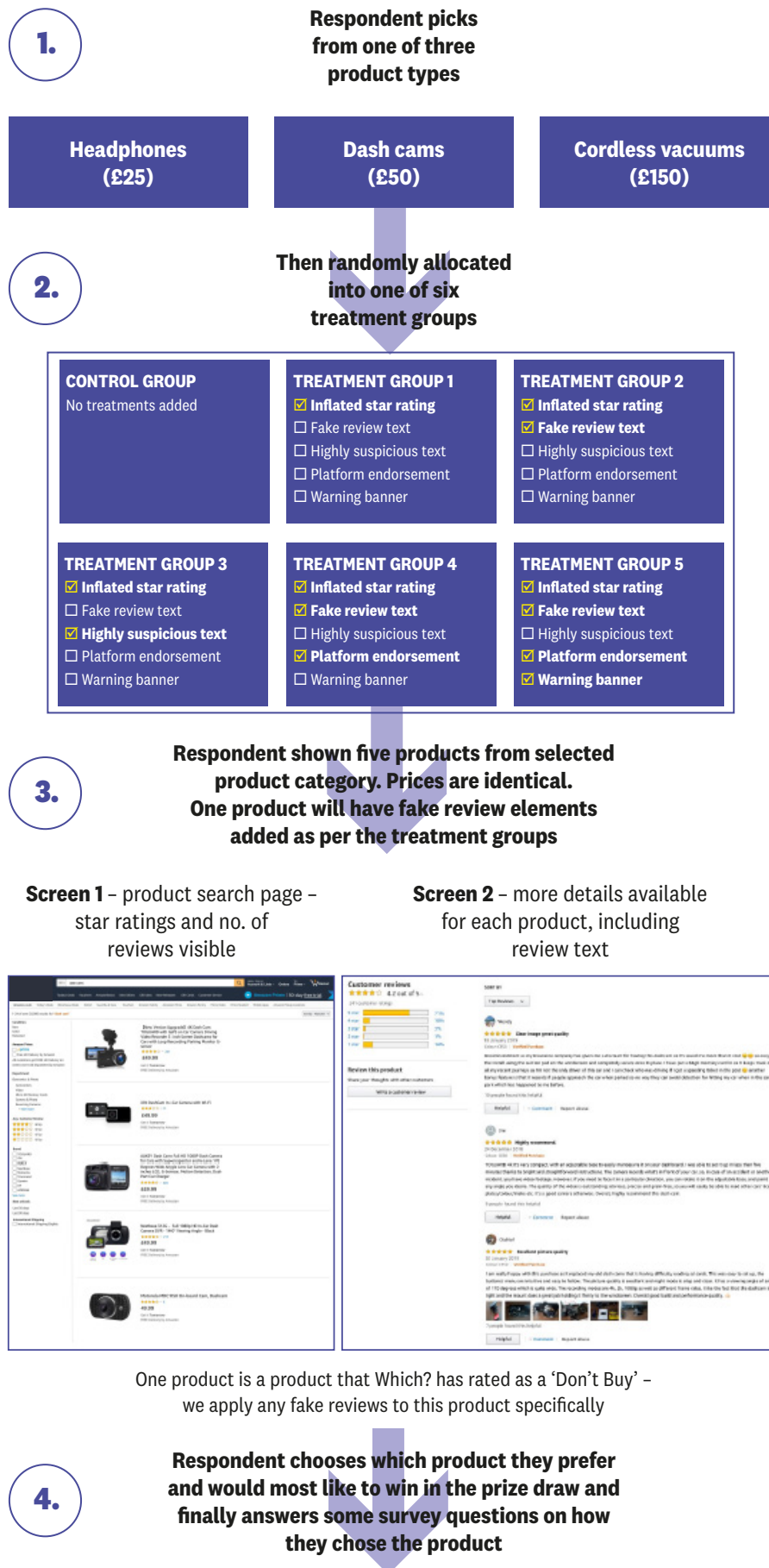
It is not necessary for the products to be of varying quality for fake reviews to create harm. If the fake reviews distort consumer choice so that they make purchases they would otherwise not have made in the absence of fake reviews, then this could constitute harm. However, by including both Best Buy and Don't Buy products we see whether consumers buy poor quality items as a result of fake reviews, and whether these sales come at the expense of high-quality items. Clearly, if this is the case then, given prices are identical, fake reviews cause harm and reduce consumer welfare. We therefore add the fake review content in our treatment groups to the Don't Buy products only.

In reality, fake reviews could be applied to either high-quality or poor quality items. However, our own investigative work has found that products with fake reviews typically perform very badly under our own rigorous testing. Empirically, there is also academic evidence that the degree of fakery increases as quality decreases i.e. the worst products have the most deception (Hu et al, 2011). Intuitively, this makes sense as those selling high-quality products should have the smallest incentives to produce fake reviews, given that the natural review content should reveal the high quality.

Participants in the experiment are initially shown a screen designed to look as though they had just searched for their product category on Amazon.co.uk, with the five products of varying quality displayed in a list. This page contains some information about the reviews of the products including the average star rating and number of reviews, but the actual content of the reviews is not visible. Participants are then given the option of reading more about any of the products by clicking through to screens designed to look like Amazon.co.uk product pages. These contain additional product information as well as the distribution of star ratings and a selection of seven written reviews. Participants are able to read as many product page screens as they like and then choose whichever product they would most like to win in the prize draw.

Before beginning the shopping task the participants were randomised into one of six experimental groups which determined the fake review treatments to which they would be exposed during the task.

Figure 2 – Experimental design

1. Respondent picks from one of three product types

| Headphones (£25) | Dash cams (£50) | Cordless vacuums (£150) |

2. Then randomly allocated into one of six treatment groups

CONTROL GROUP
No treatments added

TREATMENT GROUP 1
☑ **Inflated star rating**
☐ Fake review text
☐ Highly suspicious text
☐ Platform endorsement
☐ Warning banner

TREATMENT GROUP 2
☑ **Inflated star rating**
☑ **Fake review text**
☐ Highly suspicious text
☐ Platform endorsement
☐ Warning banner

TREATMENT GROUP 3
☑ **Inflated star rating**
☐ Fake review text
☑ **Highly suspicious text**
☐ Platform endorsement
☐ Warning banner

TREATMENT GROUP 4
☑ **Inflated star rating**
☑ **Fake review text**
☐ Highly suspicious text
☑ **Platform endorsement**
☐ Warning banner

TREATMENT GROUP 5
☑ **Inflated star rating**
☑ **Fake review text**
☐ Highly suspicious text
☑ **Platform endorsement**
☑ **Warning banner**

3. Respondent shown five products from selected product category. Prices are identical. One product will have fake review elements added as per the treatment groups

Screen 1 – product search page – star ratings and no. of reviews visible

Screen 2 – more details available for each product, including review text

One product is a product that Which? has rated as a 'Don't Buy' – we apply any fake reviews to this product specifically

4. Respondent chooses which product they prefer and would most like to win in the prize draw and finally answers some survey questions on how they chose the product

Fake reviews (inflated star ratings and fake review text)

As described above, the volume and valence of reviews affect their influence on consumer choice, but a small number of affect-rich reviews can also influence consumers. Therefore, we include treatments in the experiment to capture the effect of fakery for both of these. In all treatments the fake review content on the pages that participants saw was intended to recreate examples of suspicious reviews that Which? has seen. They were not real Amazon pages therefore it is possible that in real life these particular fake reviews may have been prevented from appearing on the site.

To begin with, the Don't Buy product is given an inflated number of five star reviews, which significantly improves its overall star rating, see Table 1 and Figure 3. This mirrors actions a company may take to artificially increase both the volume and valence of its reviews, as it either buys additional reviews, employs so-called 'reputation management' services or uses bots to post a large quantity of five star reviews on its products.

While this level of review manipulation can be hard to detect, consumers may be suspicious when unbranded products have far more reviews than branded products for which greater sales may be expected. Suspicion may also be raised as the star rating approaches the five star mark and the distribution of reviews is increasingly skewed towards the extremes.

Table 1 – product ratings for Which? don't buy products within each product category before and after inflated star rating treatment added

Don't buy product	Star rating out of 5 (without fake reviews) (volume)	Star rating out of 5 (with fake reviews) (volume)
Headphones	2.9 (155)	4.7 (2,514)
Dash cams	3.2 (11)	4.8 (1,248)
Cordless vacuums	3.8 (223)	4.8 (1,223)

Figure 3 – star ratings for headphones in the control group and treatment groups

Control group / Inflated star rating treatment
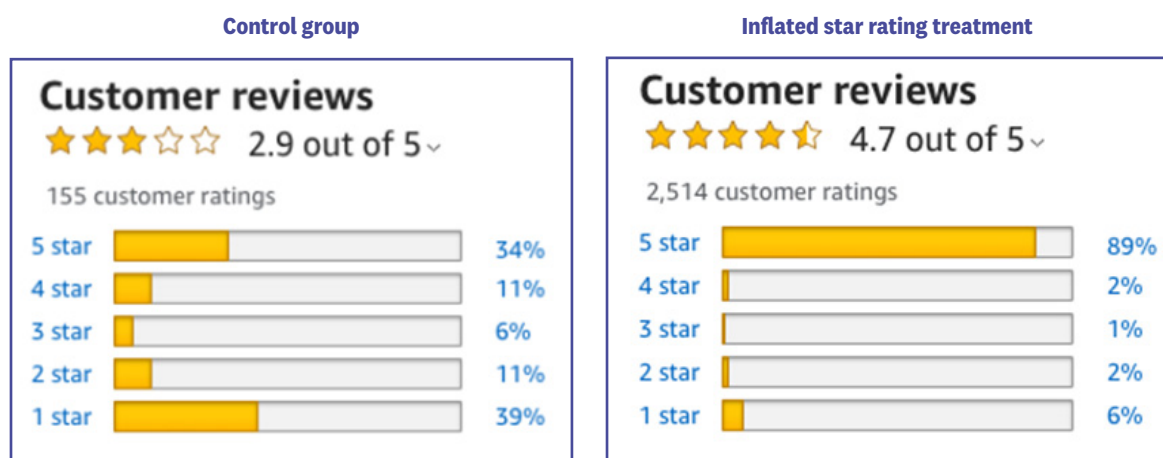
Beyond the direct effect of the star rating, we also included an effect of the fake review product moving up to the top of the search listings. While platforms algorithms for where products are displayed on a search page are generally not in the public domain, sellers using Amazon and other review platforms understand that reviews, directly or indirectly, play a role in deciding where their products are placed. Hence, we included the effect of moving up the rankings here as well.

While this does mean that we cannot separate out the effects of ranking change and star-rating change specifically, we believe in reality the two happen simultaneously and so this treatment gives us the best likely reflection of the real-life effects.

To investigate the impact of businesses creating a small number of affect-rich fake reviews, we next created reviews that had highly favourable text in general, keeping only one review that would appear to contradict the praise given in the others. In total, seven reviews were available for participants to read. In order that we did not simply recreate the effects of positive reviews more generally, we aimed to keep some elements that would raise suspicion among those very familiar with spotting fake reviews, including:

- Exaggerated language praising the products
- Repetitive phrases and formatting among reviews
- Fewer 'verified purchase' reviews
- Several reviews left on the same date
- Same reviewer leaving two reviews
- A review left by 'Amazon Customer'
- One negative review contradicting the positive feedback

These suspicious features have been spotted on many occasions across various review platforms by our investigative researchers and some also feed into algorithms employed by websites that assess the overall trustworthiness of reviews.[1] While many of these features are subtle and may take significant concentration to detect, they are reflective of real-life examples we have uncovered. However it is also worth noting that in real life such signs of suspicion may also have been picked up by Amazon's systems and prevented from appearing.

The expected effect here is not necessarily clear cut. Research on reviews more generally suggests that including positive text in reviews would improve the choice probability of a product, but if consumers are very savvy and can spot signs of suspicion then we might not see any additional effect.

Figure 4 – example of suspicious features in two reviews included in the fake review text treatment

1   For example FakeSpot or ReviewMeta
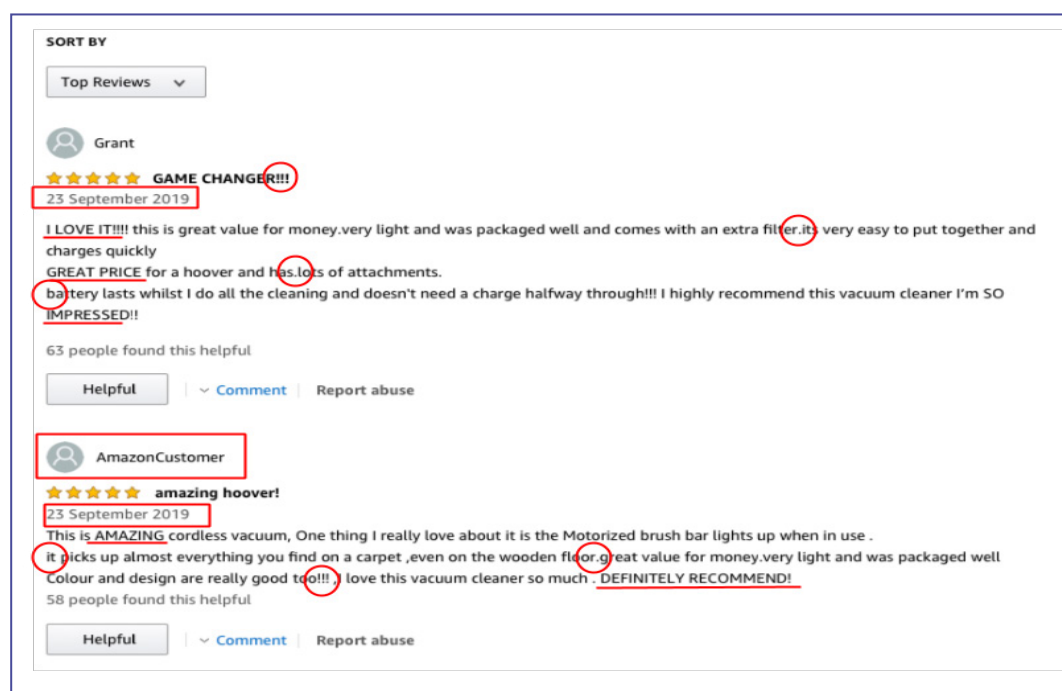
Suspicious fake review text

The elements outlined above that would raise suspicion for an expert may be too subtle for a normal consumer to detect, so we also create a treatment to test the effects of adding additional suspicious elements to the review text. Many of the features from the fake review text remained in place, but the following suspicious features identified in our investigative research were also added:

- Reviews for different products being in place. This could be symptomatic of poor black-hat services or the hijacking of reviews from other products. In our case we included two five star face cream reviews among the 7 reviews on display
- Admissions of incentivisation being offered to leave positive reviews or change negative reviews
- Negative reviews claiming that they have been offered incentives to change their reviews

Figure 5 – elements of incentivisation and reviews of other products added in the suspicious review text treatment
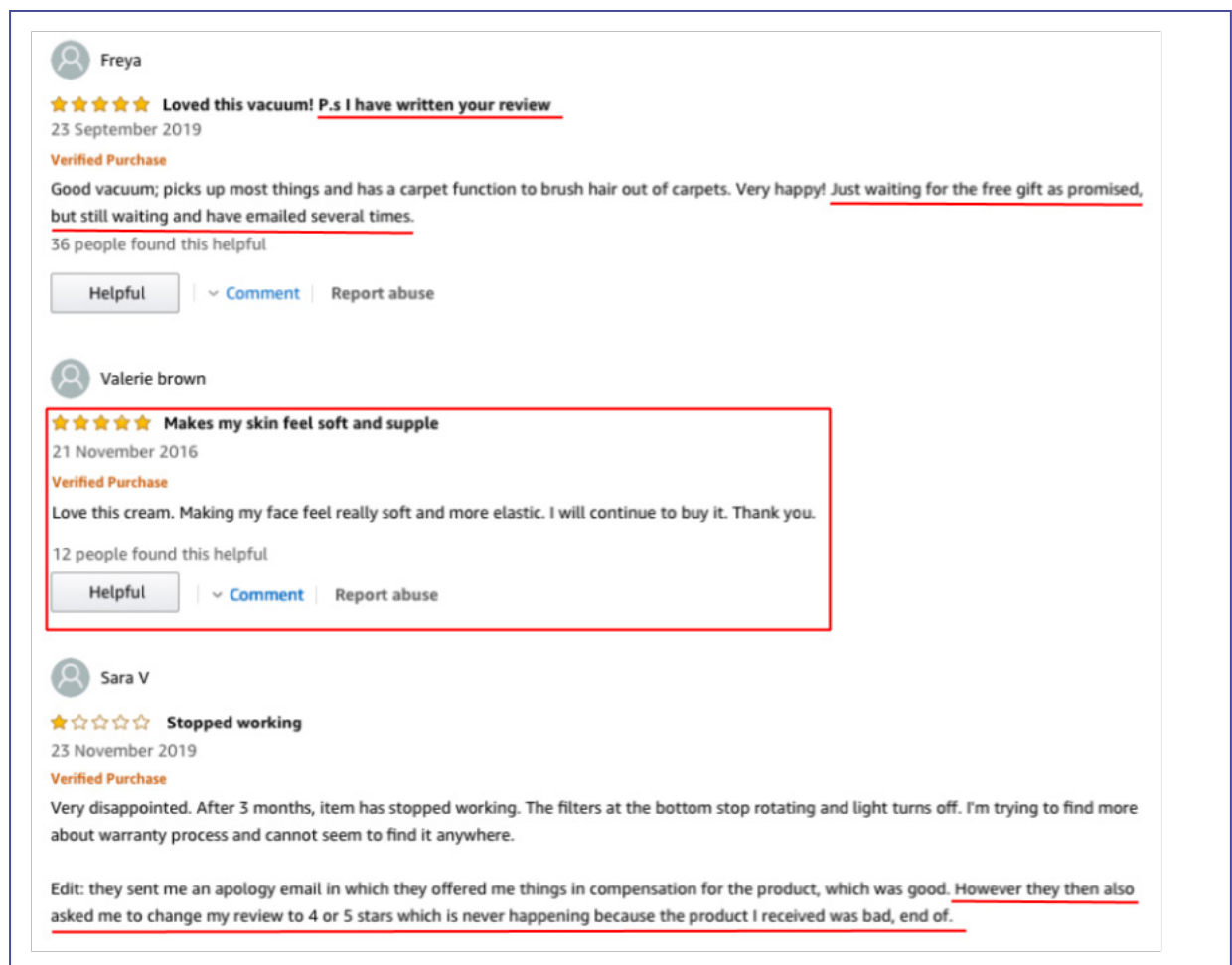
The purpose of this treatment was to see whether the effects of fake reviews would be undermined where signs of review manipulation are more obvious. If these reviews still increase choice probability of the poor quality product, then it may suggest that those seeking to manipulate their reviews do not need to be particularly sophisticated in order to successfully drive sales.

Platform endorsements

Some online review platforms give endorsement labels to products or services receiving particularly good customer feedback, providing a potential route for those manipulating reviews to extend the influence of that manipulation. In our experimental setting, we explored the impact of these endorsements on consumer choice through the inclusion of an "Amazon's Choice"-style treatment. While such endorsements are not an element of fake reviews as such, some of Which?'s recent investigative research has found instances where fake reviews had contributed to products receiving a platform endorsement.[2]

If fake reviews with a platform endorsement label can affect consumer choice even more than without it, then endorsement labels based on misleading customer feedback might have the potential to create significant additional harm.
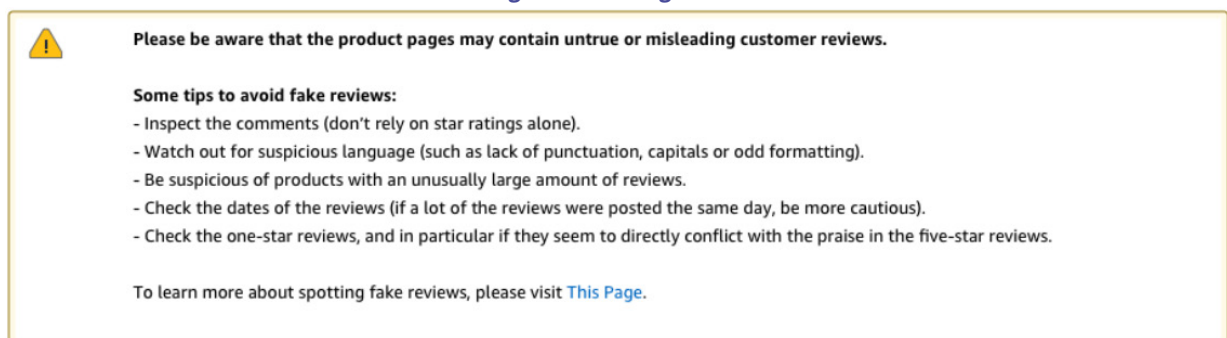
Warning banner

For our final treatment, we wanted to test whether any consumer-focused interventions to help consumers avoid the influence of fake reviews could be effective.

In the literature thus far, there is some evidence that priming consumers with a news article about fake reviews prior to them making their choice could make them less trusting of reviews and less likely to be fooled by possible deception (Munzel, 2016). However, the discussion of interventions on fake reviews has focused on how review platforms themselves can identify and remove fake reviews. Such interventions to remove fake reviews should be the priority. However, there could still be a role for raising consumer awareness to reduce their susceptibility to deceptive practices that slip the net. Ideally, such interventions should work in tandem with supply-side action rather than in place of it.

In our study we tested one simple intervention in the form of a banner at the top of all of the shopping task screens shown to the participants of the experiment. This banner contained a warning about the possible presence of fake reviews and some tips consumers could follow to avoid their influence. It was a simple and non-targeted intervention present at the top of every single page that the participants viewed. The intervention was not designed to be representative of the whole suite of consumer-focused remedies that could be adopted by review platforms but as a demonstration of how effective a simple and easy-to-adopt intervention might be in reducing the harm of fake reviews.

Figure 6 – Warning banner

⚠ **Please be aware that the product pages may contain untrue or misleading customer reviews.**

Some tips to avoid fake reviews:
- Inspect the comments (don't rely on star ratings alone).
- Watch out for suspicious language (such as lack of punctuation, capitals or odd formatting).
- Be suspicious of products with an unusually large amount of reviews.
- Check the dates of the reviews (if a lot of the reviews were posted the same day, be more cautious).
- Check the one-star reviews, and in particular if they seem to directly conflict with the praise in the five-star reviews.

To learn more about spotting fake reviews, please visit This Page.

2   Which? (2020), *Amazon's Choice badges removed after Which? Investigation*, Which? Magazine March 2020 issue

Control and treatment groups

In total the experiment has five treatment groups, each with different configurations of the treatments. Where we add fake reviews in a treatment group, these are always applied to the Don't Buy product *only*. We can then measure changes in product choices in each group against the control group (where no treatments were given) in order to find the effect of whichever treatments have been given. For example we will be able to find out whether inflating the star rating of the Don't Buy product increases the proportion of participants choosing it.

The control group and treatment groups were configured as shown in Table 2 below

Table 2 – treatments applied in the treatment groups

Group	Inflated star rating	Fake review text	Suspicious text	Platform endorsement	Warning banner
Control					
Treatment Group 1	X				
Treatment Group 2	X	X			
Treatment Group 3	X		X		
Treatment Group 4	X	X		X	
Treatment Group 5	X	X		X	X

Our approach was to add the treatment elements sequentially in each of the treatment groups e.g. to first examine the effects of inflating the star rating in Treatment Group 1 and then examining the effects of inflating the star rating and including fake text in the reviews in Treatment Group 2. By configuring the groups in this way we were able to estimate effects both by comparing treatment groups with the control group and also between treatment groups to get additive effects.

Survey questions

Beyond simply measuring the effects on whether participants in the experiment chose the Don't Buy product, we also asked participants a range of questions on their demographics, shopping habits and how they made their choice in the survey. This allows us to separate the results by different types of consumer e.g. different ages, regions or those who shop online frequently. It can also give us further insight into whether those who made their decisions in different ways might be more or less susceptible to the influence of fake reviews.

Results

Fieldwork and sample

The experiment was run in February 2020 with 9,988 respondents. Overall, our panel of participants was nationally representative across age groups, genders and regions. We also achieved a good spread between different education and income levels. Given our participants took the survey online, all of them were active users of the internet but time spent using the internet varied, with around 46% using the internet for at least 3 hours per day. Our participants were also reasonably active online shoppers, with 78% making a purchase at least once or twice per month. Similarly, 59% of the participants shop on Amazon at least once or twice per month and 82% at least once every three months, meaning that they were familiar with the general set-up of the online shopping task.

We found that all of these characteristics were balanced across our control and treatment groups, indicating that randomisation of participants was successful. This is crucial as it means we can reliably attribute differences between the choices of the groups to the treatments rather than any demographic differences between the participants in each group. Full demographic and experimental balance tables can be found in Annex 1

Control group

The distribution of choices made by participants in the control group are presented in Figure 7. In the absence of any fake reviews, just over 10% of consumers chose the low quality Don't Buy product out of the five options provided. The Best Buy was disproportionately more chosen, while the other three filler products attracted about three-fifths of all choices. This is reasonably suggestive of a retailing and review system that works well in guiding consumers towards good choices and away from poor choices, given that only around one in ten consumers decided to buy the poorest-quality products.

Figure 7 – proportion of control group respondents choosing each product quality

Product	Proportion
Don't Buy	10.5%
Filler	61.2%
Best Buy	28.3%
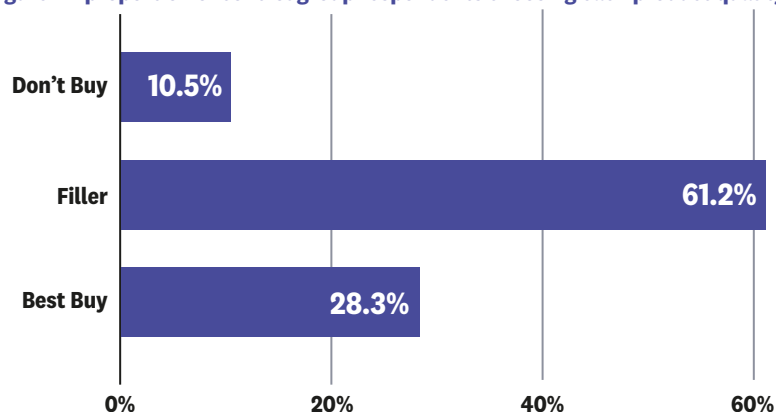
Main findings

We now present the main results in each of the treatment groups. We focus on our main outcome of interest, the proportion of all participants choosing the Don't Buy product. All of the results presented are averaged across product types and are statistically significant.

Figure 8 – proportion of participants choosing a Don't Buy product in each of the control and treatment groups
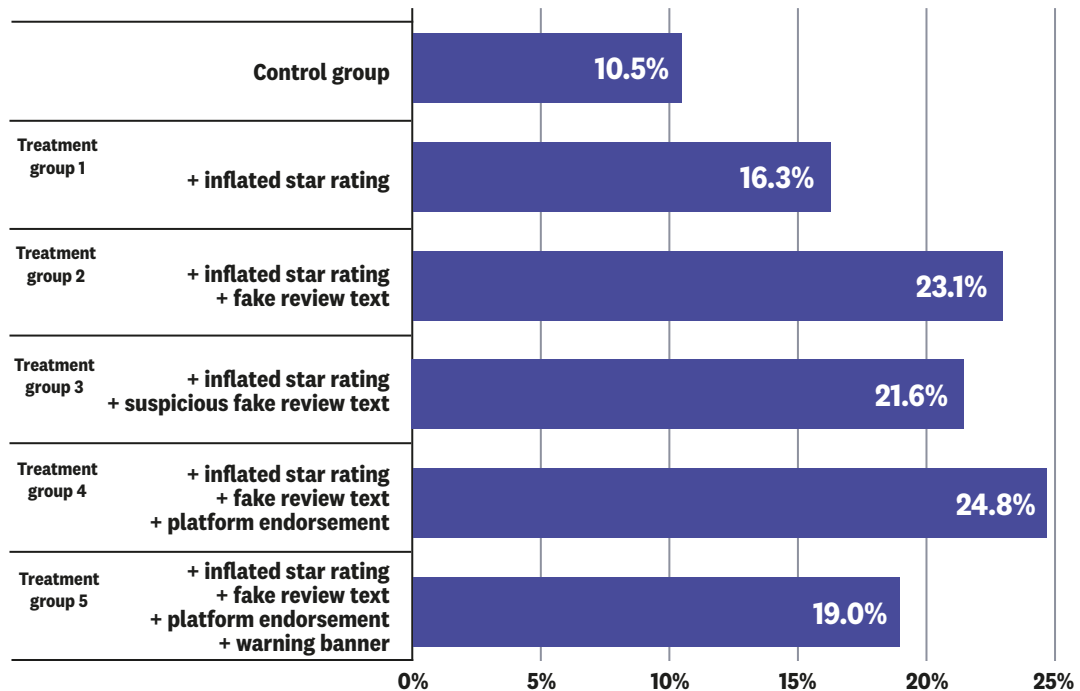
As shown in Figure 8, adding the **inflated star rating** treatment led directly to an additional 5.8pp of consumers choosing a Don't Buy product, an increase of 55% over the control group. This demonstrates that simply manipulating reviews to include a larger number of five star reviews was highly effective at driving consumers towards a low quality product, even while the text of the reviews left for those products remained largely negative. Such an effect is consistent with the literature on reviews more generally, which has found consumers use star ratings as a heuristic to judge quality and consensus of other consumers' opinions.

In combination with the inflated star rating, adding **fake review text** to the Don't Buy products increased the proportion of consumers choosing them even further. 23.1% of people chose the Don't Buy product when both fake star ratings and fake review text were added to the products, which equates to an increase of 120% in the number of people choosing the Don't Buy relative to the control group. It was also a substantial increase (+6.8pp) over the simple manipulation of the star ratings. This indicates that the production of a small number of affect-rich fake reviews can, if prominently displayed, have a large impact on the likelihood of consumers purchasing an item. Given the ease and low cost of faking small numbers of reviews, this indicates the potential harm from fake reviews could be very substantial.

Similarly sized effects were observed even when the suspiciousness of the fake review text was increased. 21.6% of participants chose the Don't Buy in our **suspicious fake review text** treatment group, only 1.5pp fewer than the previous treatment group where obvious signs of manipulation were not included, showing that fake reviews still have a large impact on consumer behaviour even when manipulation is not particularly sophisticated or well disguised.

Adding the **platform endorsement** to the inflated star ratings and (non-suspicious) fake review text further increases the proportion of people choosing the Don't Buy product though only by around 1.7 percentage points. This means that around one quarter (24.8%) of our participants in the experiment were choosing the Don't Buy product in this treatment group in comparison to only around one in ten (10.5%) in the control group.

Finally, while we were able to find large effects on behaviour from adding the fake reviews to the Don't Buy products, we also found that our simple **warning banner** intervention was reasonably effective at reducing some of this impact. However, the intervention was far from sufficient to remove the harm completely. The results show the banner had a significant impact in reducing the proportion of consumers choosing the poor product, falling from 24.8% in treatment group 4 (inflated star rating, fake review text, platform endorsement) to 19% when the banner was introduced for treatment group 5. This is a fall of 20% in the proportion choosing the Don't Buy but still represents an 81% increase over the control group.

Other findings

In addition to our main findings on the proportion of consumers choosing the Don't Buy, the survey questions attached to the choice task also allowed us to analyse other outcomes and to check whether the treatments affected the choices of some groups of consumers more than others.

Do effects differ for different groups of consumers?

Generally, the results did not vary significantly across various demographics like age, income and education level. This is indicative of fake reviews having a broad and relatively stable effect across different groups of consumers, however it is important to note that statistically significant differences are typically more difficult to detect as we drill into smaller groups and sample sizes fall. Therefore while we found few statistically significant differences, this does not necessarily mean such differences do not exist.

Beyond the standard demographic differences however, we did find that those who use Amazon more frequently (and thus most familiar with the set-up of the shopping task) were more likely to be influenced by the fake reviews. Those using Amazon more than once or twice per month were consistently 5-10pp more likely to choose the Don't Buy product when fake reviews were applied. While we cannot determine whether using an online shopping platform more frequently causes someone to be more susceptible to fake reviews, the result is interesting as a demonstration that those who we might expect to be more experienced and savvy online shoppers were less able to spot the poor quality product in the presence of fake reviews.

We also found that the effects of fake reviews were greater among those who spent longer on the experiment, apart from when only the fake stars were added. This may indicate that those consumers who spent more time actually reading the review content were more likely to be influenced by the positive comments without being able to spot the signs of fakery. This result may be important when designing remedies to help consumers avoid the effects of fake reviews, as those aimed at getting consumers to spend more time to detect reviews may be ineffective.

Results by product type

The results were mostly consistent among all three product types that participants could view. Generally, the inflated star rating led to increased choice of the Don't Buy product, with further increases for the faked review text and platform endorsement . Across all three products, increasing the suspiciousness of the fake review reduced their impact only slightly and the warning banner was only somewhat effective in reducing the proportion choosing the Don't

Buy. Across the headphones and cordless vacuum cleaner categories all of these results remain statistically significant but, given the smaller sample sizes, the dash cam results were directionally similar but generally not statistically significant other than the *inflated star rating + fake review text + platform endorsement* treatment group.

While the general structure of the results was similar, interesting differences can be observed between the product categories, particularly the headphones and cordless vacuum cleaner results where all the results retain their statistical power in isolation. The different product categories were primarily chosen to cover a range of prices, given that the literature suggests that consumers use reviews differently when products become more expensive.

In our setting, if used as a heuristic, we may expect the inflated star ratings to have a larger effect on the lower priced items (headphones, £25) and the review content possibly being more influential for the higher priced items (cordless vacuum cleaners, £150). Conversely, if consumers engage their systematic information processing for the higher priced items, they may be likely to spot the suspicious elements in the fake reviews and adjust their behaviour accordingly.

In actuality, we found that the fake reviews were actually most effective at changing consumer behaviour in our highest priced category, cordless vacuum cleaners. 9.5% of consumers chose the Don't Buy cordless vacuum cleaner in the control group, more than trebling to 29.5% once the inflated star ratings, fake text and platform endorsement treatments were added. While still strong, these effects were not as dramatic in the dash cams (15.6% to 23.64%) and headphones (9.1% to 20.5%) categories. These results across product types for Treatment Group 4 are shown in Figure 9 below but results for all treatment groups please see Annex 3.

Figure 9 – proportion choosing the Don't Buy product in the Control Group and Treatment Group 4 for each product category
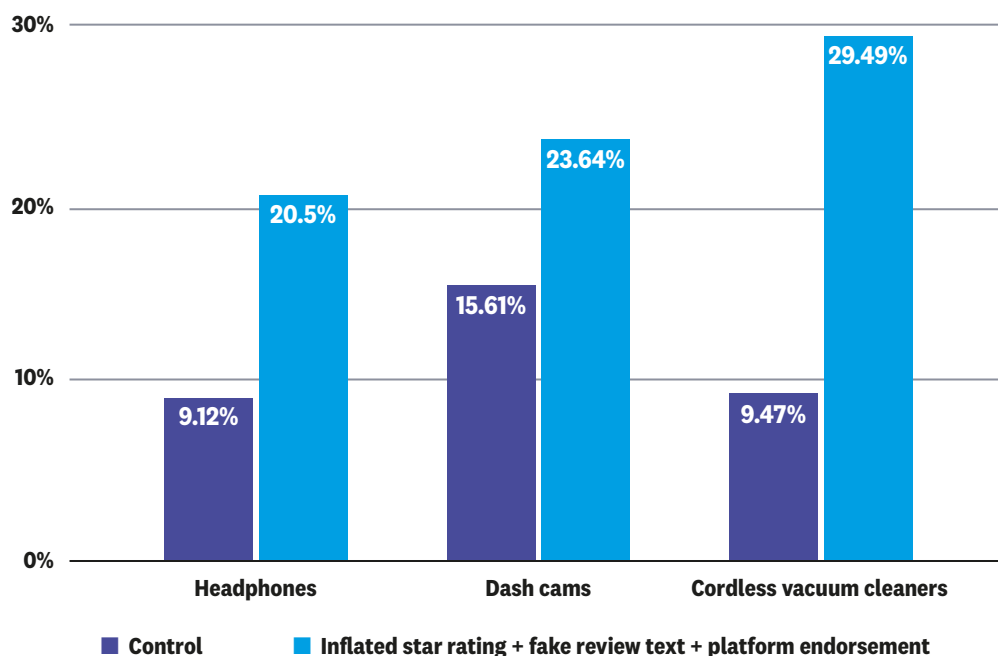
It is unclear what precisely is behind the differences in effect size between the product categories. There was, of course, variation in the information presented beyond simply price, for example branding and technical characteristics both varied across products.

However, the results show that effects are present across a range of products at different prices, levels of branding and technical specifications. The results also show that levels of harm resulting

from fake reviews could vary significantly depending on the product category and products displayed on search pages. As many as one in five of the participants in our experiment in treatment group 4 would have made a better choice of vacuum cleaner had they not been exposed to fake reviews and seller endorsement. Indeed this effect was similar (+17pp) even when clear indications of manipulation were present in the text of the reviews, demonstrating significant potential harm from fake reviews when deployed on the most susceptible products.

Proportion buying the Best Buy and filler products

Finally, we also examine which products consumers were less likely to choose when fake reviews encouraged them to choose the Don't Buy. The potential for harm would be particularly high if consumers who would have chosen a Best Buy could be misled into purchasing a Don't Buy through the inclusion of fake reviews.

The results, see Figure 10, show that some of the increase in Don't Buy purchases comes at the expense of Best Buy purchases. In Treatment Group 3 for instance, where the star rating and the text of reviews was manipulated, we saw a 12.6 percentage point increase in the proportion of consumers choosing the Don't Buy. This corresponded to a fall of 5.2 percentage points in participants choosing the Best Buy and 7.4 percentage points fewer choosing any of the three filler products.

Figure 10 - proportion choosing the Don't Buy product in the Control Group and Treatment Group 3, and where choices switched from
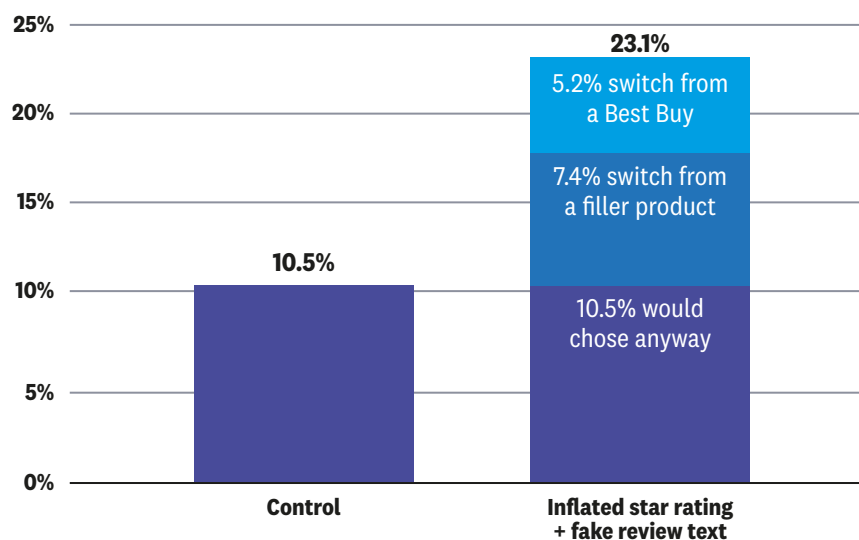
This finding indicates that a significant minority of consumers might switch their preferences away from the best choice available to the worst choice, experiencing greater harm from being misled by the reviews. The reduction in Best Buy choice was highest in Treatment Group 3 but was around 4-5 percentage points in every one of the treatment groups.

Discussion and summary

This study provided us with strong evidence that fake reviews influence consumer choices and create considerable harm. When no fake reviews were added to products, only around one in ten consumers chose a Don't Buy product among a choice of five products. When we added fake reviews to this Don't Buy, around one in four consumers chose it as their preferred option. This demonstrates that fake reviews are effective and can lead a significant number of consumers to make unequivocally poor choices. Furthermore, we found that the level of deception employed in fake reviews does not need to be sophisticated to create harm.

Our results provide an important addition to the existing research on fake reviews and reviews more generally. Other researchers have identified the presence of fake reviews, but findings on how they affect consumers are few and inconclusive. Our results are unambiguous and provide clear evidence of the significant effect that fake reviews can have.

As with any survey-based experiment, it is hard to be certain whether the results would be replicated in a real-life setting, where consumers are making actual purchases. However, we believe that it is highly likely that we would see similar effects in a real online retailing environment. The experiment was designed to closely mirror the design of the real Amazon. co.uk retail platform, including product information and much of the real review content directly from actual webpages, although the fake review content was created by us. Furthermore, by giving consumers a chance to win the product that they chose, we gave consumers an incentive to take the experiment seriously and act in a similar way to real life.

While our experiment focused quite narrowly on a single retail platform and on three particular products, we know that fake reviews are present on many different platforms on a wide variety of products and services. We expect that our main finding that fake reviews influence consumer choices would be similar across review platforms. However, we acknowledge that the specifics of the results of this study cannot be automatically generalised to these different platforms particularly those focusing on services like holidays, hotels or restaurants. Regulators and review platforms may be interested to undertake further study to see whether our results could be replicated in these different settings.

The study also adds to the small literature looking at remedies that might help consumers improve their ability to filter out deception from the reviews. We tested one simple banner added to the top of webpages and found that this could reduce the effectiveness of fake reviews. This was however far from sufficient to remove the impact entirely, and meaningful action by platforms to reduce the incidence of fake reviews will always be necessary. In addition, we have no evidence as to whether the effectiveness of such a remedy would increase or decrease over time, as consumers become more used to its messaging, and so the onus must pass to regulators and platforms themselves to test a range of consumer-focused interventions that could be introduced in order to reduce consumer harm. With access to data on real purchases on their own websites, online platforms are in the strongest position to test any potential remedies to reduce the impact of fake reviews on their own platforms. Regulators might want to consider making this mandatory in their future examinations of the topic of fake reviews.

This research serves as an important evidence base for regulators and all review platforms to acknowledge and find effective ways to reduce the impact of fake reviews on their customers. It demonstrates how large the effect on consumer behaviour can be and as such it is paramount that regulators and review platforms, such as Amazon, eBay and TripAdvisor take responsibility to make improvements to systems to remove fake reviews and stop consumers being misled.

Which? is concerned that review platforms are not doing enough to tackle the system wide problem of fake reviews. Amazon has policies in place to prevent review manipulation, including allowing users to report abuse. However our investigations continue to find examples of suspicious reviews across major websites. Therefore we welcome the regulator's investigation into whether major websites are doing enough to ensure that the measures in place are effective in ensuring that their customers are not being misled. People should be able to trust the reviews they read on these sites, without risking unwittingly purchasing a poor quality or even unsafe product. Platforms that host reviews should, as a minimum, uphold the following principles:

- Platforms should have effective and adaptive policies and processes in place for preventing and removing fake reviews.
- They should take swift and effective enforcement action against those that breach these policies, including rogue sellers, businesses that incentivise illegitimate reviews and those offering 'black hat' services that offer to manipulate platforms' policies and processes in order to inflate product ratings and reviews.
- Platforms should only accept reviews from customers who have made genuine transactions. Equally, sites should be transparent about reviews that have been incentivised and such reviews should not be biased but reflect genuine consumer experiences.
- Only real reviews for genuine transactions should be used in determining rankings and endorsements awarded by sites. Sites should be clear about how such rankings are determined.
- Review hosting sites should be transparent about their business models, including the services offered to firms that pay to use their sites, and how they display reviews.

Given the strong evidence of harm identified in our research, we welcome the CMA's investigation into how fake reviews are being used to manipulate online shoppers on major websites, and expect the regulator to take the strongest possible action against sites that fail to tackle this problem.

References

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

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

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

Clemons, E.K., Gao, G.G. and Hitt, L.M., (2006), *When online reviews meet hyperdifferentiation: A study of the craft beer industry.* Journal of management information systems, 23(2).

Floyd, K., Freling, R.,Alhoqail, S., Cho, H.Y. and Freling, T., (2014), *How online product reviews affect retail sales: A meta-analysis.* Journal of Retailing, 90(2).

Gopinath, S., Thomas, J.S. and Krishnamurthi, L., (2014), *Investigating the relationship between the content of online word of mouth, advertising, and brand performance.* Marketing Science, 33(2).

Hu, N., Bose, I., Koh, N.S. and Liu, L., (2012), *Manipulation of online reviews: An analysis of ratings, readability, and sentiments.* Decision support systems, 52(3).

Hu, N., Liu, L. and Sambamurthy, V, (2011) *Fraud detection in online consumer reviews.* Decision Support Systems, 50(3).

Imrg.org. (2016). *Over Half Of Online Sales Now Made Through Mobile Devices.* [online] Available at: https://www.imrg.org/media-and-comment/press-releases/over-half-of-online-sales-now-made-through-mobile-devices

Kostyra, D., Reiner, J., Natter, M. and Klapper, D., (2016), *Decomposing the effects of online reviews on brand, price and product attributes.* International Journal of Research in Marketing, 33.

Malthouse, E.C, Maslowska,E. and Viswanathan, V., (2017), *Do customer reviews drive purchase decisions? The moderating roles of review exposure and price*, Decision Support Systems, 98.

Munzel, A (2016), *Assisting consumers in detecting fake reviews: The role of identity information disclosure and consensus*, Journal of Retailing and Consumer Services 32.

Sun, M (2012), *How does the variance of product ratings matter?*, Management Science, 58(4).

von Helversen, B., Abramczuk, K., Kopeć, W. and Nielek, R., (2018) *Influence of consumer reviews on online purchasing decisions in older and younger adults.* Decision Support Systems, 113.

Zhang, K.Z.K, Zhao, S.J., Cheung, C.M.K and Lee, M.K.O, (2014), *Examining the influence of online reviews on consumers' decision making: A heuristic-systematic model*, Decision Support Systems, 67.

Annex 1
Sample demographics and experimental balance

	Control group	Treatment group 1	2	3	4	5	p-value
Gender							
Female	50%	52%	52%	51%	53%	50%	0.35
Male	50%	48%	48%	49%	47%	50%	0.34
Region							
East Midlands	7%	7%	8%	8%	7%	7%	0.95
East of England	10%	8%	10%	9%	9%	10%	0.64
Greater London	13%	14%	13%	13%	13%	11%	0.40
North East England	4%	4%	4%	4%	4%	5%	0.85
North West England	12%	12%	11%	11%	12%	11%	0.62
Northern Ireland	3%	3%	2%	3%	3%	3%	0.81
Scotland	8%	8%	8%	8%	7%	8%	0.85
South East England	13%	13%	15%	14%	13%	15%	0.76
South West England	10%	9%	8%	9%	9%	8%	0.36
Wales	5%	4%	5%	5%	5%	6%	0.19
West Midlands	7%	9%	9%	10%	9%	9%	0.16
Yorkshire and the Humber	7%	9%	8%	9%	9%	9%	0.61
Age							
18 to 24	12%	12%	12%	12%	12%	13%	0.98
25 to 34	16%	17%	17%	19%	17%	17%	0.37
35 to 44	18%	18%	18%	18%	18%	18%	0.99
45 to 54	18%	18%	19%	18%	18%	17%	0.82
54 to 65	15%	16%	16%	14%	15%	16%	0.56
65+	21%	19%	20%	20%	21%	20%	0.75
Income							
Up to 21k	36%	37%	35%	37%	40%	37%	0.10
Between 21k and 41k	34%	34%	33%	35%	33%	35%	0.81
Between 41k and 69k	15%	14%	16%	13%	14%	14%	0.06
69k and above	6%	7%	7%	6%	6%	6%	0.56
Education							
Primary school	1%	0%	1%	1%	1%	1%	0.21
Secondary school up to 16 years	21%	21%	21%	20%	20%	20%	0.83
Higher or secondary education	25%	25%	23%	29%	25%	26%	0.01
College or university	38%	38%	40%	37%	40%	38%	0.46
Post-graduate degree	14%	14%	14%	13%	13%	14%	0.56

	Control group	Treatment group					p-value
		1	2	3	4	5	
Daily internet usage							
Less than 2 hours	25%	24%	25%	25%	24%	24%	0.95
Between 2 and 5 hours	51%	52%	51%	52%	50%	52%	0.90
More than 5 hours	23%	22%	23%	22%	24%	22%	0.80
Frequency of purchases online							
At least once or twice per month	78%	77%	79%	77%	76%	76%	0.20
Less than once or twice per month	21%	22%	20%	22%	23%	23%	0.21
Frequency of purchases on Amazon.co.uk							
At least once or twice per month	58%	59%	60%	57%	58%	60%	0.35
Less than once or twice per month	41%	40%	39%	41%	40%	39%	0.52
Product chosen							
Headphones	40%	40%	41%	40%	40%	40%	0.99
Dash cams	19%	19%	19%	20%	20%	19%	0.94
Cordless vacuum cleaners	41%	41%	41%	40%	41%	41%	0.98
Device used							
Desktop	50%	50%	49%	50%	49%	50%	0.95
Mobile	50%	50%	52%	50%	51%	50%	0.95
Observations	1648	1667	1647	1671	1693	1662	

Annex 2

Main results table - effects on product choice

Table A2.1. Results of regression

Product choice = constant + β_1*TG1 + β_2*TG2 + β_3*TG3 + β_4*TG4 + β_5*TG5 + β

	Chose Don't Buy	Chose Best Buy	Chose 'filler'
Constant (Control group)	10.50***	28.28***	61.23***
Treatment Group 1: inflated star ratings	5.76***	-1.64	-4.12**
Treatment Group 2: Inflated star ratings and fake review text	12.57***	-5.20***	-7.37***
Treatment Group 3: Inflated star ratings and suspicious fake review text	11.05***	-4.34***	-6.71***
Treatment Group 4: Inflated star ratings and fake review text and platform endorsement	14.25***	-3.65**	-10.61***
Treatment Group 5: Inflated star ratings and fake review text and platform endorsement	8.46***	-4.39***	-4.07**

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

Annex 3

Main results by product type

Table A3.1: Headphones only results of regression

Product choice = constant + β_1*TG1 + β_2*TG2 + β_3*TG3 + β_4*TG4 + β_5*TG5 + β

	Chose Don't Buy	Chose Best Buy	Chose 'filler'
Constant (Control group)	9.12***	41.64***	49.24***
Treatment Group 1: inflated star ratings	3.91**	-3.92	0.01
Treatment Group 2: Inflated star ratings and fake review text	10.82***	-3.69	-7.13***
Treatment Group 3: Inflated star ratings and suspicious fake review text	8.74***	-5.18*	-3.56
Treatment Group 4: Inflated star ratings and fake review text and platform endorsement	11.38***	-0.79	-10.60***
Treatment Group 5: Inflated star ratings and fake review text and platform endorsement	7.85***	-0.50	-7.35***
Observations	4014	4014	4014

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

Table A3.2: Dash cams only results of regression

Product choice = constant + ß1*TG1 + ß2*TG2 + ß3*TG3 + ß4*TG4 + ß5*TG5 + ß

	Chose Don't Buy	Chose Best Buy	Chose 'filler'
Constant (Control group)	15.61***	22.29***	62.10***
Treatment Group 1: inflated star ratings	-0.46	3.26	-2.80
Treatment Group 2: Inflated star ratings and fake review text	4.39	-0.98	-3.41
Treatment Group 3: Inflated star ratings and suspicious fake review text	2.77	1.80	-4.57
Treatment Group 4: Inflated star ratings and fake review text and platform endorsement	8.03***	-0.78	-7.25*
Treatment Group 5: Inflated star ratings and fake review text and platform endorsement	2.02	-0.82	-1.20
Observations	1910	1910	1910

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

Table A3.3: Cordless vacuum cleaners only results of regression

Product choice = constant + $ß_1$*TG1 + $ß_2$*TG2 + $ß_3$*TG3 + $ß_4$*TG4 + $ß_5$*TG5 + ß

	Chose Don't Buy	Chose Best Buy	Chose 'filler'
Constant (Control group)	9.47***	18.05***	72.49***
Treatment Group 1: inflated star ratings	10.47***	-1.77	-8.70***
Treatment Group 2: Inflated star ratings and fake review text	18.14***	-9.09***	-9.05***
Treatment Group 3: Inflated star ratings and suspicious fake review text	17.37***	-6.80***	-10.57***
Treatment Group 4: Inflated star ratings and fake review text and platform endorsement	20.02***	-7.97***	-12.05***
Treatment Group 5: Inflated star ratings and fake review text and platform endorsement	12.02***	-9.86***	-2.16
Observations	4064	4064	4064

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