INVESTIGATION · MARKET AND CONSUMER INVESTIGATION

Investigating Online Reviews and Fake Feedback

Online reviews influence what people buy, which makes them a target for manipulation. This article explains how fake reviews are produced, what research shows about their scale and effects, the signals investigators use to detect them, and the laws now aimed at stopping them.

ModQu research article · 1,809 words · 18 references (APA 7th edition)

Before choosing a restaurant, a hotel, a plumber or a pair of headphones, many people read what other customers have said. Online reviews have become one of the most important sources of information in consumer markets, and research has shown that they affect sales. Chevalier and Mayzlin (2006) found that improvements in a book's reviews on one online retailer led to higher relative sales on that site, and Luca (2016) found that a one-star increase in a restaurant's Yelp rating was associated with a rise in revenue of several per cent for independent restaurants.

Because reviews matter, some businesses try to manipulate them: by writing positive reviews of their own products, paying others to do so, posting negative reviews of competitors or suppressing genuine criticism. Investigating reviews means distinguishing genuine customer experience from manufactured feedback, a task that platforms, regulators, competitors and researchers all now undertake.

How fake reviews are produced

Fake reviews come from several sources. Business owners and staff may post reviews themselves. Marketing agencies and online brokers sell batches of reviews, often written by networks of people paid per review. Some sellers offer free products or refunds in exchange for five-star reviews, a practice that blurs the line between genuine and fake. Competitors may post negative reviews to damage a rival. And automated tools, including text-generation software, can now produce plausible reviews in large numbers.

He et al. (2022) studied the market for fake reviews on a large online retailer by monitoring private social media groups in which sellers recruited reviewers. They found that sellers who bought fake reviews saw a short-term rise in ratings and sales, followed by a decline in ratings once the purchases stopped, and that the products involved tended to be of lower quality. This suggests that fake reviews mislead consumers about quality, not just about popularity.

What research shows about manipulation

Researchers have used natural experiments to estimate how common manipulation is. Mayzlin et al. (2014) compared hotel reviews on a site where anyone could post a review with reviews on a site that accepted reviews only from people who had booked through it. Independent hotels, which had stronger incentives and fewer constraints, had more five-star reviews on the open site relative to the verified site, and hotels near independent competitors had more one-star reviews, consistent with both self-promotion and attacks on rivals.

Luca and Zervas (2016) used reviews that Yelp's filtering system had flagged as suspicious and found that about 16 per cent of restaurant reviews were filtered. Restaurants were more likely to commit review fraud when their reputation was weak or recently declining, and independent restaurants were more likely to do so than chain restaurants.

Even genuine reviews can mislead. Hu et al. (2009) observed that the distribution of product ratings is often J-shaped, with many five-star ratings, some one-star ratings and few in between, because customers with strong views are more likely to write reviews. Anderson and Simester (2014) found that a notable share of reviews at a clothing retailer were written by customers who had not bought the product they reviewed, and that these reviews were more negative than average.

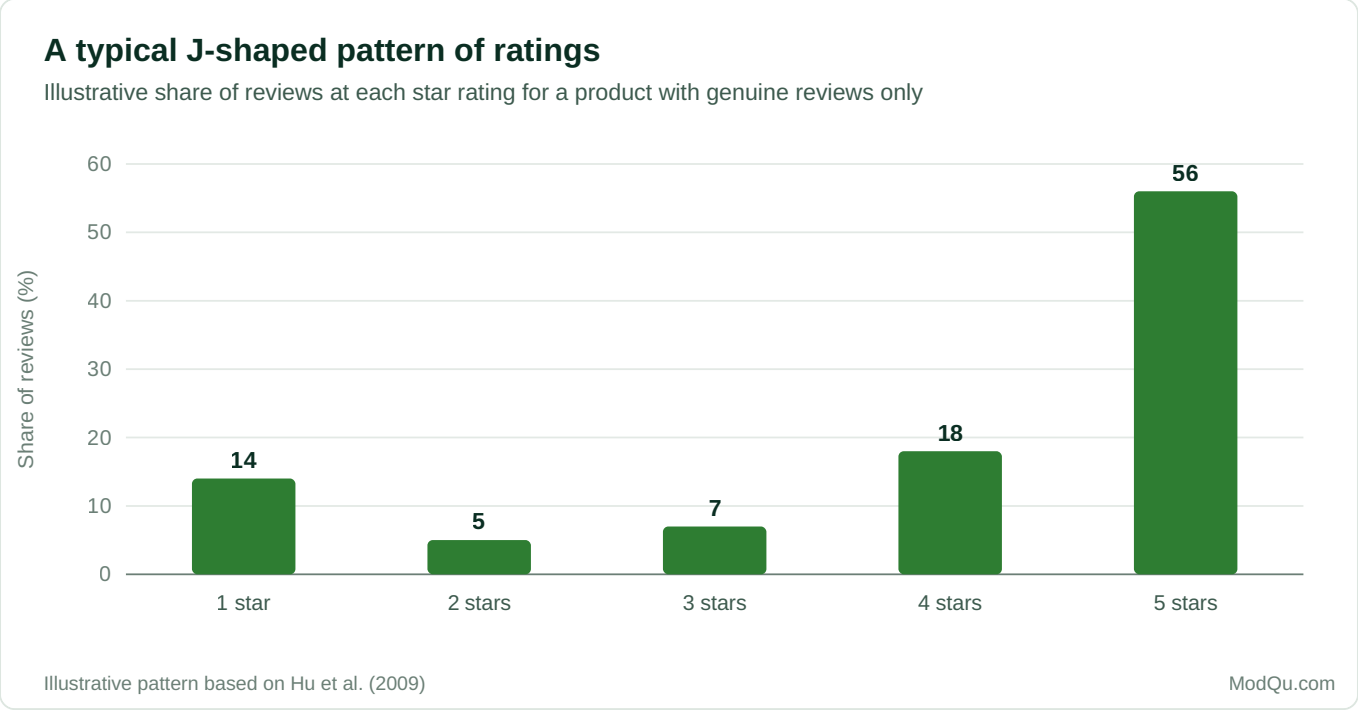


Figure 1. Because people with strong views are most likely to write reviews, genuine ratings often cluster at the extremes. An average rating hides this pattern, and a sudden change in its shape can be a sign of manipulation. Diagram by ModQu.

Signals of fake reviews

No single sign proves that a review is fake. Investigators look for combinations of signals at the level of the review text, the reviewer, the timing and the network of reviewers and products. Reviews of research on detection describe the main approaches (Wu et al., 2020).

Signals investigators use to detect fake reviews

Each is weak alone; combined, they can identify suspicious patterns

	Examples	Caution
Text	Vague praise, few specific details, repeated phrases across reviews	Genuine reviews can also be short or generic
Reviewer	New account, few reviews, all five-star, reviews across unrelated products	New customers are often new reviewers
Timing	Bursts of reviews in a short period, especially after poor ones	Launches and promotions also create bursts
Purchase	No verified purchase; reviewer received free product	Not all platforms record purchases
Network	Same reviewers across a group of sellers; shared devices or addresses	Requires platform data to establish

Based on Jindal & Liu (2008), Mukherjee et al. (2012) and Wu et al. (2020) ModQu.com

Figure 2. Fake reviews are rarely identified by one feature. Investigators look for consistent patterns across text, reviewer behaviour, timing and networks. Diagram by ModQu.

Text

Early research found that deceptive reviews differ from genuine ones in their language. Ott et al. (2011) created a set of deceptive hotel reviews by paying people to write fake positive reviews, and found that a machine learning classifier could distinguish them from genuine reviews far more accurately than human judges, who performed little better than chance. Fake reviews tended to contain fewer concrete details about the hotel's physical features and more references to the reviewer and to external context. These results are consistent with broader research showing that people are poor at detecting deception from content (Bond & DePaulo, 2006) and that cues to deception, while real, are weak and unreliable on their own (DePaulo et al., 2003). Text-based detection has become harder as text-generation tools have improved, since generated reviews can mimic the features of genuine ones (Salminen et al., 2022).

Reviewers and networks

Behavioural signals are often stronger than text. Jindal and Liu (2008) identified duplicate and near-duplicate reviews as an early sign of opinion spam. Mukherjee et al. (2012) showed that groups of reviewers who review the same products at around the same time can be identified, and that such groups are a strong indicator of coordinated manipulation. Platforms have access to data that outsiders lack, such as account creation dates, device identifiers, payment records and connections between accounts, which make network analysis much more powerful.

Timing and distribution

Sudden bursts of positive reviews, particularly following negative ones, and changes in the shape of the rating distribution can indicate manipulation. Comparing a product's ratings on platforms with different verification rules, in the manner of Mayzlin et al. (2014), can also reveal anomalies.

Beyond outright fakes

Not all manipulation involves invented reviews. Several practices distort the picture while using real customers.

Incentivised reviews. Offering customers a discount, free product or entry into a prize draw in exchange for a review tends to produce more positive reviews, especially if the reward appears to depend on the rating. Even when the incentive is disclosed, readers may not take it into account.

Review gating. Some businesses ask customers whether they were satisfied before inviting them to post a public review, and send only satisfied customers to the review site. Each review may be genuine, but the selection makes the overall rating misleading.

Suppression. Businesses may discourage or remove negative reviews, for example through contract terms that forbid criticism, threats of legal action, or platform features that allow sellers to hide or contest unfavourable reviews.

Review hijacking. On some marketplaces, sellers attach a new product to the listing of an older, well-reviewed product, so that the new product inherits ratings it never earned.

Investigators therefore ask not only whether individual reviews are real, but whether the set of reviews shown is a fair sample of customers' experience. The question is the same one that survey researchers ask about any sample: who is included, who is left out, and how does that selection affect the result?

What platforms do

Platforms are in the best position to investigate, because they hold the data that link reviews to accounts, purchases, devices and payments. Most large platforms use automated systems to filter suspicious reviews before or after publication, combined with human investigators who examine networks of accounts and sellers. Some restrict reviews to verified purchasers, label incentivised reviews, or show the distribution of ratings rather than only the average. Platforms also take legal action against review brokers and share information with regulators. Their incentives are mixed, however:

removing reviews can reduce the apparent activity and positivity of a marketplace, and filtering errors can remove genuine reviews and frustrate customers. Regulators increasingly expect platforms to explain what checks they carry out and to show that those checks are effective.

Investigating as an outsider

Consumers, journalists and competitors who suspect fake reviews do not have access to platform data, but they can still investigate. Useful steps include reading the full distribution of reviews rather than the average; reading the most recent critical reviews, which often describe problems that positive reviews ignore; examining reviewer profiles for history and diversity of reviews; looking for repeated phrases across reviews; and comparing ratings for the same business or product across several sites. Journalists have exposed review-selling networks by joining the groups where reviews are traded and documenting the transactions.

Outsiders should be cautious about conclusions. A burst of positive reviews may reflect a successful launch, and a new reviewer may be a genuine first-time customer. Allegations that a business is buying reviews can be defamatory if they are not well supported.

The legal response

Governments have begun to treat fake reviews explicitly as unlawful. The European Union amended its consumer protection rules to prohibit submitting or commissioning false consumer reviews and to require traders who publish reviews to state whether and how they check that reviews come from genuine customers (Directive (EU) 2019/2161, 2019). In the United Kingdom, the Digital Markets, Competition and Consumers Act 2024 lists commissioning, publishing or facilitating fake reviews among practices that are always unfair. In the United States, the Federal Trade Commission brought its first case against a company for paying for fake reviews on a large retail website in 2019 (Federal Trade Commission [FTC], 2019), and in 2024 adopted a rule prohibiting fake reviews and testimonials, including those generated by artificial intelligence, and the suppression of negative reviews (FTC, 2024).

These laws give regulators clearer grounds for investigation, and they place duties on platforms and businesses to take reasonable steps to prevent and remove fake reviews.

A short example

The following hypothetical case illustrates an outsider's investigation. A consumer organisation notices that several brands of wireless earbuds sold online have average ratings above 4.7 stars, despite complaints about poor battery life in its own testing. Researchers sample the reviews and find that many were posted within a few days of each other, that a large share of reviewers had reviewed products from the same group of sellers, and that some reviews mention a different product altogether, suggesting that they had been merged from an earlier listing. Researchers join social media groups advertising free products in exchange for reviews and find posts offering refunds for the same earbuds. The organisation publishes its findings, reports the sellers to the platform and the regulator, and advises consumers on how to read reviews critically. The platform removes the reviews and the listings.

Conclusion

Online reviews are valuable because they aggregate the experience of many customers, and that value is what makes them a target. Investigating reviews requires combining several kinds of evidence, since text alone rarely distinguishes fake from genuine feedback, while patterns in reviewer behaviour, timing and networks are often revealing. Research shows that manipulation is common enough to affect consumer decisions, and new laws give regulators stronger tools to act. For businesses, the lesson is that genuine reviews, including critical ones, are a source of information about how customers experience their products; for consumers, it is to read reviews as evidence to be weighed, not as verdicts.

References

Anderson, E. T., & Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research, 51*(3), 249–269. https://doi.org/10.1509/jmr.13.0209

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Digital Markets, Competition and Consumers Act 2024, c. 13 (UK).

Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules. (2019). *Official Journal of the European Union, L 328*, 7–28.

Federal Trade Commission. (2019, February 26). *FTC brings first case challenging fake paid reviews on an independent retail website* [Press release].

Federal Trade Commission. (2024). *Trade regulation rule on the use of consumer reviews and testimonials* (16 C.F.R. Part 465).

He, S., Hollenbeck, B., & Proserpio, D. (2022). The market for fake reviews. *Marketing Science, 41*(5), 896–921. https://doi.org/10.1287/mksc.2022.1353

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM, 52*(10), 144–147. https://doi.org/10.1145/1562764.1562800

Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 219–230). ACM. https://doi.org/10.1145/1341531.1341560

Luca, M. (2016). *Reviews, reputation, and revenue: The case of Yelp.com* (Harvard Business School Working Paper No. 12-016).

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science, 62*(12), 3412–3427. https://doi.org/10.1287/mnsc.2015.2304

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review, 104*(8), 2421–2455. https://doi.org/10.1257/aer.104.8.2421

Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 191–200). ACM. https://doi.org/10.1145/2187836.2187863

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 309–319). Association for Computational Linguistics.

Salminen, J., Kandpal, C., Kamel, A. M., Jung, S.-g., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services, 64*, Article 102771. https://doi.org/10.1016/j.jretconser.2021.102771

Wu, Y., Ngai, E. W. T., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems, 132*, Article 113280. https://doi.org/10.1016/j.dss.2020.113280