Online Reviews: Information Content, Drivers, and Platform Design

Tommaso Bondi, Michelangelo Rossi*

January 22, 2026

Abstract

Online ratings emerge from a multi-stage process that can systematically distort their informational content. We develop a unified framework decomposing the rating process into distinct components: experienced quality (driven by intrinsic quality, seller effort, and price), expectations formed prior to consumption, contextual influences, strategic distortions, idiosyncratic tastes, and selection into reviewing. This decomposition organizes a growing theoretical and empirical literature and clarifies how seemingly disparate findings – from fake reviews to disappointment effects to selection biases – relate to distinct stages of the data-generating process. Our framework also provides a lens for evaluating platform design interventions: effective policies target specific components of the rating process, yet many distortions remain difficult to address without introducing new trade-offs. We highlight open questions where further research is most needed.

Keywords: Online reviews, rating biases, digital platforms, platform design

*T. Bondi: Cornell Tech and SC Johnson School of Business, and CESifo. E-mail: tbondi@cornell.edu. M. Rossi: HEC Paris, and CESifo. E-mail: rossim@hec.fr.

1 Introduction

Online consumer reviews are a defining feature of modern digital marketplaces. They help buyers navigate vast product variety, reduce search frictions, and discipline sellers through reputation mechanisms. The rapid growth of online ratings has generated a large and interdisciplinary literature examining who writes reviews, how consumers interpret them, and how reviews ultimately influence demand and welfare.

Yet despite the proliferation of studies, the literature remains somewhat fragmented. Researchers in marketing, economics, information systems, and computer science have documented a wide range of phenomena – from fake reviews to herding behavior to selection biases – often using different frameworks and terminology. This fragmentation makes it difficult to assess how different findings relate to one another or to identify where the most important gaps lie.

Prior surveys (Tadelis, 2016; Pocchiari et al., 2024) have provided comprehensive coverage of these themes. Our objective is complementary but distinct: we synthesize recent research on the *drivers* of online ratings – the mechanisms that shape their informational content – and organize findings through a unified conceptual framework. Unlike earlier surveys that typically organize findings by topic or application domain, we develop a simple model of how ratings are generated. This approach helps integrate diverse empirical and theoretical regularities documented in the literature and clarifies when ratings deviate systematically from underlying quality.

Three features distinguish our approach. First, we anchor the review in a unified conceptual framework that explicitly models the rating as the outcome of a multi-stage process, making precise where different biases enter. Second, we focus on drivers and distortions – the forces that cause ratings to diverge from the quality signal a social planner might prefer – rather than on downstream effects of ratings on demand or welfare. Third, we use this framework to evaluate platform design interventions, asking which components of the rating process they target and whether they successfully restore informativeness.

2 Conceptual Framework

The rating ultimately observed by a platform is the outcome of a sequence of stages that unfold during and after consumption. We formalize these stages to organize the drivers discussed in subsequent sections.

Consumption: Experienced Quality. During the transaction, the consumer experiences realized quality:

$$Q_{ij} = q(Q_i, e_{ij}, p_{ij}), \tag{1}$$

where Q_i denotes intrinsic product quality, e_{ij} is seller effort in transaction j, and p_{ij} is the price paid. The function $q(\cdot)$ represents the production of consumption utility from these inputs. Intrinsic quality Q_i captures fixed product characteristics, while seller effort e_{ij} allows for transaction-specific variation in service or fulfillment quality. The price p_{ij} enters because consumers often evaluate experiences relative to what they paid.

Post-consumption: Internal Evaluation. After consumption, the consumer privately evaluates the experience:

$$r^*_{ij} = f\big(Q_{ij}, \mathbb{E}_{ij}[Q_{ij}], X_{ij}, B_{ij}, \theta_{ij}\big). \tag{2}$$

Here $\mathbb{E}_{ij}[Q_{ij}]$ captures expectations formed prior to consumption (e.g., through ratings, badges, or prices) and influences the evaluation through mechanisms such as disappointment or positive surprise. The term X_{ij} represents contextual and situational factors (weather, mood, social influence, reviewer identity); B_{ij} captures strategic distortions (e.g., incentives, reciprocity concerns, retaliation, managerial responses); and θ_{ij} denotes idiosyncratic tastes or reviewer stringency. The function $f(\cdot)$ maps these inputs into a scalar evaluation.

Selection into Reviewing. The platform observes the rating only if the consumer chooses to post it:

$$R_{ij} := \begin{cases} r^*_{ij} & \text{if review is posted} \\ \varnothing & \text{otherwise} \end{cases} \tag{3}$$

The reviewing decision depends systematically on the same forces that shape internal evaluations – Q_{ij}, $\mathbb{E}_{ij}[Q_{ij}]$, X_{ij}, B_{ij}, θ_{ij} – as well as additional factors such as platform-level incentives, the perceived social cost of negative feedback, and the extremity of the consumption experience. This selection is a first-order concern: the distribution of observed ratings $\{R_{ij}\}$ may differ substantially from the distribution of internal evaluations $\{r^*_{ij}\}$.

A subset of ratings does not originate from the process above. *Fake reviews* enter directly as artificial observations r^{fake}_{ij} generated for manipulation purposes rather than from consumption experiences. Although their data-generating mechanism differs, they affect the distribution of observed ratings and are incorporated as a distinct component in our framework.

When Is Q_i a Useful Benchmark? The benchmark $\bar{r}_i \to Q_i$ as the number of reviews grows large implicitly assumes that quality is fixed, ratings reflect quality rather than idiosyncratic fit, and reviewers are representative. These assumptions are more realistic in some markets than others. In “cultural markets” like movies or books, intrinsic quality is largely fixed after release, and price rarely varies across consumers. By contrast, in service markets like hotels or restaurants, quality is dynamic – it responds to seller effort, which may itself respond to reviews – and price-quality trade-offs are salient. Our framework accommodates both settings by making explicit the components that enter the rating.

We now discuss the drivers of each component in turn. Table 1 maps each driver to the affected framework component, the direction of bias, and representative papers.

3 Drivers of Experienced Quality

Consumers experience a transaction-specific quality level $Q_{ij} = q(Q_i, e_{ij}, p_{ij})$, depending on intrinsic product quality Q_i, seller effort e_{ij}, and price p_{ij}. These drivers shape experienced quality and, in turn, the baseline internal evaluation r^*_{ij}.

3.1 Seller Effort and Dynamic Quality

Seller effort e_{ij} can vary across transactions and over time, generating dynamic quality patterns. Chevalier et al. (2018) study hotel reviews and managerial responses, showing that engagement with reviews is associated with improvements in underlying service quality, which then feed into higher subsequent ratings. Ananthakrishnan et al. (2023) provide evidence that responding to customer feedback leads firms to adjust operations, rather than merely managing perceptions.

These studies highlight a feedback loop: review-driven incentives affect e_{ij}, which shifts Q_{ij}, and ultimately shapes the distribution of observed ratings R_{ij}. This dynamic complicates the interpretation of ratings as static quality measures.

3.2 Price and Value-for-Money Considerations

Ratings often reflect perceived value-for-money rather than absolute quality. When consumers pay more, they may rate a given experience more harshly; when they pay less, they may be more forgiving. This mechanism implies that Q_{ij} is decreasing in p_{ij} for fixed Q_i and e_{ij}, even though higher prices may signal higher quality in equilibrium.

Carnehl et al. (2024) formalize this trade-off and study its implications for rating system design, showing that optimal pricing strategies depend critically on how consumers weight price against quality in their evaluations. Carnehl et al. (2022) use Airbnb data to document a dominant value-for-money effect: guests respond strongly to price changes in their ratings, and strategic hosts adjust both prices and effort to manage ratings and revenues. These results highlight how p_{ij} and e_{ij} jointly shape Q_{ij} and thus the baseline evaluation r^*_{ij}.

4 Drivers of Expectations

Consumers' ratings are also shaped by expectations formed prior to consumption, $\mathbb{E}_{ij}[Q_{ij}]$. Expectations can be influenced by expert certifications, platform badges, prior ratings, and recommendation algorithms. When realized quality falls short of expectations, disappointment systematically shifts the internal evaluation r^*_{ij} downward.

4.1 Certifications, Badges, and External Signals

Platforms and third parties use signals – Michelin stars, Airbnb Superhost badges, Academy Award nominations – to highlight high-quality options. These signals help users discover better products but can backfire when expectations are not met. The mechanism is straightforward: signals raise $\mathbb{E}_{ij}[Q_{ij}]$, and when $Q_{ij} < \mathbb{E}_{ij}[Q_{ij}]$, disappointment reduces r^*_{ij} even if absolute quality is unchanged.

Li et al. (2024) find that when restaurants lose their Michelin star, review valence improves – consistent with a disappointment mechanism where high expectations translate small shortfalls into lower ratings. Once the star is removed, expectations fall and similar realizations generate more favorable evaluations. Meister and Reinholtz (2025) show that Airbnb's Superhost designation leads to lower subsequent ratings, attributing this to an interplay of inflated expectations, host behavior changes, and shifts in reviewer composition. Rossi and Schleef (2024) document a disappointment-driven penalty in movie ratings following Academy Award nominations, especially among inexperienced users: nominations raise expectations, and when the realized experience is less exceptional than anticipated, users penalize the rating.

4.2 Recommendation Systems and Belief Formation

Recommendation systems shape expectations through personalized predictions and displayed ratings. Adomavicius et al. (2013) demonstrate that system-generated ratings serve as anchors for consumers' constructed preferences: viewers' post-consumption ratings shift toward the recommendation they observed, even when the recommendation was experimentally manipulated. This anchoring effect implies that $\mathbb{E}_{ij}[Q_{ij}]$ is directly influenced by algorithmic outputs, creating a feedback loop where biased recommendations contaminate subsequent ratings.

Aridor et al. (2024) collect belief data in MovieLens and show directly that prior expectations, shaped by recommendations and popularity signals, strongly influence post-consumption evaluations. Their evidence underscores that expectation-driven mechanisms are a central driver of r^*_{ij}, operating through the gap between anticipated and realized experiences.

5 Contextual and Environmental Drivers

The internal evaluation r^*_{ij} also depends on contextual factors unrelated to product characteristics or seller effort, collected in X_{ij}: mood, weather, social influence, and reviewer identity. Although these factors need not constitute "biases" in a normative sense, they complicate inference from ratings by introducing variation in r^*_{ij} orthogonal to Q_{ij}.

5.1 Mood, Weather, and Timing

Transient emotional states can shape how consumers evaluate and report experiences. Brandes and Dover (2022) show that users are more likely to post reviews on rainy days and that these reviews tend to be more negative. Since weather at the time of reviewing does not affect the underlying consumption experience, incidental affect can distort both r^*_{ij} and the composition of those selecting into reviewing. This finding suggests that aggregate ratings may fluctuate with local weather patterns in ways unrelated to actual quality.

5.2 Social Influence and Herding

Jacobsen (2015) and Sunder et al. (2019) document herding dynamics: ratings are shaped by crowd and peer opinions, undermining the wisdom-of-the-crowd ideal. Consumers anchor on existing ratings, adjust their evaluations accordingly, or selectively choose already-popular products. These mechanisms blur the line between independent signals and socially correlated ones, potentially amplifying early rating noise into persistent biases. When early ratings happen to be positive, later consumers may interpret their own positive experiences as confirmation rather than independent evidence; when early ratings are negative, the same mechanism can create downward spirals that are difficult to reverse.

5.3 Identity and Societal Biases

Reviewer identity and societal biases affect both review content and posting decisions. Aguiar (2024) show that female-led movies receive disproportionately lower ratings from male crowd reviewers despite similar assessments from professional critics. This pattern suggests that gender bias operates through the reviewing process itself, not through differences in underlying product quality. Bayerl et al. (2024) find that women leave more favorable reviews than men, potentially due to heightened concerns about social backlash from negative feedback or different baseline standards for evaluation.

In labor platforms, Bairathi et al. (2023) show buyers give higher public ratings to male freelancers even when private satisfaction is equal, suggesting that bias operates through the reviewing process rather than actual service differences. The gap between private satisfaction and public ratings points to strategic or social considerations that differ by freelancer gender. Aneja et al. (2025) demonstrate that labeling restaurants as Black-owned increases engagement and alters reviewer composition, indicating that identity disclosures systematically affect both participation and perceived quality. These findings highlight how ratings reflect not only product quality but also the social context in which evaluations occur.

6 Strategic Distortions

Strategic drivers deliberately distort the rating process through fake reviews, incentivized reviewing, reciprocity and retaliation, and managerial responses. These distortions operate on both the intensive margin (altering r_{ij}^*) and the extensive margin (changing who reviews).

6.1 Fake Reviews

Fake reviews are intentionally crafted to manipulate perceived reputation, entering as artificial observations r_{ij}^{fake} that bypass the standard rating process. Unlike other distortions that affect genuine evaluations, fake reviews are fabricated signals with no corresponding consumption experience.

Luca and Zervas (2016) show that promotional reviews increase under intensified competition, with restaurants facing negative demand shocks particularly likely to engage in fraud. He et al. (2022b) uncover organized markets where sellers procure fakes via social media intermediaries in exchange for free products. The market is sophisticated, with specialized brokers connecting sellers seeking reviews with reviewers willing to provide them. Akesson et al. (2023) estimate consumer welfare losses of approximately \$0.12 per dollar spent, arising from misallocation of consumers to inferior products.

Detection methods have evolved from textual features (Wu et al., 2020) to network-based strategies (He et al., 2022a) that exploit the clustering of fake reviewers around common sellers, achieving high accuracy even without access to review text. Yet Adamopoulos (2024) shows that biased early reviews produce persistent distortions through recommendation algorithms, amplifying harm beyond direct effects. Even when fake reviews are eventually detected and removed, their influence on algorithmic recommendations can persist, affecting which products consumers see and purchase long after the fakes are gone.

Theoretical work reveals additional complexities. Glazer et al. (2021) show that user uncertainty about authenticity can reduce informativeness – full transparency may dominate filtering because moderation can backfire. Mostagir and Siderius (2023) find that stricter auditing may unintentionally increase sellers' willingness to pay for fakes, as reduced supply raises their marginal value.

Generative AI introduces new challenges: large language models can produce fluent text evading linguistic detection, while AI-based tools offer promise for identifying suspicious patterns. If AI-generated fakes become indistinguishable from authentic reviews, platforms may need verification-based approaches rather than content-based detection.

6.2 Incentivized and Influenced Reviews

Not all seller-driven distortions stem from outright fraud. Incentivized reviews can distort ratings more subtly, affecting both r^*_{ij} and D_{ij} through legitimate channels.

Limited participation is widely documented: Brandes et al. (2022) report that only 11% of hotel guests leave reviews on traditional platforms, compared to approximately 70% on Airbnb (Fradkin et al., 2021). These stark differences illustrate how platform design shapes D_{ij} and, consequently, the composition of observed ratings. The 11% who review are unlikely to be representative of all guests; they are more likely to have had extreme experiences (very good or very bad) that motivated the effort of writing a review. Incentivizing reviews thus holds promise for informativeness by drawing in the "silent majority" of moderate consumers.

However, incentives also influence content, and the effects are not uniform. Woolley and Sharif (2021) show that incentivized reviews contain more positive emotional language, suggesting that incentives do not merely increase volume but also shift the character of what is written. Fradkin and Holtz (2023) find that Airbnb coupons increased volume but made ratings more negative and less correlated with transaction quality – implying incentives attracted different types of reviewers rather than merely motivating existing ones to participate. Karaman (2021) finds that solicitation increases volume and reduces selection bias without altering content, distinguishing the effects of asking for reviews from paying for them. Li et al. (2020) develop a signaling model predicting that only high-quality sellers will offer rewards for truthful feedback, with supporting evidence from Taobao – suggesting that seller-initiated incentive programs may have different effects than platform-initiated ones.

6.3 Reciprocity, Retaliation, and Managerial Responses

Two-sided review systems introduce strategic dynamics where users condition reviews on anticipated counterparty reactions, creating interdependencies that distort the information revealed. This is particularly important in platforms like Airbnb, eBay, and labor marketplaces where both sides of a transaction can rate each other.

Hui et al. (2018) study eBay's 2008 policy preventing sellers from leaving negative feedback for buyers. After the change, low-quality sellers experienced reduced success or exited entirely, and overall service quality improved – demonstrating that institutional design directly influences strategic distortions. The ability of sellers to retaliate against negative buyer feedback had previously protected low-quality sellers from honest negative reviews. Fradkin et al. (2021) document that Airbnb users time reviews strategically under sequential revelation, waiting to see the other party's rating before committing to their own. Simultaneous revelation increased review rates and reduced reciprocal behavior. Still, highly positive reviews persist – Proserpio et al. (2018) attribute this to interpersonal dynamics in sharing economy transactions, where face-to-face contact creates social pressure toward leniency even when no explicit retaliation is possible.

Over time, these dynamics produce "reputation inflation." Filippas et al. (2018) document increasingly positive ratings on an online labor platform without corresponding performance improvements, suggesting that strategic considerations progressively crowd out honest assessment as participants learn to navigate the system.

Managerial responses create additional channels for strategic interaction. Proserpio and Zervas (2017) find that responses lead to higher subsequent ratings and increased review volume, potentially through both quality improvement and expectation management. Wang and Chaudhry (2018) argue that responses positively shape buyer beliefs about seller attentiveness. However, Proserpio et al. (2021) show that female reviewers become less likely to post negative feedback after responses are introduced, possibly due to anticipation of confrontational replies – illustrating how interventions can have unintended distributional consequences that affect which voices are heard.

7 Taste Heterogeneity and Aggregation

We now consider how heterogeneity in tastes – captured by θ_{ij} – affects aggregate informativeness. Even if each review faithfully reflects the reviewer's experience, differences in preferences, stringency, and selection patterns translate into systematic distortions when ratings are pooled.

7.1 Self-Selection into Consumption and Reviewing

Hu et al. (2017) distinguish selection into reviewing (underreporting bias) from selection into consumption (acquisition bias). Even if all buyers reviewed, they would differ systematically from non-buyers in their taste for the product – consumers who purchase a product typically have higher expected utility from it than those who do not. This implies $\mathbb{E}(\theta_{ij}|\text{Purchase}) > 0$, so $\bar{r}_i > Q_i$ in general.

This is not necessarily problematic if consumers care primarily about relative ratings. The challenge arises because $\mathbb{E}(\theta_{ij}|\text{Purchase})$ varies across products, so ratings mislead when selection patterns differ systematically. Products with niche appeal may attract only their natural audience, yielding high ratings that overstate quality for the broader market. Mainstream products may attract casual consumers who are less favorably disposed, yielding lower ratings than their quality warrants. Schoenmueller et al. (2020) document that buyers with strong opinions – whether positive or negative – are more likely to rate, producing the characteristic J-shaped distributions observed for many products.

7.2 Learning from Reviews and Long-Run Dynamics

Review systems create dynamic feedback loops: ratings shape beliefs, beliefs shape consumption, and consumption generates new ratings. Acemoglu et al. (2022) develop a Bayesian model showing how information revelation structure affects learning. They derive conditions under which social learning is systematically biased and, counterintuitively, demonstrate that more detailed information does not always improve outcomes – too much information can actually slow convergence to truth by inducing consumers to condition on noisy signals.

Bondi (2025) develops a model of social learning from consumer reviews, showing that reviews systematically advantage lower-quality and more polarizing products. The mechanism is taste-based self-selection: polarizing products attract consumers with strong prior tastes who then leave reviews reflecting those tastes, which in turn attract similar future buyers. In stark contrast with the winner-takes-all dynamics of classic observational learning models, social learning from opinions generates excessive choice fragmentation and long-run biases that persist even without strategic manipulation.

7.3 Reviewer Stringency and Aggregation

Bondi et al. (2024) show that experienced users consume higher-quality products and post harsher evaluations for any given quality. When ratings are pooled without adjustment, scores compress, penalizing high-quality products – ranking reversals occur for roughly 8% of movie pairs. The compression arises because stringent reviewers disproportionately consume high-quality products, leaving low-quality products to be rated by more lenient reviewers. This selection pattern means that the reviewer pool differs systematically across products in ways that distort aggregate ratings.

Dai et al. (2018) and Carayol and Jackson (2024) develop structural corrections that yield substantial informativeness gains using Yelp and wine data. Carayol and Jackson (2024) develop a method to simultaneously estimate underlying quality and reviewer reliability from rating patterns, exploiting the insight that reviewers who agree with each other provide information about their reliability on other products. Such approaches require sufficient data but offer promise for extracting signal from noisy, heterogeneous ratings.

7.4 User Responses to Aggregation

Lee et al. (2021) find that low dispersion leads users to treat the average as sufficient, while high dispersion encourages deeper review reading, especially of extreme opinions. This suggests that how platforms display rating distributions affects how much information consumers extract from reviews. When ratings are tightly clustered, consumers may rationally rely on simple summary statistics; when they are dispersed, the value of reading individual reviews increases, potentially changing which products consumers ultimately select.

8 Platform Design Implications

Our framework clarifies where biases enter and where interventions can be effective. Table 1 maps drivers to design targets.

Targeting D_{ij}: Increasing Participation. Low review rates create cold-start problems for new products and amplify selection biases for established ones. The selection problem is fundamental: the consumers who choose to review are systematically different from those who do not, and this difference varies across products in ways that distort aggregate ratings. Solicitation (Karaman, 2021) effectively reduces extremity bias by drawing in consumers with moderate experiences; monetary incentives (Fradkin and Holtz, 2023) have more ambiguous effects, potentially attracting different reviewer types rather than simply motivating existing ones. Simultaneous revelation (Fradkin et al., 2021) increases rates while reducing strategic timing, addressing both participation and strategic distortions. Platform designers must weigh the benefits of increased volume against potential distortions in content.

Targeting B_{ij}: Reducing Strategic Distortions. Detection (He et al., 2022a), policy changes (Hui et al., 2018), and revelation design all target manipulation. Network-based detection exploits structural features – such as the clustering of suspicious reviewers around common sellers – that are costly for manipulators to circumvent. The key advantage of network methods is that they identify suspicious patterns in who reviews whom, rather than relying on textual features that sophisticated fake review operations can easily disguise. However, aggressive enforcement may backfire (Mostagir and Siderius, 2023) as strategic actors adapt to new detection methods. The fundamental challenge is an arms race: as platforms improve detection, fake review markets invest in circumvention.

Targeting θ_{ij}: Adjusting for Heterogeneity. Debiasing algorithms (Bondi et al., 2024; Dai et al., 2018; Carayol and Jackson, 2024) can substantially improve informativeness but require rich data and may be difficult to implement transparently. The core idea is to estimate each reviewer's baseline stringency and adjust their ratings accordingly, but this requires observing the same reviewer across multiple products and raises questions about how to communicate adjusted ratings to users. Whether platforms should display adjusted ratings directly or provide user-controlled adjustment tools remains an open question with trade-offs between paternalism and user autonomy.

Targeting $\mathbb{E}_{ij}[Q_{ij}]$: Managing Expectations. Disappointment effects are hard to address since certifications serve legitimate discovery functions even when they inflate expectations. Platforms might provide more granular certification information or adjust criteria dynamically based on observed disappointment patterns, though this introduces complexity and potential gaming. The challenge is that the same signals that help consumers find good products also create inflated expectations that lead to disappointment.

9 Conclusion

Online ratings emerge from a multi-stage process in which experienced quality, expectations, contextual factors, strategic incentives, idiosyncratic tastes, and selection into reviewing all play a role. We have developed a unified framework making these components explicit and used it to organize a growing theoretical and empirical literature on rating biases and platform design.

Three insights emerge from our review. First, distortions enter at distinct stages of the rating process, and effective interventions must be targeted accordingly. A platform concerned about fake reviews faces a different design problem than one concerned about selection bias or disappointment effects. Second, many biases are deeply intertwined: selection into reviewing is shaped by the same forces that shape review content, and strategic behavior responds to platform policies in complex ways. This interdependence means that interventions targeting one distortion may exacerbate another. Third, some distortions – such as disappointment effects from quality signals – arise from mechanisms that also serve legitimate functions, creating inherent trade-offs for platform designers. Certifications help consumers discover high-quality products even as they inflate expectations and generate disappointment.

The rise of artificial intelligence presents both opportunities and challenges for online review systems. On the detection side, machine learning can identify suspicious patterns at unprecedented scale, while network-based methods offer promising approaches for catching sophisticated manipulation. On the generation side, large language models can produce fake reviews increasingly difficult to distinguish from genuine feedback, potentially undermining the informational foundations that make reputation systems valuable. The interaction between AI-assisted consumers – who may rely on algorithmic summaries rather than reading individual reviews – and AI-generated content creates feedback loops whose welfare implications remain largely unexplored. Understanding how these technologies reshape the information environment represents a critical frontier for future research.

Table 1: Drivers of Online Ratings: Framework Components, Biases, and Key Papers

Driver	**Component**	**Direction of Bias**	**Key Papers**
Panel A: Experienced Quality			
Price / value-for-money	Q_{ij}	Higher price → lower rating	Carnehl et al. (2024), Carnehl et al. (2022)
Seller effort	Q_{ij}	Dynamic quality; feedback loops	Chevalier et al. (2018), Ananthakrishnan et al. (2023)
Panel B: Expectations			
Certifications / badges	$\mathbb{E}_{ij}$	Disappointment penalty	Li et al. (2024), Meister & Reinholtz (2025), Rossi & Schleef (2024)
Recommendations	$\mathbb{E}_{ij}$	Anchoring; disappointment	Adomavicius et al. (2013), Aridor et al. (2024)
Panel C: Contextual Drivers			
Weather / mood	X_{ij}, D_{ij}	Negative affect → lower ratings	Brandes & Dover (2022)
Social influence	X_{ij}	Anchoring; correlated signals	Jacobsen (2015), Sunder et al. (2019)

Continued on next page

Table 1 continued

Driver	Component	Direction of Bias	Key Papers
Gender bias (reviewer)	X_{ij}, θ_{ij}	Women more favorable	Bayerl et al. (2024), Bairathi et al. (2023)
Gender bias (product)	X_{ij}, r^*_{ij}	Female-led products lower	Aguiar (2024), Proserpio et al. (2021)
Race / identity	X_{ij}, D_{ij}	Alters composition and ratings	Aneja et al. (2025)
Panel D: Strategic Distortions			
Fake reviews	r^{fake}_{ij}	Inflates ratings; erodes trust	Luca & Zervas (2016), He et al. (2022a,b), Glazer et al. (2021), Mostagir & Siderius (2023), Adamopoulos (2024), Akesson et al. (2023)
Incentivized reviews	B_{ij}, D_{ij}	More volume; mixed valence	Woolley & Sharif (2021), Fradkin & Holtz (2023), Karaman (2021), Li et al. (2020)
Reciprocity / retaliation	B_{ij}, D_{ij}	Inflates; reduces honesty	Fradkin et al. (2021), Hui et al. (2018), Proserpio et al. (2018)
Reputation inflation	B_{ij}	Upward drift	Filippas et al. (2018)

Continued on next page

Table 1 continued

Driver	**Component**	**Direction of Bias**	**Key Papers**
Managerial responses	B_{ij}, Q_{ij}, D_{ij}	Higher ratings; quality feedback	Proserpio & Zervas (2017), Wang & Chaudhry (2018), Chevalier et al. (2018)
Panel E: Taste Heterogeneity			
Self-selection	θ_{ij}	Buyers more favorable	Hu et al. (2017), Schoenmueller et al. (2020)
Stringency heterogeneity	θ_{ij}	Compresses differences	Bondi et al. (2024), Dai et al. (2018), Carayol & Jackson (2024)
Learning dynamics	θ_{ij}, D_{ij}	Fragmentation; polarization	Acemoglu et al. (2022), Bondi (2025)
Panel F: Selection into Reviewing			
Extremity bias	D_{ij}	Overweights strong opinions	Schoenmueller et al. (2020), Brandes et al. (2022)
Low participation	D_{ij}	Cold-start; entry barriers	Brandes et al. (2022)
Dispersion effects	D_{ij}	High dispersion → engagement	Lee et al. (2021)

References

Acemoglu, D., Makhdoumi, A., Malekian, A., and Ozdaglar, A. (2022). Learning from reviews: The selection effect and the speed of learning. *Econometrica*, 90(6):2857–2899.

Adamopoulos, P. (2024). The spillover effect of fraudulent reviews on product recommendations. *Management Science*, 70(12):8818–8832.

Adomavicius, G., Bockstedt, J. C., Curley, S. P., and Zhang, J. (2013). Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Information Systems Research*, 24(4):956–975.

Aguiar, L. (2024). Bad apples on Rotten Tomatoes: Critics, crowds, and gender bias. *Marketing Science*, 43(5):1021–1042.

Akesson, J., Hahn, R. W., Metcalfe, R. D., and Monti-Nussbaum, M. (2023). The impact of fake reviews on demand and welfare. NBER Working Paper.

Ananthakrishnan, U., Proserpio, D., and Sharma, S. (2023). I hear you: Does quality improve with customer voice? *Marketing Science*, 42(6):1143–1161.

Aneja, A., Luca, M., and Reshef, O. (2025). The benefits of revealing race: Evidence from minority-owned local businesses. *American Economic Review*, 115(2):660–689.

Aridor, G., Gonçalves, D., Kong, R., Kluver, D., and Konstan, J. (2024). The MovieLens beliefs dataset: Collecting pre-choice data for online recommender systems. *Proceedings of the 18th ACM Conference on Recommender Systems*.

Bairathi, M., Lambrecht, A., and Zhang, X. (2023). Gender disparity in online reputation. SSRN Working Paper.

Bayerl, A., Dover, Y., Riemer, H., and Shapira, D. (2024). Gender rating gap in online reviews. *Nature Human Behaviour*, 8:538–551.

Bondi, T. (2025). Alone, together: A model of social (mis)learning from consumer reviews. *Marketing Science*, 44(1):1–22.

Bondi, T., Rossi, M., and Stevens, R. L. (2024). The good, the bad and the picky: Consumer heterogeneity and the reversal of product ratings. *Management Science*, 70(11):7469–7489.

Brandes, L. and Dover, Y. (2022). Offline context affects online reviews: The effect of post-consumption weather. *Journal of Consumer Research*, 49(4):595–615.

Brandes, L., Godes, D., and Mayzlin, D. (2022). Extremity bias in online reviews: The role of attrition. *Journal of Marketing Research*, 59(4):675–695.

Carayol, N. and Jackson, M. O. (2024). Finding the wise and the wisdom in a crowd. *The Economic Journal*, 134(663):2712–2745.

Carnehl, C., Schaefer, M., Stenzel, A., and Tran, K. D. (2022). Value for money and selection: How pricing affects Airbnb ratings. IGIER Working Paper.

Carnehl, C., Stenzel, A., and Schmidt, P. (2024). Pricing for the stars: Dynamic pricing in the presence of rating systems. *Management Science*, 70(3):1755–1772.

Chevalier, J. A., Dover, Y., and Mayzlin, D. (2018). Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science*, 37(5):688–709.

Dai, W., Jin, G., Lee, J., and Luca, M. (2018). Aggregation of consumer ratings: An application to Yelp.com. *Quantitative Marketing and Economics*, 16:289–339.

Filippas, A., Horton, J. J., and Golden, J. (2018). Reputation inflation. *Proceedings of the 2018 ACM Conference on Economics and Computation*.

Fradkin, A. and Holtz, D. (2023). Do incentives to review help the market? *Marketing Science*, 42(5):853–865.

Fradkin, A., Grewal, E., and Holtz, D. (2021). Reciprocity and unveiling in two-sided reputation systems. *Marketing Science*, 40(6):1013–1029.

Glazer, J., Herrera, H., and Perry, M. (2021). Fake reviews. *The Economic Journal*, 131(636):1772–1787.

He, S., Hollenbeck, B., Overgoor, G., Proserpio, D., and Tosyali, A. (2022a). Detecting fake-review buyers using network structure. *Proceedings of the National Academy of Sciences*, 119(47):e2211932119.

He, S., Hollenbeck, B., and Proserpio, D. (2022b). The market for fake reviews. *Marketing Science*, 41(5):896–921.

Hu, N., Pavlou, P. A., and Zhang, J. (2017). On self-selection biases in online product reviews. *MIS Quarterly*, 41(2):449–475.

Hui, X., Saeedi, M., and Sundaresan, N. (2018). Adverse selection or moral hazard: An empirical study. *The Journal of Industrial Economics*, 66(3):610–649.

Jacobsen, G. D. (2015). Consumers, experts, and online product evaluations: Evidence from the brewing industry. *Journal of Public Economics*, 126:114–123.

Karaman, H. (2021). Online review solicitations reduce extremity bias. *Management Science*, 67(7):4420–4445.

Lee, S., Lee, S., and Baek, H. (2021). Does the dispersion of online review ratings affect review helpfulness? *Computers in Human Behavior*, 117:106670.

Li, L., Tadelis, S., and Zhou, X. (2020). Buying reputation as a signal of quality: Evidence from an online marketplace. *The RAND Journal of Economics*, 51(4):965–988.

Li, X., Deng, Y., Manchanda, P., and De Reyck, B. (2024). Can lower expert opinions lead to better consumer ratings?: The case of Michelin stars. *Management Science*, 70(8):5195–5214.

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

Meister, M. and Reinholtz, N. (2025). Quality certifications influence user-generated ratings. *Journal of Consumer Research*.

Mostagir, M. and Siderius, J. (2023). Strategic reviews. *Management Science*, 69(2):904–921.

Pocchiari, M., Proserpio, D., and Dover, Y. (2024). Online reviews: A literature review and roadmap for future research. *International Journal of Research in Marketing*, 41(4):832–858.

Proserpio, D. and Zervas, G. (2017). Online reputation management. *Marketing Science*, 36(5):645–665.

Proserpio, D., Xu, W., and Zervas, G. (2018). You get what you give: Theory and evidence of reciprocity in the sharing economy. *Quantitative Marketing and Economics*, 16:371–407.

Proserpio, D., Troncoso, I., and Valsesia, F. (2021). Does gender matter? The effect of management responses on reviewing behavior. *Marketing Science*, 40(6):1199–1213.

Rossi, M. and Schleef, F. (2024). Quality disclosures and disappointment: Evidence from the Academy nominations. CESifo Working Paper.

Schoenmueller, V., Netzer, O., and Stahl, F. (2020). The polarity of online reviews: Prevalence, drivers and implications. *Journal of Marketing Research*, 57(5):853–877.

Sunder, S., Kim, K. H., and Yorkston, E. A. (2019). What drives herding behavior in online ratings? *Journal of Marketing*, 83(6):93–112.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 8:321–340.

Wang, Y. and Chaudhry, A. (2018). When and how managers' responses to online reviews affect subsequent reviews. *Journal of Marketing Research*, 55(2):163–177.

Woolley, K. and Sharif, M. A. (2021). Incentives increase relative positivity of review content. *Journal of Marketing Research*, 58(3):539–558.

Wu, Y., Ngai, E. W., Wu, P., and Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, 132:113280.

Appendix

A.1 Notation Summary

For reference, we summarize the key notation used throughout the paper:

- Q_i: Intrinsic quality of product i
- e_{ij}: Seller effort in transaction j for product i
- p_{ij}: Price paid by consumer j for product i
- $Q_{ij} = q(Q_i, e_{ij}, p_{ij})$: Experienced quality – the realized consumption utility
- $\mathbb{E}_{ij}[Q_{ij}]$: Consumer j's prior expectation of quality before consumption
- X_{ij}: Contextual factors (weather, mood, social influence, reviewer identity)
- B_{ij}: Strategic distortions (incentives, reciprocity, retaliation)
- θ_{ij}: Idiosyncratic tastes and reviewer stringency
- r^*_{ij}: Internal evaluation – the rating the consumer would give
- D_{ij}: Indicator for whether consumer j posts a review (selection)
- R_{ij}: Observed rating ($= r^*_{ij}$ if $D_{ij} = 1$, $= \varnothing$ otherwise)
- r^{fake}_{ij}: Fake reviews that bypass the standard rating process
- $\bar{r}_i$: Average observed rating for product i

A.2 Framework Microfoundations

The framework in Section 2 can be derived from a simple model of consumer behavior. Consider a consumer j who purchases product i at price p_{ij}. The consumer's utility from consumption is:

$$U_{ij} = Q_{ij} - p_{ij} + \varepsilon_{ij}$$

where $Q_{ij} = q(Q_i, e_{ij}, p_{ij})$ is experienced quality and ε_{ij} is an idiosyncratic shock. The consumer forms an internal evaluation r^*_{ij} by comparing realized utility to expected utility:

$$r^*_{ij} = g(U_{ij} - \mathbb{E}_{ij}[U_{ij}]) + h(X_{ij}) + B_{ij} + \theta_{ij}$$

where $g(\cdot)$ captures disappointment/elation effects (with $g(0) = 0$, $g' > 0$), $h(X_{ij})$ represents contextual influences, B_{ij} captures strategic considerations, and θ_{ij} reflects baseline stringency.

The decision to post a review depends on the expected benefit relative to the cost c:

$$D_{ij} = \mathbf{1}\{\text{Expected benefit of reviewing} > c\}$$

Benefits may include altruism (helping other consumers), reciprocity (rewarding/punishing sellers), or platform incentives. This generates selection: consumers with extreme experiences or strong preferences are more likely to review.

A.3 Identification Challenges

Empirically distinguishing between framework components presents several challenges that researchers must navigate:

Expectations vs. experienced quality. When a certified product receives lower ratings, this could reflect (a) disappointment arising from inflated expectations, (b) actual quality decline due to reduced seller effort post-certification, or (c) compositional changes in the reviewing population. Separating these mechanisms requires either direct measurement of expectations (Aridor et al., 2024) or exogenous variation in certification that is orthogonal to quality.

Selection vs. rating content. Observed rating distributions confound who reviews (D_{ij}) with what ratings they give (r^*_{ij}). A shift toward more positive ratings could reflect either changes in the reviewing population or changes in how a fixed population evaluates experiences. Panel data tracking individual reviewers across products can help disentangle these channels, but selection into which products to consume remains a confound.

Strategic behavior. Reciprocity, retaliation, and incentive effects are difficult to identify because they respond to anticipations of future interactions. Natural experiments – such as the policy changes studied by Hui et al. (2018) and Fradkin et al. (2021) – provide the cleanest identification by generating exogenous shifts in the strategic environment.

Taste heterogeneity vs. quality. High ratings for niche products could indicate either genuinely high quality or favorable self-selection by consumers whose tastes align with the product. Disentangling these explanations requires observing the same consumers across multiple products to estimate individual-specific taste parameters, as in Bondi et al. (2024) and Dai et al. (2018).

A.4 Extensions of the Basic Framework

The baseline framework admits several natural extensions:

Dynamic quality. When Q_i evolves over time in response to reviews, the system exhibits feedback loops. Let $Q_i^{t+1} = \phi(Q_i^t, \bar{r}_i^t)$ where ϕ captures how sellers adjust quality in response to ratings. This formulation, which connects to the empirical work of Chevalier et al. (2018) and Ananthakrishnan et al. (2023), creates path dependence: early reviews shape quality trajectories and can have persistent effects.

Two-sided ratings. In platforms where both parties rate each other (e.g., Airbnb, Uber), ratings become strategic complements or substitutes. Let r_{ij}^* depend on the anticipated rating $\tilde{r}_{ji}$ from the counterparty:

$$r_{ij}^* = f(Q_{ij}, \mathbb{E}_{ij}[Q_{ij}], X_{ij}, \tilde{r}_{ji}, \theta_{ij})$$

This introduces coordination games and can generate equilibria with inflated ratings on both sides, as documented empirically by Fradkin et al. (2021) and Filippas et al. (2018).

Heterogeneous interpretation. Different consumers may weight the same rating differently based on their own characteristics. If consumer k interprets product i's rating as $\hat{Q}_i^k = \bar{r}_i + \delta_k$, where δ_k captures systematic optimism or pessimism, then ratings transmit information imperfectly even absent bias in their generation. This connects to the learning dynamics studied by Acemoglu et al. (2022).

Multi-attribute ratings. Many platforms elicit ratings along multiple dimensions (e.g., cleanliness, location, communication for Airbnb). Let $\mathbf{r}_{ij}^* = (r_{ij,1}^*, \ldots, r_{ij,K}^*)$ be a vector of attribute-specific ratings. The aggregation problem becomes more complex: consumers must weight attributes according to their own preferences, and platforms must decide how to summarize multidimensional information. Attribute-level ratings can improve informativeness by allowing consumers to focus on dimensions they care about, but they also increase the cognitive burden on both reviewers and readers.