Truth Warrants Increase Economic Value and Accelerate Product Sales in Digital Marketplaces

April, 2025

Abstract

Misleading advertisements in digital marketplaces deceive buyers of online goods, posing significant challenges to platform integrity and buyer protection. Existing content moderation and reputation systems have proven insufficient in addressing the root incentives that drive false advertising. This paper introduces a novel market-based economic intervention, "truth warrants" that allows online advertisers to guarantee the accuracy of their claims, and presents buyers with the ability to seek recourse when they are misled. Through a series of controlled online experiments in a competitive two-sided marketplace, we demonstrate the economic value of truth warrants in comparison to existing reputation and rating systems. The introduction of truth warrants penalizes cheating advertisers that make false claims, and promotes accountability among online advertisers, without affecting the profits or sales of honest advertisers. Empirical results ($n = 250$ users and 5208 rounds of product sales) and demonstrate that truth warrants nearly double the economic value provided by reputation signals and reduce the time it takes for the first sale of a product by nearly half the full duration of our experiment.

1 Introduction

Incentives are central to the functioning of two-sided marketplaces, where producers and consumers interact in a digital economy. Classical economic theories emphasize the importance of aligning incentives to ensure efficient market outcomes (Coase, 2013; Spence, 1973; Stiglitz and Weiss, 1981a). In modern digital platforms, consumers seek to maximize their utility from acquiring reliable goods and services, while producers seek to maximize the value gained from the provision thereof. This exchange takes place through advertisements of goods, where a good may be a physical product or a digital service. Digital advertising has turned into a market worth nearly half a trillion dollars in 2024[1], and itself underpins the platform economics for Meta, Google, Amazon, TikTok, Instagram, and a number of digital platforms with over a trillion dollar market capitalization.

The problem is the information asymmetry between producers and consumers: while a producer has complete information on the true quality of a good, they may choose to advertise a higher quality, misleading the consumer about the true quality of the good. The consumer cannot verify whether a producer's claim is true or false, given that the consumer does not have any information on the true quality of the good. Therefore, consumers can never guarantee that the goods they acquire will indeed match the quality at which they are advertised, since that information is privately held by the producer that provides them. Considering the cost of production of lower quality good is less than the same for producing a higher quality good, there is incentive for producers to mislead consumers in the near term, since selling a lower quality good under the claim of providing a high quality good results in higher gains. In physical marketplaces, this would have resulted in consumer aversion to engage with deceptive producers having knowledge of their identity, and the history of their deceptive practices, including through word-of-mouth. However on digital platforms, identities of producers may not directly be revealed beyond what they choose to proffer, and the option of anonymity allows deceptive sellers to exploit the information asymmetry and produce misleading advertisements to deceive consumers.

In this research, we propose a theoretical model of a marketplace and estimators to evaluate the economic value of truth warrants, hypothesizing that it provides a greater value than reputation signaling in current marketplaces. Thereafter, we use empirical data drawn from online experiments with 250 human participants in a two-

[1] https://www.grandviewresearch.com/industry-analysis/digital-advertising-market-report

sided digital marketplace to fit least squares regressions to evaluate the proposed hypotheses. Our results show that truth warrants offer significantly higher economic value, nearly twice ($\beta = 0.3088$ (SE $= 0.013$)) as effective in increasing product sales as reputation signals ($\beta = 0.1623$ (SE $= 0.005$)). Furthermore, they reduce the time taken for the first sale of an advertised product by nearly half of the full duration of our experiment ($\beta = -3.3471$ (SE $= 0.091$)).

1.1 Platform Interventions to Address Information Asymmetry

Digital platforms are well aware of the information asymmetry between producers and consumers and actively attempt to mitigate it in a number of ways. Below, we list the key areas of these consumer protection measures or *interventions* drawn from global consumer protection reports (OECD, 2022, 2021), further including a tangible example of each on a popular digital platform:

1. Soliciting verification information from third-party producers before permitting them to display advertisements on their platforms.

    - Digital platforms intend to create friction for deceptive sellers to reenter the marketplace since the online verification of company registration is required to be eligible as a seller on Amazon and Taobao.

2. Allowing consumers to register complaints in the case of significant damage incurred from the acquisition of the good.

    - The intention is to create friction for deceptive sellers to reenter the marketplace and the online verification of company registration is required to be eligible as a seller on Amazon and Taobao.

3. Introduction of producer reputation systems in order for consumers that have been verified to have purchased a good to have the ability to optionally add a rating to the producer, ideally based on the true quality of the good. The worse a producer's reputation gets as a result of deceptive sales to cheated consumers, the less likely their future sales are.

    - Almost every e-commerce platform has introduced 'seller ratings' gathered from verified buyers as a central feature designed to allow for transparent

feedback including user-uploaded images to display issues with the product(s). Even for multimedia platforms like TikTok, YouTube, and Instagram, users may choose to provide positive or negative engagement signals upon encountering all manner of creator-produced content on the platform.

4. Introducing community-based moderation to allow consumers to govern the outcome relating to the production or consumption of a good in the marketplace.

    - Using the 'wisdom of the crowd' is an ages-old content and the idea of 'peer juries' of platform users is operationalized to adjudicate user complaints by Chinese food delivery service Meituan with over 290 million monthly users, Idle Fish[2], a secondary market by Alibaba, and Reddit, X (formerly, Twitter), and now Meta with their X-inspired Community Notes rollout[3].

These interventions are well-intended, laboriously designed, and often expensive to deploy given that they may directly affect the bottom line for the platform businesses deploying them.

1.2 Limitations of Platform Interventions

Despite their best attempts at addressing the information asymmetry, most digital marketplaces often struggle to achieve this balance in practice, with glaring failures in fostering honesty and transparency among producers. Taking the case of each intervention above, let us enlist the obstacles to their success:

1. **Third-party verification**: While Amazon or Taobao may attempt to impose limitations to third party sellers signing up to advertise products on their platform, it is against their own business incentives as profit-seeking corporations, to impede platform growth. The tension between making third-party verification 'too hard for the average honest seller' to qualify, and 'too easy for the masterful dishonest seller' to bypass allows the latter to make their way into the system to cause consumer harm (sometimes, with multiple fake seller accounts scamming consumers and the platform to the tune of a million dollars[4]).

2. **Consumer-initiated Feedback**: Feedback systems place the onus of responsibility on the cheated consumer, who is now responsible for expending time

[2] https://www.wsj.com/articles/online-shopping-dispute-alibaba-meituan-11655489957
[3] https://about.meta.com/technologies/community-notes/
[4] https://bit.ly/seller-scams-amazon

and energy assuming they even have awareness of the appropriate channels to seek recourse. The feedback system, while helpful, faces the serious limitation of having a misplaced burden of responsibility on the consumers already harmed by the deceptively advertised good. On a separate note, feedback mechanisms work best when they have the agency to hold platforms accountable, which they often do not, as evidenced by the case of Meta's investment into an external Oversight Board, that has been criticized as being unable to successfully hold the platform to its own words, when judged by it's own 2023 reports of over 40% of its recommendations having been met with no evidence of platform implementation[5].

3. **Reputation Systems**: Reputation systems face numerous implementation challenges including the tracking and verification process for authentic consumers of products. Producers in marketplaces like Amazon and eBay buy fake reviews to solve the 'cold start' problem for new products and manipulate ratings for low quality products (He et al., 2022a; Pooja and Upadhyaya, 2024). Stepping back though, it is clear that yet again, the burden of responsibility to avoid deception falls squarely on consumers who are already deceived once by a product; only then can they offer 'verified' feedback in the form of a negative rating intended to impact the producer reputation. This fails to limit the original occurrence of consumer harm, limiting its relevance as a meaningful consumer *protection* measure in the first place, instead hoping to serve more as a soft deterrent for producer deception.

4. **Consumer-led Content Moderation**: While the idea of users as content moderators has worked out well for Reddit, Wikipedia, Stackoverflow, and reflects a potential opportunity to provide platform users with control over their own feeds, so to speak, the challenge is the lack of guarantees that consumers can censure the digital platform itself since this is a centralized intervention. While consumers may hold the power to *make* the moderation decision, the platform and its leadership ultimately control its *implementation*, leading into the same issue with Meta's apparent non-implementation of the recommendations of the Oversight Board. But beyond this, the idea of consumer-led moderation process is a time-consuming process and much of the harm is already done by

[5] https://www.oversightboard.com/wp-content/uploads/2024/03/Oversight-Board-H2-2023-Transparency-Report-March-2024.pdf

the time this moderation is applied. X's Community Notes intervention still allows a large volume of content to remain available on the platform before a correction with additional context is issued which is a critical challenge considering false information spreads much faster than true information on social media (Vosoughi et al., 2018). Interestingly, X (now, xAI) owner Elon Musk has expressed uncertainty as to the validity of Community Notes even though he previously championed the intervention[6]. He recently announced his belief that there were ways to subvert the ratings system for Community Notes[7] despite its algorithmic transparency and notes data being made available for end-to-end external auditing.

The key challenges to the issue of information asymmetry are then with **centralized authority**, **adversarial gaming by producers**, and **misplaced responsibility on consumers** resulting in no penalties applied to the production of lies while those misled end up paying the price. While many interventions attempt to reinstate balance in two-sided marketplaces, some tackling centralized authority through community governance (the Fediverse including Bluesky, Nostr, or Mastodon), others attempt to address the adversarial activities of producers, none of the existing interventions–to our knowledge–prevent the *production of lies* which exacerbates the information asymmetry and ultimately results in market failure. Notably, markets do not self-correct market failures so the consumers continue to deal with compounding harm. The lack of consequences for the production of misleading claims online implies that deceptive practices will be rewarded, eroding consumer welfare (Stiglitz and Weiss, 1981b) and trust in the marketplace.

2 Related Work

2.1 Digital Harm from Information Asymmetry

The problem of misinformation in marketplaces is a direct cause of the information asymmetry, where sellers have more knowledge about product quality than buyers. This information imbalance can lead to adverse selection, in which low-quality goods dominate the market, leading to a "marketplace for lemons" (Akerlof, 1970; Sheng et al., 2010). Extending this understanding of "lemon markets" to the misinformation context, speakers have insufficient incentives to share accurate information, as

[6] https://archive.ph/XpiDh
[7] https://archive.ph/m25yK

they stand to benefit from exaggerating their claims, thereby increasing influence, followers, or profits. Consequently, claims buyers make decisions that are not in their best interest, a phenomenon that parallels the dynamics of fake news production. Deceptive advertising practices undermine platform integrity and buyer trust while distorting decision-making processes (Sänger et al., 2016). Existing reputation systems and content moderation approaches have proven insufficient to address these challenges comprehensively (Wright et al., 2008; Kutabish et al., 2023; Kenning et al., 2018). Clearing platforms of misleading claims improves the "social welfare", benefiting honest sellers through the profits they make, and honest buyers through the "utility" they gain from a purchased product. The economic equilibrium between sellers and buyers determines the total social welfare in the marketplace. In modern e-commerce marketplaces, the advertisement of false product claims distort buyers' purchasing decisions, leading buyers to make suboptimal consumption choices, or *decision error* (Rao, 2022; Fong et al., 2024; **?**). Similarly, in social media marketplaces, the lack of consequences for the production and amplification of falsehoods materially accelerates the spread of false and misleading information (Vosoughi et al., 2018; Allcott and Gentzkow, 2017; Van Alstyne, 2021; Mazar and Ariely, 2006; Tucker et al., 2018) resulting in harm to not only consumers, *but also honest producers*. The platform X, for example, has been in the limelight for suing departing advertisers on the one hand, while plagued by user complaints about widespread and incessant spam, hate speech, and explicit content on the other[8]. The widespread prevalence of fake ads and review scams on Amazon (He et al., 2022b), genocide-inciting misinformation on Meta and Telegram (Stevenson, 2018; Crystal, 2023) provide a stark example of market failures that result in a failure to improve social welfare for both honest producers and consumers of online information.

2.2 Combating False Advertising

Past research has explored various incentive structures to mitigate these issues, for example, elevating the role of reputation systems in promoting honesty among sellers (Jiao et al., 2021; Luca, 2017). Platforms design various consumer protection with the intention to effectively limit the impact of misleading claims that result in decision error for buyers (Kozyreva et al., 2024; Mehta, 2023). Reducing misleading claims in digital advertisement markets requires innovative solutions that balance user auton-

[8]Kari Paul, "Advertisers axe corporate responsibility scheme after lawsuit from Musk's X," Aug. 8, 2024.

omy with accountability. Traditional approaches to combating misleading claims in such two-sided marketplaces have shown limited efficacy (Saltz et al., 2021; Kozyreva et al., 2024; Green et al., 2023; Tay et al., 2023), with the key gap being the burden of removal of false claims is still placed on on consumers of information or the platform (centralized authority) rather than on the author producing it. Guo found that when confronted with false advertising, buyers tend to rely on preconceived beliefs, and debunking false claims is only effective if done systematically across the market (Guo et al., 2023). Similar findings were reflected in theoretical work within simulated marketplaces (Liu et al., 2012; Zhang and Cohen, 2007). In fact, research has shown that interactive visualizations of reputation data can improve users' ability to detect fraudulent behavior, highlighting the importance of interface design in combating misinformation (Sänger et al., 2016; Hughes et al., 2024; Kutabish et al., 2023; Kim et al., 2016). The design of real-time online games for user research has provided an empirical understanding and replicating the dynamics of online marketplaces through high-fidelity models of digital platforms (Miller et al., 2024; Nacke et al., 2023; Almaatouq et al., 2021a). Despite these efforts, there remains a significant gap in understanding how to design market interventions that not only curb dishonest behavior, but also enhance overall social welfare.

3 Key Contributions

In order to address the limitations of existing marketplace interventions to mitigate misleading claims, we design a novel **truth warrants** mechanism to align economic incentives with truthful behavior while addressing the limitations of existing reputation systems. We develop an interactive online experiment with human participants as consumers in a digital marketplace and test our intervention to demonstrate statistically significant techniques to clear false advertisements from online marketplaces. Our market design framework draws inspiration from economic principles and social computing theories to create a more balanced and effective ecosystem for information exchange (Shen et al., 2012). We raise source incentives to provide honest claims and give recipients a new signal to help them distinguish honest advertised claims from misleading ones. Our interventions, grounded in economic theories, avoid censorship, avoid delays of crowdsourcing, and preserve user autonomy and platform neutrality. By requiring advertisers to stake financial bonds on the accuracy of their claims, warranting introduces an economic disincentive for false advertising. This

approach aligns with behavioral economics principles and complements existing reputation systems by promoting accountability among sellers (Sheng et al., 2010). Our research advances prior art with a novel governance mechanism aimed at realigning incentives in digital marketplaces, advancing economic theories in two-sided markets (Eisenmann et al., 2006). We provide an empirical study testing an intervention that prioritizes consumer welfare, and penalizes dishonest sellers while simultaneously supporting honest sellers such that it improves overall social welfare in the marketplace, called 'truth warrants' (Lin, 2024; Van Alstyne et al., 2023; Alstyne and W, 2021). Our approach addresses the critical issue of market distortion caused by misplaced incentives and contributes to the ongoing discourse on economic models for platform intervention. Our study investigates how governance interventions in digital marketplaces be designed to improve social welfare without limiting individual agency. We explore the question, **can truth warrants provide economic value to drive online sales above traditional rating and reputation systems prevalent in digital marketplaces?**

Since economic value is tied directly to two outcomes in traditional marketplaces– profit for the producers, and volume of sales achieved–we examine both of these in our experiments. We enlist two key hypotheses. These questions are rooted in the broader context of information economics, where the goal is to create innovative market structures that enhance transparency, fairness, and efficiency.

3.0.1 H1: Truth warrants provide an increased economic value compared to traditional reputation systems like seller ratings.

Our arguments underlying **H1** for greater economic value of truth warrants over reputation signals in digital marketplaces center on their ability to directly align incentives, reduce information asymmetry, and impose immediate financial consequences for dishonesty. Warrants act as costly signals requiring sellers to escrow funds proportional to the potential consumer loss, which dishonest producers cannot afford to risk, and therefore, should deter false advertising more effectively than reputation systems. Reputation systems, while useful, impose diffuse, delayed penalties and fail to internalize the externalities of deception, whereas we expect warrants to create a self-enforcing equilibrium where honest sellers thrive and consumers recover losses through challenges.

3.0.2 H2: New sellers in the warrants marketplace without a reputation can use truth warrants to sell products faster

We hypothesize that truth warrants accelerate sales for producers by serving as credible, costly signals that reduce consumer uncertainty and build trust more effectively than reputation systems alone. By requiring producers to escrow funds proportional to potential consumer losses, warrants create immediate financial penalties for dishonesty, deterring false advertising and incentivizing truthful claims. We expect that integrating financial commitments with compliance tools streamlines consumer decisions, reducing purchase hesitation. Thus, warrants should enhance market efficiency by shifting incentives toward honesty, directly accelerating sales velocity while reducing fraud-related friction.

4 The Marketplace Model

4.1 Constructing the Model

In this section, we construct a theoretical marketplace consisting of producers advertising products to consumers, each with their independent incentive structures. Consumers purchase products based on an advertisement displayed to them by a producer, who may have chosen to honestly advertise its quality or dishonestly do so, in an attempt to deceive gullible consumers. As discussed in the introduction, and as a reflection of modern digital marketplaces, our marketplace is equipped with a feedback mechanism that allows consumers to rate their experiences, which, in turn, influences the visibility and success of producers within the market. Our experiment involves the application of a truth warrants *intervention* that introduces accountability to the *production* of misleading claims as an alternative to traditional reputation systems that place responsibility on those affected to discern misleading claims. The marketplace offers a competitive setting for producers to advertise products for sale to consumers in several rounds, with the goal of a producer being to maximize their profit ϕ while that of a consumer being to maximize their utility η gained from the purchase of a product. Correspondingly, for our baseline or control condition, we construct a 'reputation' market which is reflective of the affordances of a modern-day e-commerce marketplace offering a consumer-ratings system that result in the creation of a brand reputation applied to the producer. Producers are represented by a brand and assigned a rating based on 'thumbs up' and 'thumbs down' ratings

offered by consumers post-purchase of a product based on an advertisement by the producer's brand. For the treatment condition, we introduce truth warrants into the reputation market, due to which a producer may optionally elect to escrow an additional amount of money in exchange for a signal in the form of a product label for their advertised claim, thus creating the 'warrant' market. A warranted claim in a product advertisement allows consumers to challenge the producer's claim in order to potentially win the escrowed amount, if it is determined that the warranted claim was false (in this case, a lower product quality than as warranted). The determination of the claim is done by a peer jury Yang (2023); Hua (2022), following the decentralized governance model with truly anonymized voting. In our initial experiments presented in this paper, we replace the peer jury with an "oracle" in the marketplace and adjudicate all challenges with this *theoretically optimal* peer jury. This allows us to evaluate the effects of truth warrants and reputation signals under optimal adjudication conditions. For the e-commerce marketplace that we conduct experiments within and use as a running example for subsequent quantitative analysis, producers are represented by advertisers that are selling a product (termed "sellers") while consumers are represented by users that acquire products in exchange for money they pay to a seller (termed "buyers").

4.2 Platform Economics

The economics of the platform involves products at two qualities: high and low. At each quality level, there is a production cost, selling price, and value gained from its purchase. There are several rounds of sales wherein sellers determine the true quality of the product to produce, advertise it[9].

1. **Production Costs**:
    - c_H: Cost to produce high-quality goods
    - c_L: Cost to produce low-quality goods ($c_H > c_L$)
2. **Consumer Valuations**:
    - v_H: Value of high-quality goods to consumers

[9]by default, all advertisements present a product claim to be of high quality, as previously discussed. No sensible advertisement would claim a product to be explicitly and unironically of "low quality" to a buyer so we can safely impose this design constraint on the experiment without loss of generality

- v_L: Value of low-quality goods to consumers ($v_H > v_L$)

3. **Prices**:

   - s_H: Selling price for high-quality goods
   - s_L: Selling price for low-quality goods ($s_H > s_L$)

4. **Warrant Parameters**:

   - w: Escrowed warrant amount
   - α: Fraction of w paid by consumers to challenge claims ($0 < \alpha < 1$)
   - p: Probability a dishonest producer is challenged and loses w (assumed $p = 1$ if challenges always succeed)

In the control condition without warrants, the consumer is shown a set of 7 products, in which they can only see the product, it's price, the seller's brand name, and seller's reputation (number of thumbs up, and number of thumbs down). In the treatment condition with warrants, the consumer is shown a set of 7 products, in which they can see not only the product, its price, the seller's brand name, reputation, and whether a label indicating that the product claim is warranted, but also a history of prior warrants issued by a producer for their advertised claims and a count of the number of challenges that they lost to buyers of their products. A producer can choose to produce a higher or lower quality product and a consumer can choose to make up to 3 product purchases in any given round based on the visible advertisements and the availability of capital in their wallet (which is replenished each round, allowing them at least 2 and a maximum of 3 purchases). All advertisements make the claim that the product is high quality and in all aesthetic regard, all product advertisements look the same, so a consumer may be misled by an advertisement for a low-quality product given that it is not materially different from the advertisement for a high quality product, beyond the seller brand name, seller reputation. This is where we hypothesize they may rely on credibility signals such as reputation, history of warranting products, and the advertised price of the product. [10]

4.3 Pricing Truth Warrants

In this section, we conduct the following analysis of the model described in Section 4.1.

[10] A caveat to note is that similar to e-commerce marketplaces like Amazon and eBay, sellers in our market can reset their reputation to 0 by electing to switch their brand in each round–with the cost being the loss of all reputation and warrant history for the seller in the marketplace.

4.4 Producer Profit from Dishonesty

A dishonest producer selling low-quality goods as high-quality earns:

$$\phi_{\text{dishonest}} = s_H - c_L$$

To deter dishonesty, the net profit after losing the warrant must be negative:

$$\phi_{\text{net}} = \phi_{\text{dishonest}} - p \cdot w < 0$$

Rearranging for w:

$$w > \frac{s_H - c_L}{p}$$

If challenges always succeed as in our case with the *optimal peer jury* ($p = 1$):

$$w > s_H - c_L \quad \Rightarrow \quad \boxed{w_{\min} = s_H - c_L + \epsilon}$$

where $\epsilon > 0$ ensures strict inequality.

4.5 Consumer Utility

When cheated, a consumer's utility is:

$$\eta_{\text{cheated}} = v_L - s_H$$

After challenging (cost = αw) and winning w:

$$\eta_{\text{challenge}} = (v_L - s_H) - \alpha w + w = v_L - s_H + w(1 - \alpha)$$

For challenges to be rational, consumers must make more money than the loss in utility that they suffered through falsely warranted product claims that deceived them into making a low quality product purchase:

$$\eta_{\text{challenge}} > \eta_{\text{cheated}} \quad \Rightarrow \quad w(1 - \alpha) > 0$$

This holds if $w > 0$ and $\alpha < 1$.

4.6 Platform Incentives

The platform retains challenge fees (αw) and ensures market efficiency.

Note: Digital platforms are run by leaders that are obligated to *maximize shareholder value*[11]. Often, theory fails practice because academic researchers fail to acknowledge that there are practical considerations that platforms make when considering the deployment of an intervention. There is a need to shift the current focus in how we think about evaluating interventions Tay et al. (2023), and beyond this, in their design and deployment preceding evaluation in order to ensure that we incentivize the adoption of the intervention, and therein avoid the traditional "cold start" problem. Academics often state that there needs to be the provision of external stimuli (such as regulatory compliance) in order to *arm-twist* platforms into changing their questionable online safety practices, and while that may be true, it designates the entirety of responsibility onto a policymaking apparatus that trails most platform teams in innovation, thereby rendering its controls limited barring egregious harms. Keeping with our stated academic responsibility to design the incentive structure of this intervention to align with its adoption by platforms, we attempt to make the design of the economics underlying truth warrants a nominally financially rewarding outcome for platforms, in addition to the benefits it provides for consumer protection. In this way, we kill two birds with one stone, incentivizing the adoption of this technology intervention at scale, while simultaneously improving consumer protection for online users.

4.6.1 Minimum Warrant Price for Market Efficiency

Theorem 4.1. *: To ensure dishonest producers face net losses when caught, the warrant must satisfy:*

$$w > s_H - c_L$$

Proof:

1. Dishonest Profit: $\phi_{dishonest} = s_H - c_L$.
2. Net Profit After Warrant Loss: $\phi_{net} = (s_H - c_L) - w$.
3. For $\phi_{net} < 0$:

$$s_H - c_L - w < 0 \quad \Rightarrow \quad w > s_H - c_L$$

Consumer Challenge Rationality: Challenges occur if $w(1-\alpha) > 0$, which holds for $\alpha < 1$.

[11] or at least seem to provide this as a justification when questioned by Congress

4.6.2 Implications

1. Warrant Floor: $w_{min} = s_H - c_L$ ensures dishonesty is unprofitable.
2. Consumer Protection: Challenges recover w, offsetting losses ($v_L - s_H + w(1 - \alpha)$).
3. Platform Stability: Fees (αw) fund adjudication.

5 Virtual Marketplace Experiment

We investigate the effects of warrant in online advertising markets using a reproducible virtual game framework (Almaatouq et al., 2021b). In particular we employ a gamified model of a two-sided market comprising sellers and buyers of commercial experience goods (i.e., products whose qualities can only be fully assessed after purchase, through use or consumption). The sellers aim to maximize profits by selling products that buyers will purchase, while the buyers seek to maximize the value gained from the purchase of advertised products.

We sample 250 human participants located in the United States, in a preregistered, IRB-approved online experiment, having them play up to 7 rounds of the sales game as buyers, allocating 125 participants randomly to each of the treatment and control arms representing the warrant and reputation market, respectively. The results presented reflect the gameplay statistics from nearly 5208 rounds of games played in the marketplace that were valid.

Our experiment presents two types of markets, and participants in our experiment are randomly assigned to one of them: the reputation market or the warrant market. The reputation market reflects the current state of digital platforms like eBay, Facebook Marketplace, or Amazon, where sellers are rated after transactions but buyers cannot be certain they will recover losses when they purchase a falsely advertised or counterfeit product. In this type of market, reputation ratings can influence future purchases, but they do little to penalize misleading sellers in the present.

By contrast, the warrant market (our intervention) offers sellers the voluntary option to signal the truth of their claims by offering a "truth warrant." Warrants are labels attached to product advertisements indicating that the seller has escrowed an amount of money proportional to the value of the buyer's wrong choice i.e. the difference in value between a good and bad product. If a buyer is cheated, after purchasing the product they can challenge the seller's claim, and receive the escrowed

value. Warrants history and adjudication is made public, with buyers able to view how many claims the seller has warranted and how many claims were adjudicated to have been false. This system creates an independent mechanism for buyers to hold sellers accountable and recover damages from falsely advertised products, without relying on the seller's goodwill or future buyer feedback.

The setup of the reputation and warrant markets is otherwise identical, ensuring that the only variable is the presence or absence of warrant. Using warrant is entirely voluntary. In the Warrants Market, buyers can see whether a product is warranted and may challenge false claims after purchase. This introduces a strategic choice for sellers regarding whether to offer a warrant, while buyers must decide whether to trust unwarranted products based on the quality of their brands.

The primary outcome variables in this experiment include, as discussed, the economic value of warrants in terms of the sold product stock, and the round number of the first sale of a product by a seller.

We also examine the volume of false claims and social welfare. The former is defined by the fraction of low quality products advertised as high relative to all ads in the marketplace. The latter is defined by buyer utility and seller profit, which includes the endogenous decision to warrant and the endogenous decision to produce high or low quality. In our model, sellers have higher profit margins when they cheat but only on condition of a sale. Production without sale incurs a loss and high quality production is more costly than low quality production. Buyers benefit by getting the best deal, conditional on not getting cheated. Table 1 gives the explicit values. By comparing the reputation and warrant markets, we aim to evaluate how these mechanisms influence economic value and the time taken to make a successful sale.

5.1 Game Setup

After providing consent, participants proceed to a tutorial that introduces the marketplace interface and decision-making process. Participants are then randomly assigned to the role of either buyer or seller, with each role receiving tailored instructions. Sellers aim to maximize profits by managing product quality, advertising, and reputation, while buyers seek to maximize value by purchasing high-quality products and avoiding deception (see Figure 4 for instructions and an example screen).

Each participant is assigned to play either the seller or buyer role for the duration of the experiment. In addition to human participants, automated bots are included to generate a competitive, dynamic environment. Sellers compete against buyer bots,

while buyers interact with seller bots.

The marketplace operates in 7 sequential rounds, giving participants the opportunity to learn from past actions and adapt their strategies. Each round follows a structured process with each player making their respective choices in an interactive (sequential) fashion; sellers first, then buyers. In seller phase, sellers choose product quality and set the advertisement quality for buyers to view. Sellers also decide if they want to add a warrant to their advertisement by placing an escrow deposit as a credibility signal. In the buyer phase, buyers observe a set of products with key information such as price, seller reputation, and any "warranted" badges and decide to either purchase a product or skip the round. In feedback and brand switching phase, if the true product quality does not match the advertised quality, buyers can challenge the claim (only in the Warrants Market) for a small fee. buyers also rate sellers, and seller reputations are updated accordingly. Sellers then decide whether to switch to a new brand to reset their reputation, particularly if negative feedback is received. Finally, we update the profits for sellers and give feedback to buyers to show whether they were "cheated" and may recover funds via challenges in the warrant market. This round process repeated 7 times for all participants.

Participants are randomly assigned to one of two market types: reputation market and warrant market. In the reputation market, sellers decide at the end of each round whether to maintain their current brand or switch to reset their reputation. buyers rate sellers based on whether the advertised product quality matches the actual production quality. Positive ratings reward honest sellers, while negative ratings penalize those who engage in deception, influencing future buyer decisions.

In the Warrants Market, additional features incentivize honest claims. Sellers have the option to warrant their advertisement claims by placing an escrow amount, signaled to buyers via a "warranted" badge. If a buyer is misled by a warranted claim, they can challenge it for a small fee. Successful challenges result in the buyer recovering the escrowed amount, while honest sellers retain their deposit, reinforcing their credibility. buyers can see which products are warranted during their purchasing decisions, adding another layer of trust to the marketplace. Unlike the reputation market, the Warrants Market allows for proactive mechanisms to hold sellers accountable in real-time rather than relying solely on buyer ratings.

These values define player strategies. Listing profits in increasing order yields: producing low and warranted high ($\pi = 2$), producing high but not warranting ($\pi = 4$), producing high and warranting ($\pi = 6$), producing low and claiming high ($\pi =$

Product Quality	User Value	Price	Prod'n Cost
Low	6	10	2
Low (warranted)	6	12	2 + (14-6)
High	14	10	6
High (warranted)	14	12	6

Table 1. Distribution of experimental values for low and high quality products. Warranting the claim that a low quality product has high quality incurs the cost of decision error based on deceiving the buyer. A challenge renders this a non-dominant strategy compared to other seller strategies.

8). Listing buyer utilities in increasing order is: getting cheated ($u = -4$), buying warranted ($u = 2$ regardless of quality), buying high without paying for the warrant ($u = 4$). If the buyer believes they'll be cheated, they should not buy. If the buyer believes the product is high quality, they should try to buy it but not pay for the warrant. If the buyer is risk averse, and a warrant is available, they should buy the warranted product. The brands market has no warranting, only reputations, in which case payoffs are defined by the two rows without warranting.

5.2 Rating System

The seller rating system is tied to the seller's current brand and serves as a signal to other buyers in the marketplace, helping them assess the trustworthiness of a given seller based on previous buyer experiences. This challenge mechanism enables cheated buyers to recover the costs they incurred for subpar products advertised as high-quality. If the product's claims are accurate, buyers have no reason to challenge it, and products without warranting cannot be challenged at all. In this system, the rational behavior for buyer bots is to always challenge misleading advertisements when cheated, as this maximizes their utility by recovering their losses. Although a buyer choice in the game design, this rating allocation is automated, given that true quality of the product is revealed post-purchase. So if it is a low-quality product, we assume the user would want to rate it poorly, having been cheated by the seller, and paying a higher amount for a subpar product. And if the advertised purchase is a high quality product, the rating for the seller should be higher given the accurate advertisement for it. We focus on the buyer marketplace, and in order to create a meaningful comparison between the two marketplaces—the reputation market and the warranting Market—we simulate the interaction between various types of strategic sellers and buyers.

5.3 Brand Changes

Each of the adaptive bot sellers in the marketplace is represented by a *brand* with a corresponding brand name and brand rating, reflecting the seller's reputation. At the end of every round of sales, a seller has the option to costlessly change their brand. The option to make brand changes costless is intended to make it a dominant choice for any seller with a net negative reputation at the end of a given round of sales.

Modeling brand *changes* in this marketplace is critical to capturing real-world dynamics where dishonest sellers strategically manipulate reputation systems. The ability for producers to switch brands—effectively resetting their reputations—mirrors common fraudulent practices in digital marketplaces, such as sellers creating new accounts after receiving negative reviews. This behavior undermines traditional reputation systems, as it allows bad actors to evade accountability while retaining the benefits of a "clean slate." By incorporating brand changes into the experimental design, the study evaluates whether truth warrants can mitigate this vulnerability.

Brand switching introduces a strategic tension between short-term gains from deception and long-term trust-building. In reputation-only markets, producers face incentives to engage in "reputation mining": build credibility through initial honest behavior, exploit it with deceptive practices, then reset their brand to avoid penalties. This creates a cyclical pattern of fraud reflecting real-world examples of fraud on Amazon[12], that erodes consumer trust systemically, as buyers cannot distinguish between genuinely new sellers and rebranded cheaters. We expect the warrants market to disrupt this cycle by tying financial commitments (escrowed funds) to specific claims rather than seller identities. Even if a producer rebrands, the cost of warranting false claims remains prohibitive, making deception a non-dominant strategy.

The inclusion of brand changes also tests the comparative resilience of warrants versus reputation signals. While reputation systems suffer from "cold-start" problems for new/rebranded sellers, warrants provide an immediate credibility signal independent of historical performance. This is particularly valuable in markets with high seller turnover or frequent rebranding.

[12] *Central District of California | Hacienda Heights Man Admits Bilking Amazon in $1.3 Million Refund Scam and Will Plead Guilty to Federal Fraud Charge | United States Department of Justice* (2022)

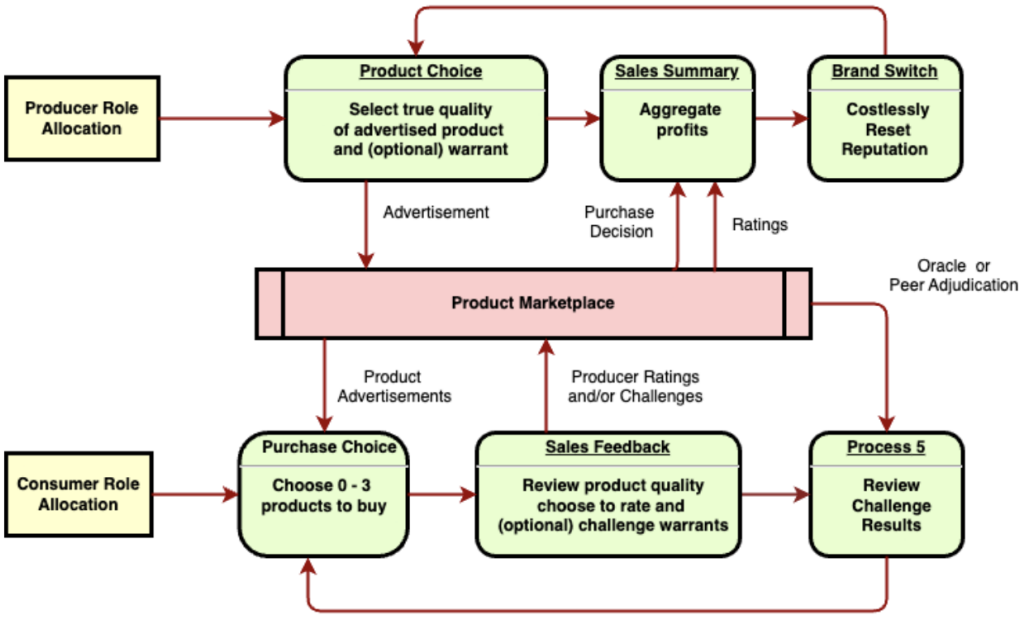

Figure 1. The workflow diagram reflecting the sequence of actions taken by buyers (consumers) and sellers (producers) in this two-sided marketplace (for a single round in the game, repeated across each of 7 rounds).

6 Analysis

We encode the hypotheses and game concepts into our statistical models for estimating the economic value provided by truth warrants as well as the reduction in time to the first sale that a producer makes using truth warrants. For each hypothesis, we enlist the independent variable, dependent variables, and controls in a section corresponding to each below.

6.1 H1: Warrants provide greater economic value than reputation signals

We employ an ordinary least squares regression to estimate the coefficients of the dependent variables provided below.

6.1.1 Dependent Variable

- $Sales_{i,t}$ = Number of units sold for product i in round t

6.1.2 Key Independent Variables

:

1. $Warrant_{i,t} =$

$$\begin{cases} 1 \text{ if product } i \text{ has a warrant in round } t \\ 0 \text{ otherwise} \end{cases}$$

2. $Reputation_{i,t}$ = Producer's net rating (positive negative) at start of round t

6.1.3 Control Variables

- $Price_{i,t}$ = Product price
- $WarrantHistory_{i,t}$ = Number of prior warrants by producer
- $Player_i$ = Player fixed effects
- $Round_t$ = Round fixed-effects for rounds 1–7

We arrive at the corresponding regression formulation to evaluate **H1**.

$$\text{Sales}_{i,t} = \beta_0 + \beta_1 \text{Warrant}_{i,t} + \beta_2 \text{Reputation}_{i,t} + \beta_3 (\text{Warrant}_{i,t} \times \text{Reputation}_{i,t}) + \gamma \cdot \text{Controls} + \epsilon_{i,t}$$

1. Primary Test: $H_0 : \beta_1 = \beta_2$ (Warrant Reputation effect)
2. Alternate: $H_A : \beta_1 > \beta_2$ (Warrant ¿ Reputation)

6.2 H2: Warrants Accelerate Sales for Producers

Let $z_i \in \{1, 2, ..., 7\}$ = Round when producer i made their first sale.

6.2.1 Independent Variables

1. $WarrantApplied_i =$

$$\begin{cases} 1 \text{ if product had warrant when first sold} \\ 0 \text{ otherwise} \end{cases}$$

2. $WarrantHistory_i$ = Number of previous warrants applied by producer

3. $Reputation_i$ = Producer's average rating at start of game

6.2.2 Control Variables

- $Price_i$ = Product price
- $Brand_i$ = Brand fixed effects
- $Round_t$ = Time fixed effects

The regression model we fit to estimate the coefficients that help test **H2** are:

$$z_i = \beta_0 + \beta_1 \text{WarrantApplied}_i + \beta_2 \text{WarrantHistory}_i + \beta_3 \text{Reputation}_i + \gamma \cdot \text{Controls} + \epsilon_i$$

We expect that:

1. Warrant Acceleration: $\beta_1 < 0$ (warranted products sell faster)
2. Warrant vs Reputation: $|\beta_1| > |\beta_3|$ (warrants > reputation in achieving product sales in earlier rounds)

7 Results

We ran a preregistered experimental study recruiting 125 participants per condition (Warrants, Reputation) and had each participant play 7 rounds in the game, resulting in a dataset of 868 samples per condition, and 1736 in total. There were 124 admissible games where seller bots played a total of 5208 rounds in the game. We consider two conditions: Reputation market and Warrants Market where the truth warrant is optionally available to sellers. The availability of warranted advertisement means a product can sell for 20% higher price than it otherwise could. We find evidence supporting the claims for improvement of profits for honest producers as well as the acceleration in the sales of products. We discuss the results in detail below, and offer explanations for their interpretation.

We sample 250 human participants located in the United States, in a preregistered, IRB-approved online experiment, having them play up to 7 rounds of an online

Player Role	Player Instructions
buyer	Your goal is to purchase high-quality products without being cheated. Your utility decreases when you are cheated. You have the option to challenge sellers and attempt to recover money if you've been misled.
	Challenges cost a nominal amount to initiate, regardless of the outcome of the challenge.
seller	Your goal is to maximize your score by generating profits from sales. You may choose to create false advertisements to deceive buyers. You may switch "brands" to reset your reputation.
Doing nothing may seem a safe option, but it will not earn you any points!	

Table 2. Instructions for buyers and sellers

product sales game as buyers, allocating 125 participants randomly to each of the treatment and control arms representing the "Warrants" and "Reputation" market, respectively–with human buyers choosing to buy up to 3 products of 7 in each round as shown in Fig. 4 where they play against 7 automated 'bot' sellers producing a single product each, based on a varied range of strategies shown in 3.

The results presented reflect the gameplay statistics (honest and dishonest sales and avg. profits in Figs. 5 and 6) from nearly 5208 rounds of games played in the marketplace that were valid.

1. **Economic Value of Truth Warrants**: The act of warranting significantly increases sales in the warrants market at nearly double the magnitude ($\mu = 0.31(\sigma = 0.01)$) when compared to ratings ($\mu = 0.1623(\sigma = 0.005)$) Honest advertisements contribute to a significantly higher social welfare even though 'reputation mining' has a significant negative impact on social welfare (Fig. 2).

2. **Time Taken for the First Sale**: Ratings are a significant accelerator of sales, however warranting accelerates sales even more. Sellers that do not warrant in the marketplace with the option to do so take significantly longer to achieve sales; sales are also significantly slower in the control markets without the option to warrant (Fig. 3).

Dep. Variable:	SoldStock	**R-squared:**	0.189
Model:	OLS	**Adj. R-squared:**	0.189
Method:	Least Squares	**F-statistic:**	771.7
Date:	Thu, 27 Feb 2025	**Prob (F-statistic):**	0.00
Time:	03:52:23	**Log-Likelihood:**	-6075.2
No. Observations:	10578	**AIC:**	1.216e+04
Df Residuals:	10572	**BIC:**	1.221e+04
Df Model:	5		
Covariance Type:	cluster		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.3729	0.006	60.839	0.000	0.361	0.385
Warranted[T.True]	0.3088	0.013	24.024	0.000	0.284	0.334
NetRatings	0.1623	0.005	33.506	0.000	0.153	0.172
NetRatings:Warr...	0.0125	0.017	0.738	0.460	-0.021	0.046
Condition	-0.1874	0.010	-18.628	0.000	-0.207	-0.168
NetRatings:Condition	-0.0145	0.016	-0.887	0.375	-0.047	0.018

Omnibus:	1856.406	**Durbin-Watson:**	2.122
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1231.460
Skew:	0.721	**Prob(JB):**	3.91e-268
Kurtosis:	2.155	**Cond. No.**	6.58

Figure 2. OLS Regression Results for H1 about the economic value of warranting for sales in the warrants marketplace. Warranting significantly increases the sales by nearly by double the amount, as compared to ratings provided to sellers in the marketplace.

Dep. Variable:	FirstSale	**R-squared:**	0.109
Model:	OLS	**Adj. R-squared:**	0.109
Method:	Least Squares	**F-statistic:**	684.8
Date:	Thu, 27 Feb 2025	**Prob (F-statistic):**	9.31e-221
Time:	01:40:23	**Log-Likelihood:**	-27851.
No. Observations:	10578	**AIC:**	5.571e+04
Df Residuals:	10575	**BIC:**	5.573e+04
Df Model:	2		
Covariance Type:	cluster		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	5.0828	0.061	83.510	0.000	4.963	5.202
Warranted[T.True]	-3.3471	0.091	-36.768	0.000	-3.526	-3.169
Condition	1.5025	0.090	16.638	0.000	1.325	1.680

Omnibus:	10078.362	**Durbin-Watson:**	1.973
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	735.217
Skew:	0.228	**Prob(JB):**	2.24e-160
Kurtosis:	1.791	**Cond. No.**	3.91

Figure 3. OLS Regression Results for H2 about the time taken for a seller bot's first sale. Note that while the sales slow down in the warrants market, the act of warranting an advertisement achieves a significant reduction in the marketplace. That said, it is important to note that given there is only one human buyer, most products do not get sold in this marketplace.

Seller Bots
The perpetually honest seller that warrants their claims each round.
The perpetual cheat that always produces low-quality products.
The 'goldfish' seller that switches the product quality on receiving a sale, while otherwise not switching the quality.
The 'bait-and-switch' seller that starts with a high quality product and switches true product quality after each successful sale.
The 'politician' that sells honest high-quality products for two rounds before switching to low-quality products until they make one sale, before reverting to their original strategy.
The reformed cheat, that starts with cheating, but converts to perpetually honest on receiving a sale.
The 'honest opportunist', that is honest until the penultimate round, after which they switch quality for the last round.

Table 3. We design seller bots to follow a varying set of patterns in selecting whether to advertise honestly or dishonestly in the digital marketplace in order to achieve their sales goals and profits.

8 Limitations and Future Work

One of the questions we asked is, if profits for sellers are higher in the Warrants Market, then why might buyer utility be lower given the increase in honest production. Upon reviewing details, we do indeed find a lower utility for buyers in the Warrants Market compared to the Reputation Market. It turned out that we kept the capital the same but increased the price for warranted products so the cost of the warranted product ended up being transferred over to the buyer, thereby reducing their overall utility in the game. The volume of sales in the marketplace with warrants are 11.4% lower than in the control, and thus have a lower contribution to social welfare. This is because our design set a price premium of 20% for warranted products while keeping the same capital available to the consumer, limiting their sales capacity by 16.7%

[13]. It is therefore an unexpectedly positive result for actual sales to only reduce 11.4%. In our effort to test buyer willingness to pay for increased certainty, we found a strongly positive result, but this demand for warranted products decreased their net gain owing to the increase in total price. Second, the prospect of winning an extra payment, by claiming the warrant, led a fraction of buyers to challenge warranted claims even when they were true. Challenging true claims reduced their earnings, which never occurs in the Reputation Market. These facts imply that the design successfully punishes attempts to game the system but also that the interface must explain more clearly that dishonest *buyer* behavior is costly. Honesty benefits *both* sides of the market.

9 Conclusion

Warranted claims significantly increase the likelihood of product sales as compared to standard reputation systems. Effect sizes are especially pronounced for higher rated products and honestly advertised claims. We also find evidence that market entry is easier for new products, represented as previously unseen brands without ratings. The more credible signal, is not sbject to "reputation spending," which appears to contribute to willingness to purchase. Further, the total volume of false claims falls in the market where truth warrants are present and offered as a voluntary option. [xxx needs statistical testing!!] Our work bridges economic theory and user interface design, investigating a new method of enhancing the resiliency and trustworthiness of online information environments. We design a novel two-sided marketplace and report on user studies focused on limiting misleading information in the market. By integrating these mechanisms into platform design, we can create more trustworthy and resilient online marketplaces.

References

Akerlof, George A, "The market for "lemons": Quality uncertainty and the market mechanism," *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Allcott, Hunt and Matthew Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, 2017, *31* (2), 211–236.

[13]At a price of 12 instead of 10, consumers can purchase a maximum of 5 instead of 6 products every 2 rounds with a capital of 60, reducing maximum sales by 1/6th

Almaatouq, Abdullah, Joshua Becker, James P Houghton, Nicolas Paton, Duncan J Watts, and Mark E Whiting, "Empirica: a virtual lab for high-throughput macro-level experiments," *Behavior Research Methods*, 2021, *53* (5), 2158–2171.

_ , _ , James P. Houghton, Nicolas Paton, Duncan J. Watts, and Mark E. Whiting, "Empirica: a virtual lab for high-throughput macro-level experiments," *Behavior Research Methods*, 2021, *53* (5), 2158–2171.

Alstyne, Marshall Van, Michael D. Smith, and Ha-Sung Lin, "Improving Section 230, Preserving Democracy, and Protecting Free Speech," *Communications of the ACM*, 2023, *66* (4), 26–28.

Alstyne, Marshall W Van, "Free Speech, Platforms & The Fake News Problem," *ssrn.com (December 31, 2021)*, 2021.

Alstyne, Van and Marshall W, "Free Speech, Platforms & The Fake News Problem," December 2021.

Central District of California | Hacienda Heights Man Admits Bilking Amazon in $1.3 Million Refund Scam and Will Plead Guilty to Federal Fraud Charge | United States Department of Justice

***Central District of California | Hacienda Heights Man Admits Bilking Amazon in $1.3 Million Refund Scam and Will Plead Guilty to Federal Fraud Charge | United States Department of Justice**, March 2022.*

Coase, R. H., *"The Problem of Social Cost," The Journal of Law and Economics, November 2013, 56 (4), 837–877. Publisher: The University of Chicago Press.*

Crystal, Caroline, *"Facebook, Telegram, and the Ongoing Struggle Against Online Hate Speech," September 2023.*

Eisenmann, Thomas R., Geoffrey Parker, and Marshall W. Van Alstyne, *"Strategies for Two Sided Markets," October 2006.*

Fong, Jessica, Tong Guo, and Anita Rao, *"Debunking misinformation about consumer products: Effects on beliefs and purchase behavior," Journal of Marketing Research, 2024, 61 (4), 659–681.*

Green, Yasmin, Andrew Gully, Yoel Roth, Abhishek Roy, Joshua A Tucker, and Alicia Wanless, *"Evidence-based misinformation interventions: Challenges and opportunities for measurement and collaboration," Cargengie Endowment for International Peace. January, 2023, 9.*

Guo, Tong et al., *"Debunking Misinformation About Consumer Products: Effects on Beliefs and Purchase Behavior," Journal of Marketing Research, 2023.*

He, Sherry, Brett Hollenbeck, and Davide Proserpio, *"The market for fake reviews," Marketing Science, 2022, 41 (5), 896–921.*

_ , _ , and _ , *"The Market for Fake Reviews," October 2022.*

Hua, Sha, *"When Online Shoppers Feel Cheated, It's Time to Go to Crab Court," Wall Street Journal, June 2022.*

Hughes, Emelia May, Renee Wang, Prerna Juneja, Tony W Li, Tanushree Mitra, and Amy X Zhang, *"Viblio: Introducing Credibility Signals and Citations to Video-Sharing Platforms," in "Proceedings of the CHI Conference on Human Factors in Computing Systems" 2024, pp. 1–20.*

Jiao, Ruohuang, Wojtek Przepiorka, and Vincent Buskens, *"Reputation effects in peer-to-peer online markets: A meta-analysis," Social Science Research, March 2021, 95, 102522.*

Kenning, Michael P., Ryan Kelly, and Simon L. Jones, *"Supporting Credibility Assessment of News in Social Media using Star Ratings and Alternate Sources," in "Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems" CHI EA '18 Association for Computing Machinery New York, NY, USA April 2018, pp. 1–6.*

Kim, Jennifer G., Ha Kyung Kong, Karrie Karahalios, Wai-Tat Fu, and Hwajung Hong, *"The Power of Collective Endorsements: Credibility Factors in Medical Crowdfunding Campaigns," in "Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems" CHI '16 Association for Computing Machinery New York, NY, USA May 2016, pp. 4538–4549.*

Kozyreva, Anastasia, Philipp Lorenz-Spreen, Stefan M. Herzog, Ullrich K. H. Ecker, Stephan Lewandowsky, Ralph Hertwig, Ayesha Ali, Joe

Bak-Coleman, Sarit Barzilai, Melisa Basol, Adam J. Berinsky, Cornelia Betsch, John Cook, Lisa K. Fazio, Michael Geers, Andrew M. Guess, Haifeng Huang, Horacio Larreguy, Rakoen Maertens, Folco Panizza, Gordon Pennycook, David G. Rand, Steve Rathje, Jason Reifler, Philipp Schmid, Mark Smith, Briony Swire-Thompson, Paula Szewach, Sander van der Linden, and Sam Wineburg, *"Toolbox of individual-level interventions against online misinformation," Nature Human Behaviour, June 2024, 8 (6), 1044–1052. Publisher: Nature Publishing Group.*

Kutabish, Saleh, Ana Maria Soares, and Beatriz Casais, *"The Influence of Online Ratings and Reviews in Consumer Buying Behavior: A Systematic Literature Review," in Rim Jallouli, Mohamed Anis Bach Tobji, Meriam Belkhir, Ana Maria Soares, and Beatriz Casais, eds., Digital Economy. Emerging Technologies and Business Innovation, Springer International Publishing Cham 2023, pp. 113–136.*

Lin, Ha-Sung, *"Towards Implementation of Warrant-Based Content Self-Moderation," Electronic Markets, 2024, 34 (1), 1–12.*

Liu, Yuan, Jie Zhang, and Qin Li, *"Design of an incentive mechanism to promote honesty in e-marketplaces with limited inventory," in "Proceedings of the 14th Annual International Conference on Electronic Commerce" 2012, pp. 54–61.*

Luca, Michael, *"Designing Online Marketplaces: Trust and Reputation Mechanisms," Innovation Policy and the Economy, January 2017, 17, 77–93. Publisher: The University of Chicago Press.*

Mazar, Nina and Dan Ariely, *"Dishonesty in everyday life and its policy implications," Journal of public policy & Marketing, 2006, 25 (1), 117–126.*

Mehta, Swapneel, *"Towards Informed Interventions to Limit the Effects of Misleading Information on Social Networks." PhD dissertation, New York University 2023.*

Miller, Josh Aaron, Kutub Gandhi, Matthew Alexander Whitby, Mehmet Kosa, Seth Cooper, Elisa D. Mekler, and Ioanna Iacovides, *"A Design Framework for Reflective Play," in "Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems" CHI '24 Association for Computing Machinery New York, NY, USA May 2024, pp. 1–21.*

Nacke, Lennart E., Pejman Mirza-Babaei, and Anders Drachen, *"User Experience Design and Research in Games," in "Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems" CHI EA '23 Association for Computing Machinery New York, NY, USA April 2023, pp. 1–3.*

OECD, *"The role of online marketplaces in enhancing consumer protection," Going Digital Toolkit Notes 7 April 2021. Series: Going Digital Toolkit Notes Volume: 7.*

_ , *"The role of online marketplaces in protecting and empowering consumers: Country and business survey findings," OECD Digital Economy Papers 329 July 2022. Series: OECD Digital Economy Papers Volume: 329.*

Pooja, K. and Pallavi Upadhyaya, *"What makes an online review credible? A systematic review of the literature and future research directions," Management Review Quarterly, June 2024, 74 (2), 627–659.*

Rao, Anita, *"Deceptive claims using fake news advertising: The impact on consumers," Journal of Marketing Research, 2022, 59 (3), 534–554.*

Saltz, Emily, Soubhik Barari, Claire Leibowicz, and Claire Wardle, *"Misinformation interventions are common, divisive, and poorly understood," Harvard Kennedy School Misinformation Review, October 2021.*

Shen, Dawei, Marshall Van Alstyne, Andrew Lippman, and Hind Benbya, *"Barter: mechanism design for a market incented wisdom exchange," in "Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work" CSCW '12 Association for Computing Machinery New York, NY, USA February 2012, pp. 275–284.*

Sheng, Steve, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs, *"Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions," in "Proceedings of the SIGCHI Conference on Human Factors in Computing Systems" CHI '10 Association for Computing Machinery New York, NY, USA April 2010, pp. 373–382.*

Spence, Michael, *"Job Market Signaling*," The Quarterly Journal of Economics, August 1973, 87 (3), 355–374.*

Stevenson, Alexandra, *"Facebook Admits It Was Used to Incite Violence in Myanmar - The New York Times," November 2018.*

Stiglitz, Joseph E. and Andrew Weiss, *"Credit Rationing in Markets with Imperfect Information," 1981, 71 (3), 393–410.*

Stiglitz, Joseph E and Andrew Weiss, *"Credit rationing in markets with imperfect information," The American economic review, 1981, 71 (3), 393–410.*

Sänger, Johannes, Norman Hänsch, Brian Glass, Zinaida Benenson, Robert Landwirth, and M. Angela Sasse, *"Look Before You Leap: Improving the Users' Ability to Detect Fraud in Electronic Marketplaces," in "Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems" CHI '16 Association for Computing Machinery New York, NY, USA May 2016, pp. 3870–3882.*

Tay, Li Qian, Stephan Lewandowsky, Mark J Hurlstone, Tim Kurz, and Ullrich KH Ecker, *"A focus shift in the evaluation of misinformation interventions," Harvard Kennedy School Misinformation Review, 2023.*

Tucker, Joshua A, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan, *"Social media, political polarization, and political disinformation: A review of the scientific literature," 2018.*

Vosoughi, Soroush, Deb Roy, and Sinan Aral, *"The spread of true and false news online," Science, March 2018, 359 (6380), 1146–1151. Publisher: American Association for the Advancement of Science.*

Wright, Alyssa, Pattie Maes, and Hiroshi Ishii, *"Social resonance: balancing reputation with tangible design," in "CHI '08 Extended Abstracts on Human Factors in Computing Systems" CHI EA '08 Association for Computing Machinery New York, NY, USA April 2008, pp. 3387–3392.*

Yang, Zeyi, *"Users are doling out justice on a Chinese food delivery app | MIT Technology Review," December 2023.*

Zhang, Jie and Robin Cohen, *"Design of a mechanism for promoting honesty in e-marketplaces," in "PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE," Vol. 22 Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999 2007, p. 1495.*

10 Appendix

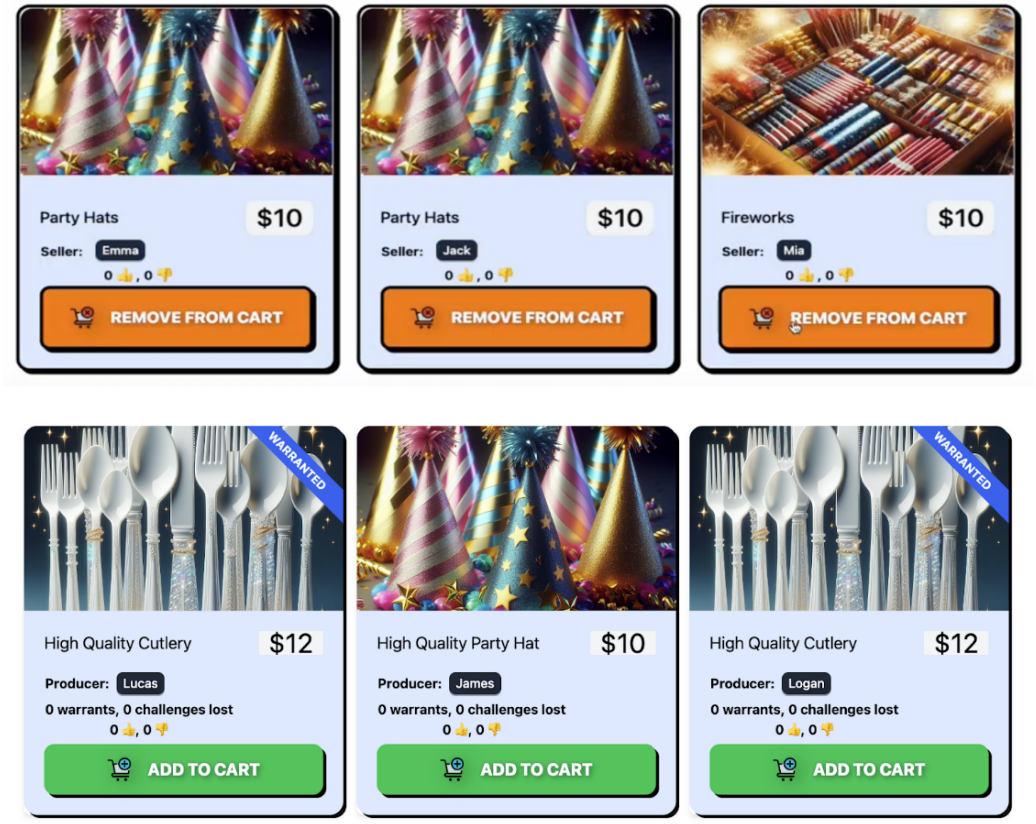

Figure 4. Pictured in the 3-product upper image above we have the control condition (Reputation Market) without truth warrants where the buyer can only see the price (P), Seller's name, and Seller's reputation. In the 3-product lower image we depict the treatment (truth warrant). The buyer is shown a set of 7 products (3 products pictured for brevity), in which they can see not only the product, its price, the Seller name, reputation, and history of warrants and challenges.

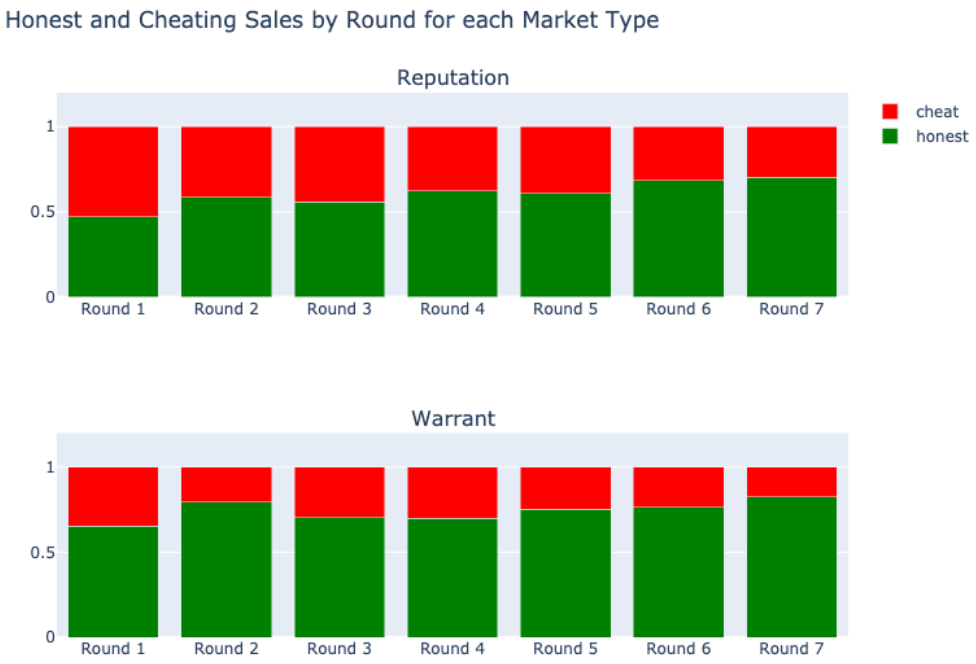

Figure 5. Total sales generated in each round by seller bots in the Warrants (treatment) market and Reputation market (control).

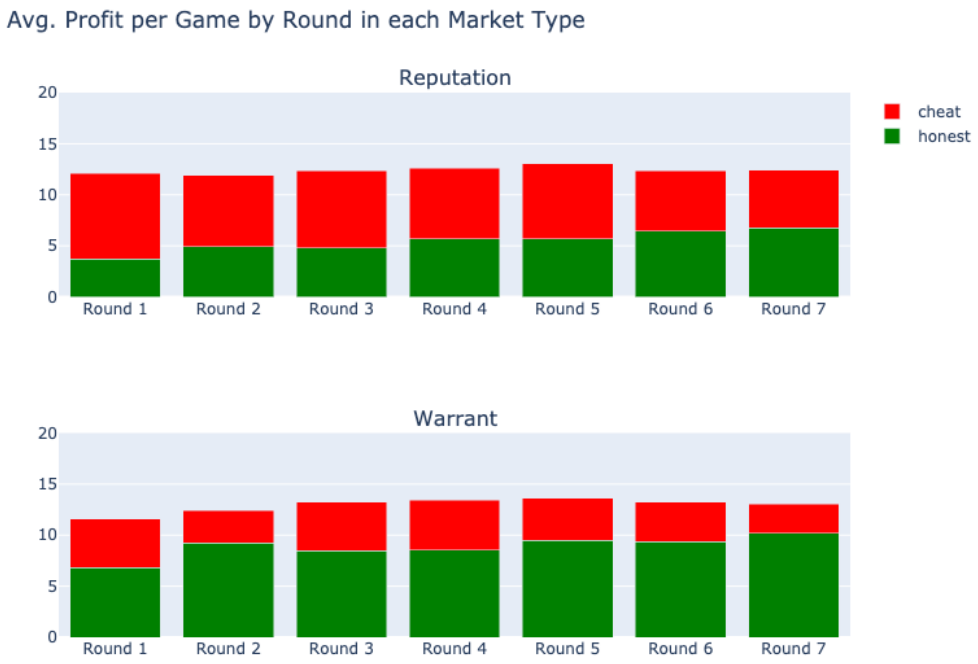

Figure 6. Total profits generated in each round by seller bots in the treatment (warrants) market and control (reputation) market (control).