

Investigating Nudges toward Related Sellers on E-commerce Marketplaces: A Case Study on Amazon

ABHISEK DASH, Max Planck Institute for Software Systems, Germany

ABHIJNAN CHAKRABORTY, Indian Institute of Technology Kharagpur, India

SAPTARSHI GHOSH, Indian Institute of Technology Kharagpur, India

ANIMESH MUKHERJEE, Indian Institute of Technology Kharagpur, India

KRISHNA P. GUMMADI, Max Planck Institute for Software Systems, Germany

E-commerce marketplaces provide business opportunities to millions of sellers worldwide. Some of these sellers have special relationships with the marketplace by virtue of using their subsidiary services (e.g., fulfillment and/or shipping services provided by the marketplace) – we refer to such sellers collectively as *Related Sellers*. When multiple sellers offer to sell the same product, the marketplace helps a customer in selecting an offer (by a seller) through (a) a default offer selection algorithm, (b) showing features about each of the offers and the corresponding sellers (price, seller performance metrics, seller’s number of ratings etc.), and (c) finally evaluating the sellers along these features. In this paper, we perform an end-to-end investigation into how the above apparatus can nudge customers toward the *Related Sellers* on Amazon’s four different marketplaces in India, USA, Germany and France. We find that given explicit choices, customers’ preferred offers and algorithmically selected offers can be significantly different. We highlight that Amazon is adopting different performance metric evaluation policies for different sellers, potentially benefiting *Related Sellers*. For instance, such policies result in notable discrepancy between the actual performance metric and the presented performance metric of *Related Sellers*. We further observe that among the seller-centric features visible to customers, sellers’ number of ratings influences their decisions the most, yet it may not reflect the true quality of service by the seller, rather reflecting the scale at which the seller operates, thereby implicitly steering customers toward larger *Related Sellers*. Moreover, when customers are shown the rectified metrics for the different sellers, their preference toward *Related Sellers* is almost halved. We believe our findings will inform and encourage further deliberation toward more effective governance of such design choices and policies adopted by e-commerce marketplaces.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**; *Empirical studies in HCI*.

Additional Key Words and Phrases: Choice architecture; Nudges; E-commerce marketplace; Related sellers; Algorithmic auditing ; Preferential treatment

ACM Reference Format:

Abhisek Dash, Abhijnan Chakraborty, Saptarshi Ghosh, Animesh Mukherjee, and Krishna P. Gummadi. 2024. Investigating Nudges toward Related Sellers on E-commerce Marketplaces: A Case Study on Amazon. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 455 (November 2024), 31 pages. <https://doi.org/10.1145/3686994>

Authors’ Contact Information: Abhisek Dash, Max Planck Institute for Software Systems, Germany; Abhijnan Chakraborty, Indian Institute of Technology Kharagpur, India; Saptarshi Ghosh, Indian Institute of Technology Kharagpur, India; Animesh Mukherjee, Indian Institute of Technology Kharagpur, India; Krishna P. Gummadi, Max Planck Institute for Software Systems, Germany.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2573-0142/2024/11-ART455
<https://doi.org/10.1145/3686994>

1 Introduction

E-commerce marketplaces like Amazon, Walmart and Alibaba have emerged as essential online platforms providing livelihood to millions of sellers and producers [30]. Some of these sellers may have special relationships with the marketplace organization for using their subsidiary services [12, 31, 35]. For example, sellers can be of three broad types in Amazon marketplace: (a) **Amazon fulfilled**—orders are fulfilled by Amazon; (b) **Amazon shipped**—orders are only delivered by Amazon; (c) **Merchant fulfilled**—orders are fulfilled by the seller and delivered by other third party delivery service provider (e.g., FedEx, DHL etc.). Moreover, Amazon itself can sell products on its marketplace. In countries, where it is not legally allowed, it sells products via **Special Merchants (SMs)** where Amazon has equity stakes [28, 31]. For instance, there are two SMs in Amazon India marketplace (Amazon.in) – *Cloudtail India* and *Appario Retail Private Ltd.* Collectively, we refer to all these sellers (Amazon fulfilled, Amazon shipped, Amazon and its Special Merchants) having special relationships with Amazon marketplace as **Related Sellers**. Table 1 summarizes the different kind of sellers on Amazon marketplace and their special relationship with Amazon.

The inception of these special relationships and the monetary incentive it provides for marketplace platforms have triggered policymakers to look into the anti-competitive prowess of algorithms and policies deployed by e-commerce marketplaces [47–49]. Potential preferential treatment toward sellers using Amazon’s subsidiary services has been at the core of some of these investigations. For instance, in the Antitrust subcommittee hearings of 2019, Amazon was asked if it takes into account which sellers use its fulfillment services in its decision making. Amazon denied saying its algorithms do *not* consider such factors in their decision making [48]. In contrast, the Federal Trade Commission (FTC) of the USA has filed a lawsuit against Amazon in September 2023, where one of the primary arguments is centered around the nuance of preferential treatment of sellers using Amazon’s fulfillment services [17]. While such allegations and claims continue to emerge, to the best of our knowledge, there has been no formal end-to-end study or systematic audit of such practices encompassing the different stakeholders on Amazon or any other e-commerce marketplace. Note that these algorithms and design choices are proprietary and their effects are subtle. Hence, understanding such systems and their effects might require involved treatment of the research questions by breaking them down to different granularities. This paper attempts to bridge the mentioned gap by performing a third-party audit of the different choice architectures deployed on Amazon for potential preferential treatment (if any) toward *Related Sellers*.

1.1 Important choice architectures and research questions

In Amazon, multiple sellers can offer to sell the same product. Consequently, when a customer views a product page, the platform presents a choice architecture to help in deciding which offer (or corresponding seller) to select. ‘Choice architecture’ broadly describes the way different choices are presented to customers in a multi-choice setup [25, 42, 54]. In Amazon, such choice architecture has three different facets (a) default offer selection (Buy Box) algorithm, (b) offer listing page, (c) evaluation of seller performance metrics. Next, we discuss each of these facets and pinpoint the fairness concerns to contextualise the present work.

Buy Box algorithm: On each Amazon product page, one can observe a rectangular box containing two buttons – ‘add to cart’ and ‘buy now’ (see Figure 1(a)). This box is popularly referred to as Amazon Buy Box. When multiple sellers offer the same product, a proprietary algorithm (Amazon Buy Box algorithm) determines which seller should be selected as a default seller i.e., the winner of the Buy Box. Given the prime positioning of the Buy Box (and the prominently displayed buttons inside it), it accounts for more than 80% sales on Amazon [33]. Thus, *winning the Buy Box or becoming the featured offer for a product directly impacts the revenue of the sellers*. Hence,

Seller category	Their relationship with Amazon
Special Merchants	Amazon, the seller in the USA, Germany and France marketplaces. Sellers where Amazon has equity stakes in India, e.g., Cloudtail India and Appario Retail Pvt. Ltd. They use Amazon's fulfillment service for handling their logistics
Amazon fulfilled	Sellers who use Amazon's fulfillment service wherein Amazon picks, packs, and ships orders on behalf of the seller.
Amazon shipped	Sellers who pick, and pack their products on their own; however they use Amazon's delivery services to ship their products.
Merchant fulfilled	Sellers who pick, pack and ship orders on their own or with the help of any other third party service provider, e.g., DHL, FedEx etc. They do not use Amazon's subsidiary services for handling their logistics.

Table 1. Different categories of sellers on Amazon marketplace and the way they handle orders. Collectively, we refer to the sellers in the first three categories as *Related Sellers*.

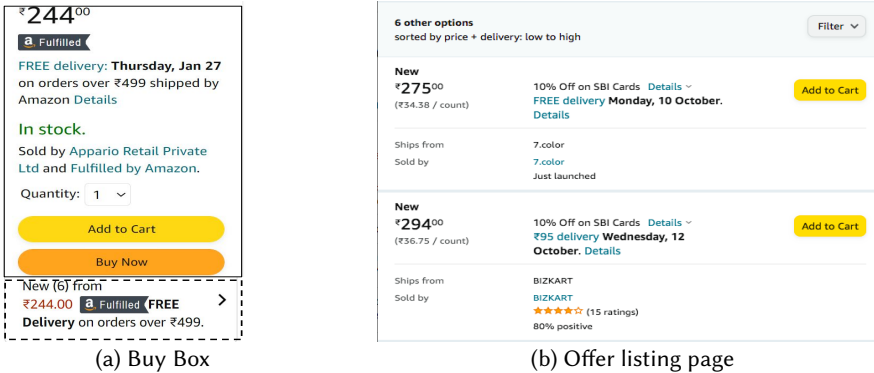


Fig. 1. (a) An example of a Buy Box on Amazon. The small rectangle box (dashed), is the link to redirect customers to a separate 'offer listing page'. (b) The offer listing page, where the offers are sorted as per price + delivery charges.

unsurprisingly, preferential treatment in Buy Box offer selection has been a matter of concern among policymakers across the globe [17, 23, 47–49]. However, Amazon has maintained that its predictions reflect customers' preferences [48]. This leads to the first set of research questions (RQs) studied in this paper:

RQ1 (a): How frequently does Amazon (or a related seller) win the Buy Box when it competes?

RQ1 (b): Do Amazon's algorithmic choices match the customers' choices when they get to decide an offer among the competing offers?

Offer listing page: For every product with multiple interested sellers, Amazon presents a separate offer listing page with all offers listed for the customers to compare (see Figure 1(b)). For each seller, apart from the price and delivery details, different performance metrics such as their average user rating, % of positive feedback, and number of ratings received (#Ratings) are also displayed. These metrics are meant to convey the customer the quality of service a seller provides. However, #Ratings does not necessarily indicate the true quality; rather it reflects the scale at which the corresponding seller operates. A seller that extensively sells products in multiple categories but has

lower quality of service, may well have significantly higher #Ratings as compared to a niche seller providing good quality service operating in less diverse product categories. Thus, showcasing such an un-normalised attribute may steer the customer preference toward **larger** sellers (majority of whom are *Related Sellers*). This deliberation brings us to our second research question:

RQ2: Which information on the offer listing pages influence customers' decision-making the most?

Evaluation of performance metrics: Moreover, platforms may have explicit policies which may put their *Related Sellers* in a position of strength. For example, Amazon has a policy that it can **strike-through** negative feedback received by *Related Sellers* [7]. As per Amazon's policy, a negative feedback concerning delivery experience can be struck through for sellers who opt for its fulfillment or delivery services (more details in Section 3.4). In these cases, such negative feedback does not reflect on their performance metrics. However, these policies are in place only for sellers using Amazon's delivery and fulfillment services i.e., *Related Sellers*. For other sellers (merchant fulfilled sellers), no such policies are in place. Hence for the exact same issue faced by a customer, different sellers on Amazon may be treated differently based on their special relationship with Amazon. This raises concerns of disparate treatment [3] towards the sellers and is also questionable as per the recently enforced Digital Markets Act (DMA) in the European Union [11]. In fact, FTC's new antitrust suite on Amazon focuses on Amazon leveraging its power to reward online sellers who uses its subsidiary services [16]. Given the importance of seller performance metrics in the Buy Box algorithm [1], such misrepresentation may create an uneven playing field for sellers (in favor of *Related Sellers*) by steering customer preference toward them. Hence to quantify and understand the awareness and effect of this review strike-through policy, our final set of RQs are as follows:

RQ3 (a): What are the effects of Amazon's review strike-through policies on seller performance metrics reported on Amazon?

RQ3 (b): How aware are customers about these review strike-through policies and how do they interpret them?

RQ3 (c): What are the effects of Amazon's review strike-through policy on customers' choices?

1.2 Our contributions

In this work, we design a web-scraping and collect the Buy Box information for multiple products and sellers from Amazon's marketplaces in four different countries: Amazon.in (India), Amazon.com (USA), Amazon.de (Germany) and Amazon.fr (France). In total, we collect and analyse data of more than 75,000+ Buy Box competitions across 100 most frequently searched queries on Amazon. Our analyses over the collected data show that Amazon wins a large fraction of the Buy Box competitions where it competes with other sellers. To further understand if this situation truly reflects customers' preferences, we conduct a survey among 200 participants (50 from each of the countries). For a set of 60 products, we show them the offers details (from different sellers) and ask them to give us their preferred choices. We observe that the customer choices elicited in the survey do *not* match with the algorithmic choices of the Buy Box algorithm for the surveyed products in a significant fraction of cases. During the survey, customers also expressed that different seller performance metrics shown on the offer listing pages influence their decisions. To investigate the correctness of the reported seller performance metrics, we collect seller feedback of top-1000 most active sellers on all four marketplaces and conduct an analysis of over 4M seller feedback. Our analyses show that Amazon's review strike-through policy results in significant discrepancy between the reported and actual performance metrics of *Related Sellers*. To further quantify the effect of such policies on customers' decision-making, we repeat the surveys by showing them the rectified metrics. In this counterfactual setting, we observe that preference toward Amazon and its *Related Sellers* further reduces among customers. Our major findings are summarized below:

- **Amazon Buy Box prefer Amazon and its special merchants in the analysed competitions:**

We observe that Amazon and its SMs enjoy a Buy Box win-rate of upwards of 80% in our collected dataset. In particular, for the data collected for USA and India, they also win more than 25% cases, *despite not offering the lowest price for the said product*. Our conducted survey for 60 products across 200 participants consistently points at significant difference between algorithmic decision and customer preferences across countries. While customers prefer Amazon and its SMs in 31% cases, Amazon Buy Box algorithm prefer these sellers for similar competitions in nearly 80% cases across countries (as per the surveyed competitions). Often respondents mention seller performance metrics such as percentage of positive feedback, #Ratings as the most influential features.

- **Amazon's review strike-through impacts seller performance metrics:** Amazon's strike-through policy has serious ramifications on the seller performance metrics of *Related Sellers* across countries. Especially, the performance metrics of Amazon fulfilled sellers, in our dataset, drop by nearly 20% when we include the struck-through reviews in Amazon India. More worryingly, 78% of the respondents, we surveyed, were not even aware of such strike-through policies of Amazon.

- **Seller performance metrics have significant impact on customers' decisions:** Mere removal of #Ratings feature reduces the frequency of customers preferring Amazon SMs by nearly 40%. Furthermore, when customers are shown the rectified metrics (which takes into account the struck through reviews), the first preference votes toward Amazon SMs get almost halved and that for Amazon fulfilled sellers reduce by 16.67%.

Organization of the paper: The rest of the paper is organised as follows. Section 2 discusses some of the prior works in different relevant strands of research. Section 3 elaborates on the main choice architectures in Amazon marketplace and discusses the potential concerning them to further motivate the work. Section 4 gives an overview of the methods adopted in this paper. Sections 5, 6, 7 and 8 elaborate more about each of the methods adopted, along with their insights. Specifically, Section 5 investigates potential preferences towards Related Sellers in the Buy Box predictions, Section 6 compares the Buy Box algorithmic decisions with customer preferences, Section 7 studies the effects of Amazon's policy of strike-through seller feedback, and Section 8 quantifies the impact of seller performance metrics on customer decisions. Finally, we conclude with a discussion summarizing the contributions, limitations, recommendations along with insights for different relevant stakeholders in the e-commerce ecosystem.

2 Related work

To place our work in the perspective of prior research, we present a brief overview of the following lines of research: (a) nudges for influencing customer behavior; and (b) fairness in multi-sided digital marketplaces.

Nudges to influence customer behavior: Nudges are interventions in the presentation of choices to customers that steer them toward a desirable action [54]. The effectiveness of the same has been elaborated in many different contexts ranging from enabling students to eat more healthy food to selecting better future plans in terms of savings [54]. Similarly, nudges have been studied to be effective in steering users' decision making [52] and for enabling equitable donation [42]. At the same time, many prior works have studied how companies exploit customers' cognitive limitations and biases for their profit [20]. What information is available or accessible [27] and how the information is framed [52, 55] plays a significant part in their decision making process. Lately, such market manipulation has been intensified by digital marketplaces [6]. For example, unlike physical stores of yesteryear, digital marketplaces have improved the scale and sophistication of what customer behaviour information they capture and retain and how do they use it [31]. At the same time, they also deploy algorithms trained on such user behaviour to mediate the interaction

Paper	Primary goal	Customer Preference	Seller Opportunity	Platform's Influence	Platform's Policy	Effect
Dash et al. [13]	Investigate bias in sponsored product recommendation toward Amazon Private Label products.	✗	✗	✓	✗	✗
Yin and Jeffries [57]	Investigates bias in product search toward Amazon Private Label products.	✗	✗	✓	✗	✗
Dash et al. [14]	Investigates fairness and interpretability issues of conversational voice search in e-commerce from customers and producers perspective.	✓	✓	✗	✗	✗
Chen et al. [10]	Investigates usage and effect of algorithmic pricing on Amazon Buy Box offer selection.	✗	✓	✗	✗	✓
He et al. [22]	Investigates solicitation and effect of fake product reviews on sales of different products on Amazon.	✗	✓	✗	✓	✓
This work	<i>Investigates existence of preferential treatment toward related sellers on Amazon through different design and policy choices.</i>	✓	✓	✓	✓	✓

Table 2. **Prior audit works on different algorithmic systems of the Amazon marketplace, and their fundamental differences with the current work. Our work incorporates customer preference, sellers' opportunities, effect of the platform and its policies in an end-to-end investigation of potential preferential treatment toward related sellers. The prior works focused on only a subset of these facets in different contexts.**

between different stakeholders on the marketplace [13, 14]. Recently studies have found some inadvertent consequences of such algorithmic mediation in terms of increasing biases [13, 32], spreading misinformation [24, 26] and hate-speech [37, 50] and so on. Similarly, other studies have indicated how certain information e.g., social influence, or ratings / reviews [34, 38, 41] if shown on shopping websites can influence the buying behaviour of customers. To this end, the current study attempts to look into the Amazon offer listing page and how the presentation and calculation of different seller performance metrics have the ability to nudge customers toward different sellers.

Fairness in multi-sided digital marketplaces: In multiple online settings the interests of both the stakeholders are of utmost importance [4, 5]. Considering the two-sided setting, many nuanced algorithms have been developed considering fairness toward both customers and producers [18, 19, 39, 40, 43, 44]. Most of these algorithms generate recommendations by optimizing for multiple objectives pertaining to customer satisfaction, exposure of producers, and sometimes even platform based business objectives. They try to reconcile these objectives under constraints such that several fairness conditions are satisfied.

Present study and prior works on Amazon platform: While the prior works are relevant to our work in broader scheme of things, there have been a few prior audits on the Amazon marketplace, that are most closely related to the current work. We list these works in Table 2. Table 2 shows the primary goals of the individual works (including the current study) in perspective of five different aspects: (a) preference of customers, (b) opportunity of sellers, (c) influence of the platforms, (d) Amazon's policy, and (e) their effects. The study of Dash et al. [13] investigated and proposed methodologies to quantify biases toward Amazon's private label products in sponsored related item recommendations i.e., studied the influence of platform in product recommendations. Along the same lines, Yin and Jeffries [57] studied similar influence and biases in the search infrastructure of Amazon. While both these works investigated the influence of Amazon, both were in the context of Amazon private label products and *not* on the sellers who use Amazon's subsidiary services. Dash et al. [14] performed a study to understand the fairness and interpretability issues in conversational

e-commerce search through Amazon Alexa. This study encompassed customer preferences and seller opportunities. However, this study did not consider special relationships of the entities with the marketplace (i.e., Amazon). Similarly, two prior works discussed seller opportunities and effect of different algorithmic systems or policies on Amazon. Chen et al. [10] also studied Amazon's Buy Box algorithm (similar to what is done in the present study). However, their primary goal was to understand the effects of algorithmic pricing and *not* the question of potential preferential treatment toward *Related Sellers*. Finally, in another recent work He et al. [22], the authors studied the event of solicitation of fake product reviews on Amazon and its effects on sales of the different products. Although our work also considers feedback reviews on Amazon, we consider *seller reviews* as opposed to product reviews in the prior work [22].

Thus, while each of the above-mentioned prior works studied different aspects in the context of different algorithms deployed on Amazon, none of them touched upon all the aspects mentioned above. The current work, on the other hand, tries to investigate potential preferential treatment in Amazon's choice architecture toward *Related Sellers* considering all of the mentioned aspects.

3 Amazon's Choice architecture and related (Un)Fairness concerns

Next, we describe the necessary background about the Amazon platform and its choice architectures.

3.1 Sellers on Amazon marketplace

Amazon has nearly 3 million active sellers across the globe [30]. As mentioned in the introduction, these sellers can be categorized depending on how they handle their delivery logistics. Table 1 summarizes the different kind of sellers active on Amazon marketplaces. Notice that Amazon itself also sells products on its own marketplace in many countries. However, in India, owing to Foreign Direct Investment (FDI) policy of the government, Amazon is not allowed to sell products on Amazon India. Thus, they created joint ventures – Cloudtail India and Appario Retail Private Ltd¹ – to sell products on Amazon while abiding by the rule of the land [28, 45]. As per reports [28], Amazon internally refers to these two sellers as 'Special Merchants'². Collectively, we refer to sellers who have any degree of special relationship with Amazon (i.e., Special Merchants, Amazon Fulfilled and Amazon Shipped) as *Related Sellers*. Note that the degree of special relationships vary from category to category: Special Merchants>>Amazon Fulfilled>>Amazon Shipped.

(Un)Fairness concerns: Since Amazon directly (or through its special merchants in India) sells to customers, it is in direct competition with other third party sellers on its marketplaces. Similarly, Amazon also charges fees from the Amazon fulfilled sellers and Amazon shipped sellers in return of its fulfillment and delivery services [2]. Thus, higher sales for such *Related Sellers* services also provides additional monetary incentive for Amazon to prefer them through its choice architecture. Such vertical integration in the functioning of the marketplace has raised concerns among policymakers across the globe [47–49].

Throughout the remainder of the section, we shall discuss the different facets of Amazon's choice architecture and describe the fairness concerns in their context.

3.2 Amazon Buy Box

Multiple sellers (including Amazon's *Related Sellers*) can offer to sell the same product on Amazon. These sellers compete to win the Buy Box (or become the featured offer) for that product when a

¹While Cloudtail ceased to offer products on Amazon in 2022, Appario remains in the business of selling products on Amazon. Both companies were sellers on Amazon India while the data presented in this article was collected.

²We are including Amazon, the seller, in the same umbrella term for the purpose of this paper. Henceforth, any reference to Special Merchant (SM) would mean Amazon (the seller) in its USA, France and Germany marketplaces, and Cloudtail and Appario in its Indian marketplace.

customer views the product page. In such cases, the Buy Box algorithm selects which seller's offer is displayed on the Buy Box of the product page. Customers typically purchase most products on Amazon through the *Buy Box* or *featured offer* on the Amazon product pages [33]. (see Figure 1(a)).

(Un)Fairness concerns: By selecting the default seller, Buy Box provides sellers the opportunity to enhanced sales and revenue [14]. Thus, if the Buy Box winners are not rotated properly (i.e., if the winner distribution is very skewed), only a particular seller or group of sellers will get the opportunity to sales, and a significant set of other competing sellers for the same product will be deprived of such opportunities. Coming from the customers' side, if a better offer at a lower price or from a seller with better quality of service is available and the Buy Box algorithm selects a different seller by default it may lead the customer to get the same product at a higher price or from a lower quality seller and potentially lead to customer dissatisfaction. To this end, recently, Amazon's Buy Box algorithm has come under intense scrutiny by policymakers across the globe [47–49]. In a recent antitrust subcommittee hearing in the USA, Amazon was asked a number of explicit questions pertaining to the Buy Box algorithm [49]. Amazon stated that it refines the predictions to reflect customer preferences by considering factors such as price, fulfillment speed, seller performance etc. Amazon has also stated that it *does not* consider profitability in the design of the Buy Box algorithm [49].

3.3 Offer listing page on Amazon

While the winner of the Buy Box is shown on the product page, if a customer wants to compare all the offers for a particular product (s)he needs to click a link on the product page which reads 'new (xx) from'. This directs to a web-page where the rest of the offers are listed in an ordered fashion. Figure 1(b) shows an instance an offer listing page. By default, the offers are shown in the increasing order of their (price + delivery). Apart from showing the seller name, their price offer and delivery options this page also shows several seller performance metrics, e.g., average user ratings, % of positive feedback over past 12 months, lifetime number of ratings of the sellers etc.

(Un)Fairness concerns: Although some of the metrics are normalised and quantify the quality of service, features such as number of ratings (#Ratings) are un-normalised and reflect the number of feedback ratings received by the corresponding seller in its entire lifetime on Amazon. Implicitly, #Ratings feature (the way it is being used) is biased toward sellers who operate at larger scale, i.e., sellers who sell multitude of products across multiple categories. Hence showing such a feature may steer customers away from small-scale sellers who sell products from a specific category.

3.4 Seller performance metrics on Amazon

The aforementioned performance metrics are evaluated from the seller feedback received by the corresponding sellers. Each of the seller feedback comes with a rating and an associated review. Any feedback with ratings 4 stars or above is a positive feedback; feedback with 2 stars or below is a negative feedback while feedback with 3 stars is considered to be neutral [56]. Thus % of positive feedback is the ratio of number of feedback with 4 stars or above and total number of feedback. Usually Amazon shows the % of positive feedback of the past 12 months for active sellers with more than 10 buyer feedback in the same time interval. Average user rating is then evaluated as the mean of all the ratings received by the seller. Along with these normalised metrics, Amazon also shows the un-normalised lifetime number of ratings received by sellers across all categories [56].

Amazon, the seller, has no associated metrics: The interesting thing to note here is that such seller performance metrics are only available for third-party sellers on Amazon. Amazon, the seller does not have any dedicated seller profile page, nor does it have any list of feedback as shown in

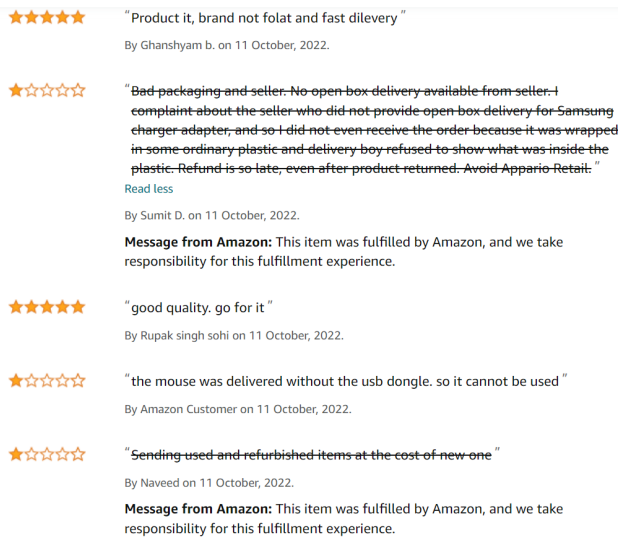


Fig. 2. Example of review striking through: the second and the last reviews have been struck through. A message from Amazon is mentioned taking responsibility of the issue. The percentage of positive feedback drops from 67% to 40% if we include the struck through reviews in the evaluation.

Figure 2. In Amazon’s USA, German and France marketplaces when Amazon competes with other third party sellers, customers do not get to see its seller performance metrics.

Amazon’s feedback strike-through policy Amazon gives this opportunity to some Related sellers that if they request to strike-through a seller feedback, Amazon may oblige. For example, Amazon may strike-through a seller feedback if “*the entire comment relates explicitly to delivery experience for an order fulfilled by Amazon (FBA)*” or “*the entire comment is related to a delayed or undelivered order, which you shipped on time by using Buy Shipping*” [7]. In both cases, in addition to the strike-through, Amazon puts up a message taking responsibility for the inconvenience faced by the customer. Furthermore, Amazon also states in its FBA advertisements that such negative reviews (which have been struck through) will **not** reflect on sellers’ performance metrics [8, 9].

Figure 2 shows examples of struck-through reviews where the second and the last feedback have been struck-through by Amazon. While the second feedback mentions packaging and refund issues, the last one complains about used product being delivered. Notice, if we disregard those two feedback completely (which is what Amazon does for related sellers), the seller has a 67% positive feedback (2 feedback with 4 stars or above out of 3). However, if we consider all the reviews (i.e., now 2 out of 5 are positive), the percentage of positive feedback drops to 40% (reduction of 27%).

(Un)Fairness concerns due to special policy for Related Sellers: The strike-through policy is different for different sellers based on the special relationship that they share with Amazon and is unavailable for independent sellers (e.g., merchant fulfilled sellers). Hence, for the same customer inconvenience, sellers from different categories may be treated differently by Amazon. Given that a struck through review does not affect the performance metrics of the sellers, this puts Related Sellers in a beneficial position. If we disentangle Amazon – the e-commerce platform from Amazon – the fulfillment service provider, then, Amazon – the platform should not put differential policies to ensure fair competition in the marketplace. Such policies not only violates the doctrine of *disparate treatment* [3] but also locks horn with some of the suggestions that the recently agreed upon Digital Marketing Act has (Article 6.5 of DMA) [11] which asks gatekeeper platforms (such as

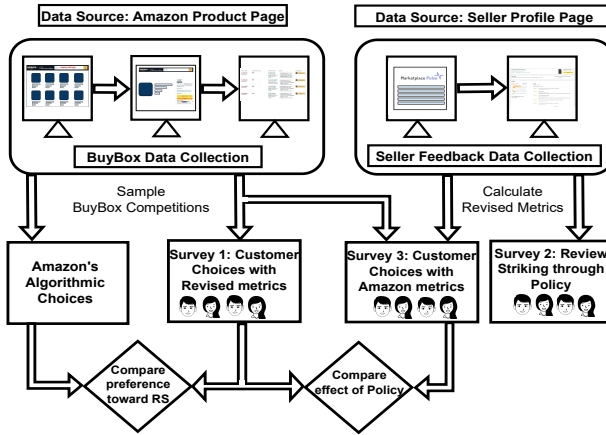


Fig. 3. Block diagram summarizing the methodology adopted in this work. RS: Related Sellers.

Amazon) to refrain from favorably treating products and services provided by themselves. It states, “The gatekeeper shall not treat more favourably, in ranking and related indexing and crawling, services and products offered by the gatekeeper itself than similar services or products of a third party.” In the USA too, the FTC plans to bring up a new antitrust suite on Amazon for leveraging its power to reward online sellers who uses Amazon’s subsidiary services [16].

Such different policies for evaluation of seller performance metric essentially makes customers vulnerable to select sellers with lower quality of service because they are being shown the misrepresented performance metrics of sellers (see the example in Section 3.4). Further, if such negative feedback do not reflect on the performance metric of sellers on Amazon then that affects the accountability and the quality of service to be provided to customers. This may also start a feedback loop wherein Amazon Buy Box algorithm may potentially select *Related Sellers* who take services from Amazon – the fulfillment service provider. In turn, Amazon – the platform may strike-through the negative feedback such sellers may end up having; this in turn will boost their overall performance metric. Given Amazon Buy Box algorithm also gives importance to seller performance metrics [1], these sellers will have higher likelihood of being selected as default sellers and so on.

However, amid all these concerns, Amazon has consistently denied the presence of any systemic preferential treatments toward itself and its *Related Sellers* [29, 48, 49]. *The lack of any systematic investigation of such design choices makes it difficult to detect evidences for such preferential treatment (if any). The methods developed in this work can be viewed as a first important step to bridge this gap.*

4 An overview of the research methodology

Figure 3 shows an overview of the methodology adopted in this work to investigate for potential preferential treatment toward Amazon’s *Related Sellers*. The primary steps in the process are – (I) Buy Box data collection, (II) survey 1: customer choices with Amazon metrics, (III) seller feedback data collection, (IV) survey 2: review strike-through policy, and (V) survey 3: customer choices with rectified metrics. In the rest of this section, we elaborate each of the steps.

Step I: Buy Box data collection: To begin our analyses, we collect data from Amazon marketplaces functioning in four different countries: India (Amazon.in), USA (Amazon.com), Germany (Amazon.de), and France (Amazon.fr). Our selection of these marketplaces is motivated by the recent popular antitrust investigations that Amazon faced in each of these countries and the fact

that they are among the biggest and emerging marketplaces of Amazon in the respective continents [28, 47–49, 51]. First, we conduct a product search on an Amazon platform, then visit the product page of each search result and collect the Buy Box competition data if multiple sellers offer to sell the same product. Since Buy Box winners tend to change with time, we collect data for different Buy Box competitions over a period of two weeks. In the first step, we shall look into these competitions and see how frequently does Amazon (or its SMs in India) win Buy Box competitions when it competes for them.

Step II: Survey 1: Customer choices with Amazon metrics: Even if Amazon and its SMs win a larger fraction of Buy Box competitions, one argument can be that the algorithmic decisions are reflective of the customers' preferences. In fact, this argument has been a common response of Amazon across different antitrust hearings (that their algorithmic decisions are based on customer preferences) [48]. To verify if there is any inherent preference of customers toward Amazon and its special merchants, we conduct a customer survey over a sample of competitions drawn from our collected dataset. We get the customer choices on which seller/offer they would prefer for these competitions. Then we compare the customer choices elicited from this survey with the frequency of times those Buy Boxes were actually won by Amazon and its special merchants in our collected dataset (over a period of two weeks). If the customer choices and algorithmic choices are not similar to each other, then we can conclude that Amazon's algorithmic decisions do not necessarily reflect the customer preferences. During the same survey, we also ask the respondents regarding which features of the different offers (i.e., their sellers) influences their decision the most.

Step III: Seller feedback data collection: Next, we check how the Amazon policies affect the sellers. Seller performance metrics are important in Amazon Buy Box algorithmic decision-making [1]. Similarly, given the amount of seller performance metrics shown on offer listing pages, these metrics can affect customers' decision making as well. However, Amazon's review strike-through policy can play a pivotal role in this context, as it is different for different type of sellers based on how they handle their delivery logistics. To understand how review strike-through can affect the seller performance metrics, we collect the seller feedback garnered by 1,000 active sellers on all the four Amazon marketplaces stated above. We evaluate different seller performance metrics for the corresponding sellers from their seller feedback data, and compare it with their Amazon reported metrics for the past one year to quantify the effect of Amazon's policies.

Step IV: Survey 2: Amazon's strike-through policy: While the method described in Step III brings out the effect of the review strike-through policy on the metrics, it does not shed light on the knowledge of customers (who use these metrics for their decision making) about such policies. To this end, we conduct another customer survey among respondents to understand their knowledge and interpretation of Amazon's strike-through policies. In this survey, we particularly evaluate Amazon's strike-through policy along three dimensions from the perspective of customers: (a) transparency of the policy, (b) interpretability of the policy, and (c) fairness of the policy.

Step V: Survey 3: Customer choices with rectified metrics: Although the survey in Step I may bring out what features influence customers' decision making according to their self-reported explanations, it is important to quantify the exact effects. For example, if in Step III, a significant discrepancy is observed between Amazon's reported metrics and the re-evaluated or rectified seller performance metrics, then there can be some potential effects of such discrepancies on customer decision making as well. Therefore, we conduct surveys similar to Step I by putting customers in a *counterfactual setup*, e.g., where #Ratings is not shown or where seller performance metrics are shown including the struck-through reviews (rectified performance metrics for brevity).

Thus, through empirical observations from the Amazon Buy Box competitions and customer surveys at each step, this work aims to bring out preferential treatment (if any) toward Amazon

Country	# Competitions	# Sellers
India (Amazon.in)	20,553	2766
USA (Amazon.com)	12,864	2577
Germany (Amazon.de)	14,784	1695
France (Amazon.fr)	29,237	2337

Table 3. Statistics of the dataset collected for buy box competitions across Amazon’s different marketplaces. For each country, we state the number of number of competitions ($\langle \text{product}, \text{timestamp} \rangle$ pairs), and the total number of distinct sellers competing in the different competitions.

Related Sellers due to Amazon’s choice presentation apparatus. In the remainder of this paper, we shall discuss each of the steps with greater details followed by the insights from each of the steps.

5 RQ1 (a): Preference toward *Related Sellers* in Buy Box predictions

We start by looking at the outcomes of Buy Box algorithm and their preference (if any) toward *Related Sellers*. In the beginning, we focus on the preference toward sellers with the strongest special relationship, i.e., *Special Merchants*: which refers to Cloudtail and Appario in the context of Amazon India and Amazon, the seller in all its other marketplaces. Our motivation stems from the concerns raised in the US antitrust hearings where policymakers are concerned about preferential treatment of Amazon toward itself as a seller [48, 49]. During the hearings, Amazon was categorically asked if its Buy Box algorithm prefers itself over other third party sellers and Amazon had denied having any preferential treatment in reply.

5.1 Dataset gathered for Buy Box competitions

As a proof of concept, we collect data from Amazon marketplaces functioning in four different countries: India (Amazon.in), USA (Amazon.com), Germany (Amazon.de), and France (Amazon.fr). Given the popularity of product search on e-commerce platforms [53], we proceed to collect our dataset by navigating the product pages following the links from the search results. To this end, we searched a set of 100 distinct most searched keywords (queries) on Amazon [21] and collected buy box competition data for a period of two weeks across all the four Amazon marketplaces. The data collection methodology is elaborated in Appendix A. Upon searching a query on Amazon, a list of product offers are shown on multiple result pages (SERPs), with several results per page. When you click on the products, Amazon redirects customer to the product detail page. We consider the seller who features on the product details page by default to be the Buy Box winner. As Buy Box winners change with time, we make it a point to *repeatedly* visit product pages at different timestamps. Upon reaching a product page, we collect product metadata, e.g., average user ratings, number of reviews, title, brand, the seller that won the Buy Box during the crawl (as was shown in Figure 1(a)). If the link to the ‘offers listing page’ exists, i.e., if multiple sellers are offering the product, we further simulate an automated click on the link and collect metadata for all offers (as shown in Figure 1(b)). For each offer on the offer listing page, we collect their price offered, delivery charges (if specified), seller performance (avg. rating, % of positive feedback, etc.), delivery options (FBA or not), etc.

Competition: For many products on Amazon (also in our dataset), more than one seller may not be present and therefore there is *no* competition for the Buy Box. For the purpose of investigating the Buy Box algorithm, we only consider instances where there exist multiple sellers competing for the Buy Box. We call such instances *competitions*. A competition is a $\langle \text{product}, \text{timestamp} \rangle$ pair, i.e., we consider each instance of visiting a particular product page as a different competition. We

Country	% Buy Box won	% Buy Box won OOP
India	87.74%	29.15%
USA	86.12%	26.39%
Germany	85.35%	13.91%
France	79.98%	12.45%

Table 4. Buy Box win rate of Amazon and SMs in competitions across different countries and % of out-of-position Buy Box won in our collected data. We observe that SMs won 80% or more competitions wherein they competed for the buy box. In India and USA, more than 25% of the buy boxes were won by Amazon when they were not offering the product at the lowest price either. The percentage is almost halved in the German and France marketplaces.

analyse over 70K competitions on Amazon over a period of 2 weeks. Table 3 states the list countries, and for each country, the number of competitions and the number of sellers competing in them.

5.2 How frequently do SMs win Buy Boxes?

When we inspect the competitions where Amazon’s special merchants compete in our dataset, we observe that a huge majority of the Buy Box competitions are eventually won by the Amazon SMs. The Amazon SMs win more than 80% of all competitions where they are involved in our dataset (Table 4). This observation is consistent across all the four Amazon marketplaces we analyzed. Moreover, we observe that 11%, 13.22%, 17.74%, and 26.23% of other sellers competing with Amazon SMs get to win at least one Buy Box competition throughout our data collection period in the India, USA, Germany and France marketplaces respectively. In other words, as per the competitions analysed, Buy Box winners are more distributed in Germany and France marketplaces than in the Indian and USA marketplaces.

A natural hypothesis to explain the above observation is that the Amazon SMs may be providing better offers than the other sellers, e.g., they may be selling the products at lower prices, and hence they are winning most of the Buy Boxes. However, we observe that in many cases, the SMs winning the Buy Box do *not* offer the best (price + delivery charges) as mentioned in the offer listing page. We consider such instances as *out-of-position (OOP)* Buy Box wins. Readers can refer to Appendix B for an example of out of position Buy Box win.

Amazon SMs have a large fraction of out-of-position Buy Box wins: We analysed what percentage of Buy Box wins of Amazon SMs are potentially out-of-position, i.e., where SMs are winning the Buy Box even though they are not offering the best offer. The break-up of out-of-position Buy Box wins are shown in the last column of Table 4. In our collected dataset, we observe that in India and USA, the Amazon SMs win out of position in more than 25% of the times they win the competition. However, the percentages reduce to almost half of it when it comes to European marketplaces such as Germany and France. Although quantitatively the percentages are within 30%, but considering the very high win rate of the Amazon SMs even such percentages account for a significant number of Buy Box competitions. Given the Buy Box algorithm almost always selects a default, and the effectiveness of such default selections in nudging customers toward the selected seller, it is important to think through such algorithmic decisions thoroughly. On one hand, such outcomes can be potentially unfair to the *competing sellers* who are offering the product at competitive prices and depend on Amazon’s algorithms for their revenue. On the other hand, it can be potentially unfair to the *customers* especially if the same product was purchasable from a different seller at a better price.

Takeaways from this section: The primary takeaways from this section are as follows:

- We observe that Amazon and its special merchants enjoy a Buy Box win-rate of upwards of 80% whenever they compete for the Buy Box (in our collected dataset).
- In particular, for the data collected for USA and India, they also win more than 25% cases, *despite not offering the lowest price for the said product.*

6 RQ1 (b): Do customers prefer Amazon SM over other sellers?

Even with the results presented earlier in this section, a question / counter-argument remains – *what if the customers on Amazon have an implicit preference toward Amazon SMs, and the Amazon Buy Box algorithm is replicating that preference as per the feedback that it is receiving from customers’ behavior?* In other words, the choice of Buy Box algorithm should match with (or deviate as less as possible from) the decisions of customers. If it does, then even if the choices are in favor of SMs even though they do not offer the lowest price (or any related seller for that matter), there may not be any existence of preferential treatment in the decision making.

6.1 Survey setup

For conducting these surveys, we chose 200 participants from India, USA, Germany and France (50 from each) who have a high approval rate ($\geq 98\%$) on Prolific crowdsourcing platform (<https://prolific.co/>). Our motivation for a general set of respondents stems from the fact that anyone can purchase and/or potentially sell on Amazon, and hence can potentially be affected by the algorithms and design choices of the platform. Each respondent was remunerated at a rate of 9 GBP per hour which is recommended by Prolific to be a good and ethical rate [46]. We also selected a balanced gender break up for the survey resulting in 49% female and 51% male participants taking the survey. During our survey, 81% of the participants reported to be regular shoppers on Amazon.

We randomly selected 60 products (15 from each marketplaces) where even though Amazon SMs did not offer the lowest price, yet they ended up winning the Buy Box for these products in the first iteration of the data collection. For each such product, we showed the respondents the top-4 offers shown on the Amazon’s offer listing page for that product. In case the Amazon SM that won the Buy Box is not within the top-4 offers, we showed top-3 along with the Amazon SM. For each of the four selected offers, we show all the features that are shown on Amazon’s offer listing page – name of the seller, price offered (including delivery charges, if any), average user rating of the seller, percentage of positive feedback of the seller, number of ratings received, and delivery option (FBA or not). Note that for ‘Amazon, the seller’ in its USA, German and France marketplaces, none of the performance metrics e.g., average rating, percentage of positive feedback or number of ratings etc., are available and Amazon itself does not show the performance metrics for ‘Amazon, the seller’. Hence, we leave those places blank for ‘Amazon, the seller’ in the survey as well.

Given the four offers, we asked the respondents the following questions:
“Suppose you are willing to buy a ‘<product name>’ on Amazon and these are the offers from different sellers for the same product. Which one would you prefer to buy from?”

Further to understand which feature(s) is(are) important for the customers to make their decision, we also asked them the following: *“Briefly reason about the order of your preferences.”*

We pose the first question as a multiple choice grid such that each respondent could provide an ordered preference for the four offers. We repeated this exercise for 15 different products (where Amazon SMs won the Buy Box out-of-position) from each of Amazon’s four marketplaces and the competition (set of 4 offers) for each product was evaluated by 50 respondents resulting in a total of 3000 evaluations.

Observations: For a particular respondent and a particular product, we consider a first preference vote (i.e., rank 1 vote) to be the offer chosen by the respondent for the said product. In other words,

had there been no algorithmic system and the customers were to select the offers themselves, a first preference vote essentially means a vote of willingness to buy from that particular seller. Out of the 3000 evaluations, in only 31% cases respondents chose the offer from Amazon SMs as the first preference. In contrast, if we consider the winner of the Buy Box for the 60 products surveyed at different time points that we observed during our data collection, more than 80% of the times the Buy Box had been won by the Amazon SMs. Thus, there is a big difference between what customers prefer, and what the Amazon Buy Box algorithm selects, with respect to Amazon SMs.

6.2 Observations segregated by countries

While the above observation itself is striking, it is important to study the trends segregated across countries because of the different ways Amazon might operate in different countries, and the different ways customers of different countries interact with the marketplace.

India: Out of the 750 evaluations done by customers of India, 407 (i.e., nearly 54%) times the Amazon Special Merchants (i.e., Cloudtail or Appario) were given the first preference votes. However, out of all the competitions in our dataset, Amazon SMs won the Buy Box in nearly 90% cases. Figure 4(a) shows, for each of the 15 products included in the survey (along x -axis) (i) the fraction of times the Amazon SM actually won the Buy Box according to Amazon's algorithm as observed over the period of our data collection indicated by **squares**, and (ii) the fraction of respondents in our survey who voted for the Amazon SMs (first preference vote) indicated by **triangles**. For 10 of the 15 products, only a single seller, i.e., an Amazon SM, won the Buy Box in all competitions. However, when the competitions were decided by respondents in our survey, in none of the cases all the first preference votes went to Amazon SMs (or any single seller).

USA: Out of the 750 evaluations done by customers in the USA, 157 (i.e., nearly 21%) times Amazon, the seller was given the first preference votes. However, out of all the competitions in our dataset, Amazon won the Buy Box in 92% cases. Figure 4(b) shows the observation segregated across the different products surveyed for Amazon's USA marketplace. For 7 of the 15 products, only a single seller, i.e., Amazon, won the Buy Box in all competitions. In all the products (barring product 6), the first preference vote is significantly lower for Amazon.

Germany: Out of the 750 evaluations done by customers from Germany, 153 (i.e., nearly 20%) times Amazon, the seller was given the first preference votes. However, out of all the competitions in our dataset, Amazon won the Buy Box in 74% cases. Figure 4(c) shows the observation segregated across the different products surveyed for Amazon's Germany marketplace. For only 2 of the 15 products, only a single seller, i.e., Amazon, won the Buy Box in all competitions. Notice, this corroborate with our initial observations wherein we pointed that Buy Box winners are relatively more distributed in Germany. Apart from products 2, 5 and 6 in all the rest of the cases, we again see considerable drop in the first preference votes for Amazon as opposed to its frequency of winning the Buy Box.

France: Out of the 750 evaluations done by customers from France, 214 (i.e., nearly 28%) times Amazon, the seller was given the first preference votes. However, out of all the competitions in our dataset, Amazon won the Buy Box in 83% cases. Figure 4(d) shows the observation segregated across the different products surveyed for Amazon. fr. For 7 of the 15 products, only a single seller, i.e., Amazon, won the Buy Box in all competitions. Apart from products 6 and 15, in all the cases here also the drop off in the two percentages (Buy Box selection vs respondent selection) are significant.

6.3 RQ2: Thematic analyses of the answers mentioning the reason for preferences

As stated in Section 6.1, we also asked each of the respondents to briefly reason about their preference orders (to which they could answer using long answer texts). Note that user preferences are subjective in nature. Hence, to verify the truthfulness of the responses that respondents provided,

minute analysis of the reasons is essential. To this end, we went over the responses manually and figured out the following five prominent themes in the reasons provided by respondents.

Seller performance metrics: The ‘percentage of positive feedback’ and ‘average user rating’ of the sellers were mentioned by many respondents for their reasons to give first preference votes. Almost half of the responses from Indian participants, 30% of the responses from the USA and Germany, and 25% of the responses from France, mentioned these two performance metrics as important features in their consideration. These observations underline the importance of seller performance metrics in customers’ decision making.

Price: Nearly 60% of the respondents from USA and France mentioned price to be an influential factor in their decision making. This percentage is slightly lower for Germany (50%) and India (40%). This observation not only highlights the importance of price in decision making of customers, but also corroborates the findings of prior works [10] where it was shown that price is among the influential factors in algorithmic decision making as well.

Delivery option: Although not explicitly mentioned as many times as some of the other features, 81.2% of the Indian respondents and 66% of the respondents from USA stated that they opt for sellers whose offers are fulfilled by Amazon (including the Amazon SMs). Some of the responses mention FBA to be a preference, e.g., ‘*Seller C has highest number of positive feedback and it is FBA also which is a positive*’; ‘*Preference is based on the number of ratings and positive feedback %*. Also if it is Fulfilled by Amazon it receives a higher preference’. Such preference toward Amazon FBA options is understandable since it almost always ensures a faster delivery too. However, this percentage drops to 37% and 38% in Germany and France respectively, indicating that customers in many countries may not necessarily have any inherent preference for sellers fulfilled by Amazon.

Number of ratings (# Ratings): Surprisingly, 60% of the responses from Indian respondents explicitly included the mention of #Ratings to be one of the most influential features in their decision making. Note that the two Amazon SMs in Amazon.in sell products across all available categories and hence their #Ratings is of the order of 10^6 ; while for majority of the other sellers this is of the order of 10^3 or 10^2 . Thus, the mere mention of #Ratings itself drives customers toward these sellers more. We did not find #Ratings being mentioned prominently in the surveys conducted for other countries.

Amazon, the seller: Another interesting observation in the survey of respondents from USA, Germany and France was the mention of ‘Amazon’ in the reasoning. In 22%, 20% and 28% cases respondents explicitly mentioned Amazon in their explanation in those countries respectively. Some of the explanations are, ‘*I selected Amazon because of their good reputation. I then selected based on best price.*’; ‘*I chose Amazon because I trust them the most. I then selected options based on price*’, etc. In other words, some customers (nearly 1 out of 5) associate trust and reputation with Amazon (the seller) in countries where it sells to customers directly. However, nearly 80% of respondents selected other third-party sellers, showing that they trust other sellers as well.

Consistency in responses: To further quantify the relative effect of each individual features on the preference of respondents and see if the respondents’ reasons corroborate with their decisions we fit a random forest regression model.³ The independent variables are price (normalised by the minimum price within a competition), average rating of the seller, its percentage of positive feedback, its number of ratings and the delivery option (FBA or Not). Given these features, the model is trained to predict the number of times respondents give the first preference votes to the corresponding offers from different sellers. The feature importance analysis of the random forest regression model trained on the dataset mostly corroborates with the answers provided by respondents in the survey. The relative ranking of the different features as calculated by the

³<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

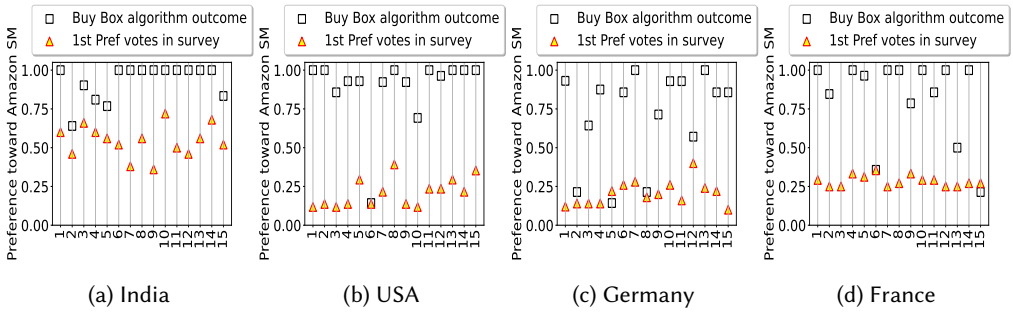


Fig. 4. Preference toward Amazon SMs for the 15 products surveyed, (i) in Buy Box selection outcome (squares), (ii) first preference of respondents when all features were shown (triangles) in its (a) Indian, (b) American, (c) German, and (d) France marketplaces. We consistently observe drop in the customers’ preference toward SMs as opposed to the algorithm’s preference.

Rank	India	USA	Germany	France
1st	#ratings	relative price	% of +ve feedback	relative price
2nd	% of +ve feedback	% of +ve feedback	relative price	% of +ve feedback
3rd	average user rating	delivery option	#ratings	delivery option
4th	relative price	#ratings	average user rating	average user rating
5th	delivery option	average user rating	delivery option	#ratings

Table 5. Feature importance ranking in respondents’ first preference votes as observed in the survey response based on a random forest regression model. We observe that % of positive feedback (a seller performance metric) is among the top-2 most influential feature across all the countries. While in India #ratings is found to be most important, the importance of relative price is higher in the rest of the three countries.

random forest regression model is summarised in Table 5. For India, #Ratings is found to be the most influential feature, followed by percentage of positive feedback. For USA, relative price and percentage of positive feedback are found to be the top most influential features in our regression analysis which are among the top features mentioned by the respondents too. Much like in USA for France, relative price and percentage of positive feedback are found to be the top most influential features and the feature ranking corroborates with the explanations provided by the respondents too. But in case of Germany, although more people mentioned price in their responses, the random forest analyses shows the relative importance of percentage of positive feedback is higher.

Takeaways from this section: The primary takeaway from this section are as follows:

- Some customers associate trust and / or reputation with Amazon, the seller in countries where Amazon sells on its marketplaces. Yet, in a significant fraction of cases most customers opt for offers from sellers other than Amazon SMs.
- In contrast, Amazon’s Buy Box algorithm pre-selects Amazon SMs in a significantly higher frequency for similar offers as per our dataset.
- In India and USA, customers prefer options from Amazon fulfilled sellers (i.e., when the offer is fulfilled by Amazon) more than in European countries like Germany or France.
- Lastly, the self reported explanations and a subsequent analyses on random forest regression model suggest that customers give more importance to the reported seller performance metrics

	India			USA		
Number of	AF	AS	MF	AF	AS	MF
sellers	693	260	47	900	33	67
reviews	462,443	63,869	19,327	439,088	16,665	31,490
Low rated	131,664	8,909	2,978	57,567	2,038	3,262
struck-through	97,603	2,001	0	47,287	304	0
% struck-through (out of low rated)	74.13	22.46	0	82.14	14.92	0
	Germany			France		
Number of	AF	AS	MF	AF	AS	MF
sellers	745	165	90	782	88	130
reviews	1,404,880	356,565	170,795	779,471	125,931	194,700
Low rated	102,774	33,908	12,944	65,284	15,762	18,452
struck-through	57,723	1,927	0	42,010	570	0
% struck-through (out of low rated)	56.16	5.68	0	64.35	3.62	0

Table 6. Statistics of seller feedback data collected for top-1000 active sellers across four of Amazon’s marketplaces. We observe that more than half of the low rated feedback that Amazon fulfilled sellers garner are eventually struck through in all of its marketplaces.

– specifically, (1) % of positive feedback, and (2) number of ratings (primarily in India) – and the price on the offer listing pages.

7 Striking-through of seller feedback on Amazon

Given the importance of seller performance metric e.g., percentage of positive feedback, average user rating etc., in this section we focus on these seller performance metrics and how they are evaluated for each individual seller.

7.1 Dataset collected for seller feedback

We collected the seller feedback garnered by 1000 active sellers on Amazon.in, Amazon.com, Amazon.de and Amazon.fr. We extracted the set of sellers from a list curated by Marketplace Pulse [36], where the sellers are sorted as per the number of feedback received in last 30 days. We collected the seller feedback given to these sellers i.e., the rating (out of 5 stars), the associated review text for each seller feedback (as shown in Figure 2) and if it was struck through by Amazon or not. The data collection methodology is detailed in Appendix C. Our dataset consists of over 4M reviews received by 4000 Amazon sellers in the last 12 months across the four countries. Our set of 4000 sellers includes 3120 Amazon Fulfilled sellers, 546 Amazon shipped sellers and 334 Merchant Fulfilled sellers. A brief overview of the dataset collected is shown in Table 6. The numbers in the table should be read as follows: for Amazon fulfilled sellers in the Indian marketplace (first column), the total number of reviews collected was 462,443. Out of these 131,664 were low rated reviews i.e., reviews where the associated rating was 3 stars or below. Note that these are the potential candidates for review striking through since such striking through would ensure that the average user rating and the percentage of positive feedback is unaffected for the seller and is therefore advantageous for the seller. Finally, 97,603 out of the low rated reviews were actually struck through by Amazon.

7.2 RQ3 (a): Impact of striking-through negative feedback

As discussed in Section 3.4, Amazon has policies in place that allows its *Related Sellers* to apply for striking through some of the low rated seller feedback that they receive. As per Amazon's policies, this striking through also will not reflect on the seller performance metrics. Table 6 shows that nearly 28% of the feedback collected for Amazon fulfilled sellers had a low rating associated with it for its Indian marketplace in our dataset. Out of these, 74.13% were struck-through by Amazon (97,603 low-rated reviews were struck through out of 131,664, thus 74.13%). If we consider the percentage of low rated reviews across the different countries, the performance of Amazon fulfilled sellers in general improves in USA, Germany and France (28% low rated reviews in India as compared to 13%, 7% and 8% low rated reviews respectively for the other countries). However, out of the low rated reviews 82.14%, 56.16%, and 64.35% reviews were struck through for USA, Germany and France respectively. Sellers who use Amazon's delivery services for shipping their product only (i.e., Amazon shipped (AS) type sellers) has very comparable performance to that of Amazon fulfilled sellers across countries. However, the percentage of negative reviews being struck through reduces to 22.46%, 14.92%, 5.68%, 3.62% (out of all low rated reviews) in the respective countries. For the Merchant-fulfilled sellers, who fulfill their order *without* any assistance of Amazon, no low rated reviews were struck through as this functionality is unavailable for these group of sellers.

Discrepancy in the performance metric of *Related Sellers*: Figure 5 shows the percentage of positive feedback calculated with and without the struck-through reviews for some of the top active sellers across the four Amazon marketplaces. The height of the bar for each seller denotes the metric reported by Amazon; while the height of the red portion of each bar is the calculated score by considering all the rating scores (including those which are struck-through). The number on top of each bar denotes the magnitude of drop / increase in the metric. For example, the first seller in Figure 5(a), Appario Retail (an Amazon SM) reports a % of positive feedback of 87% over the past 12 months. However, when we calculate the same including the reviews which have been struck-through, the metric drops to 61%, i.e., a 26% drop. Notice, all of the sellers shown in the figure are Amazon fulfilled and are among the largest active sellers on the respective marketplaces.

Next, we perform the analyses with sellers grouped together in categories based on how they handle their delivery logistics. Figure 6 shows the box plot of % of positive feedback of all the sellers in the different groups. In Figure 6(a, c, e and g), we show the distribution of the % of positive feedback of the sellers as reported by Amazon (i.e., evaluated excluding the struck-through reviews). In Figure 6(b, d, f and h), we show the distribution of the % of positive feedback of the sellers evaluated including the struck-through reviews. Notice, the distributions where we exclude the struck through reviews suggest the % of positive feedback and hence the quality of service of Amazon fulfilled sellers are the best among the three types. However, as soon as we include the struck through reviews in the evaluation, the median percentage of positive feedback for Amazon fulfilled sellers drops by 20%, 11%, 5%, and 7% for India, USA, Germany and France respectively. In fact, for India, USA and Germany, the median value for the Amazon fulfilled sellers actually drops below that of Merchant fulfilled sellers. In addition, a student's *t*-test confirmed that all these drops were statistically significant. In other words, if not for the review strike-through policy of Amazon, the percentage of positive feedback of merchant fulfilled sellers are in fact better than that of Amazon fulfilled sellers, across several countries.

Note that adaptation of such differential policies based on special relationship with the marketplace is in direct violation of the prescription of the Digital Markets Act (see Section 3.4 for more details). At the same time, our experiment shows that the review strike through policy improves the seller performance metric of the *Related Sellers* (esp. Amazon fulfilled sellers) significantly. Given the importance that seller performance metrics have in the decision making process of both

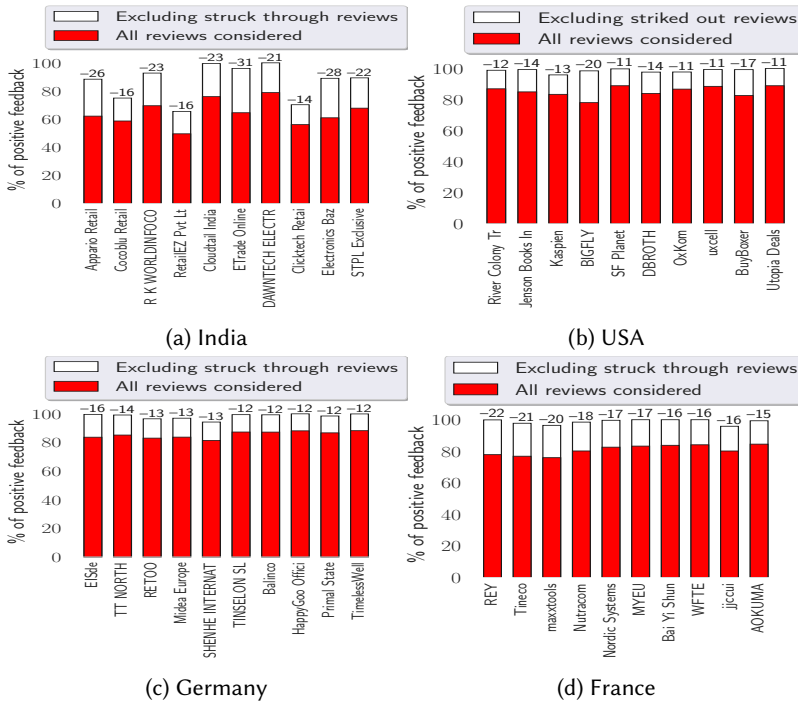


Fig. 5. Discrepancy in the % of +ve feedback of some of the top sellers on Amazon's (a) India, (b) USA, (c) Germany, and (d) France marketplaces. All these sellers have special relationship with Amazon. Amazon significantly over-represents performance metrics for sellers having special relationships.

customers (as discussed in Section 6) and in the Buy Box algorithm [1]– such discrepancy and therefore such policies can potentially result in nudging customers more toward *Related Sellers* as compared to other potentially more deserving third party sellers.

7.3 RQ3 (b): Transparency, interpretability and fairness of strike-through policy

To understand transparency, customers' interpretability and perception of Amazon's feedback strike-through, we conducted another survey on Prolific. We recruited 100 participants on Prolific who are fluent in English and have a very high approval rate ($\geq 98\%$) in prior participated surveys. Each respondent was remunerated at a rate of 9 GBP per hour which is recommended by Prolific to be a good and ethical rate [46]. The gender distribution was balanced at 50-50 for this survey. Most of the respondents are familiar with Amazon e-commerce platform and regularly shop on Amazon too. Next, we discuss the insights from the survey.

Transparency of Amazon's strike-through policy: We ask the respondents to put on the hats of customers of Amazon and answer if they were aware of feedback striking-through being done on Amazon before taking our survey. *78 out of the 100 participants stated that they were not aware of such striking-through on Amazon before taking our survey* (even though they self-report to be regular shoppers on Amazon). This lack of awareness of respondents about this important policy hints at lack of transparency on the part of Amazon marketplace for the strike-through policy they adopt for evaluating the seller performance metrics. In fact, one would not be aware of this policy unless he/she goes through the seller profile page of some of the sellers where they will be able to see seller feedback as shown in Figure 2.

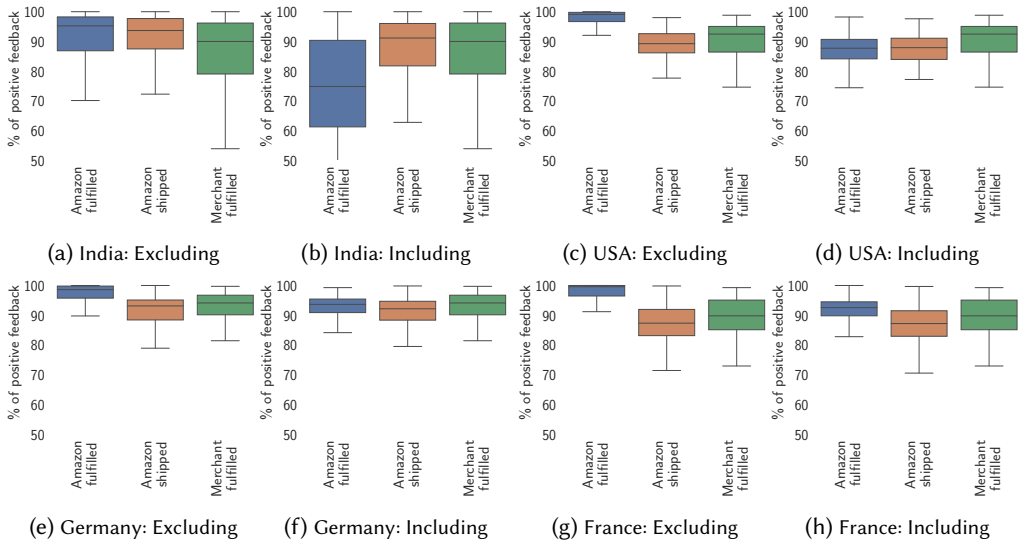


Fig. 6. % of positive feedback of AF, AS and MF sellers calculated excluding and including struck-through reviews for the different Amazon marketplaces. The distribution of AF sellers are very different (as shown in the first box plots in the consecutive figures) highlighting the effect of review striking through on seller performance metrics of *Related Sellers*.

Interpretation of Amazon’s strike-through policy: In the second part of the survey we ask our respondents to evaluate if the policy is being implemented properly. Notice Amazon mentions if the entire comment relates to *delivery experience* then Amazon may strike-through such reviews for Amazon fulfilled sellers. ‘Delivery experience’– is an umbrella term which may include an entire array of issues. We came up with eight distinct types of issues, from a manual inspection of a random sample of struck-through reviews. These issues are – (i) order not delivered, (ii) other delivery issues, (iii) wrong product delivered, (iv) damaged product delivered, (v) return related issues, (vi) refund related issues, (vii) product quality related issues, (viii) canceled order. We wanted to understand which issue (out of these eight) respondents deem to be within the scope of ‘Delivery experience’ and to what degree. To this end, we randomly selected 30 struck-through reviews written in English received by Amazon fulfilled sellers. These reviews correspond to different issues as mentioned above.

During the survey, we show Amazon’s strike-through policy along with a brief description about the “Fulfilled by Amazon” concept to the survey participant. Then we ask the following question: *Review: “<Text of the struck-through review>”*

Amazon has struck-through the above review. Do you think it should have been struck-through as per Amazon’s review strike-through policy?

Further, a follow-up question was also asked as to *why* they think a particular review should / should not be struck-through.

Observation: We received a total of 3000 evaluations for the 30 struck-through reviews. In 1826 cases respondents agreed that the corresponding reviews should have been struck-through as per Amazon’s policy (as it is worded currently). *In the remaining 39% of cases respondents did not agree with Amazon striking-through a particular review.* Table 7 shows examples of reviews that have

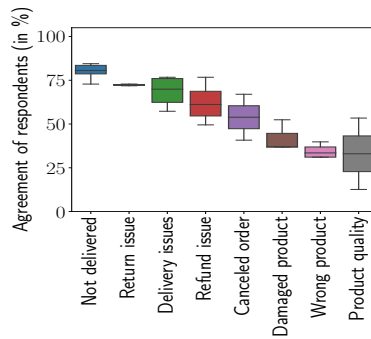


Fig. 7. Percentage of respondents who agreed that reviews mentioning the corresponding issue should be struck-through as per Amazon's policy. The agreement is high for delivery-related issues, but very low for issues mentioning wrong products delivered or product quality.

Struck-through reviews for AF sellers	Reasons why they should not be struck-through (as per respondents)
We returned this item and never received any refund. The refund has been pending for over a month now. Very bad service.	Others should be aware of such potential risks.
Even the replacement I received is wrong, you have mentioned c-type cable but sending it an a-type cable.	It was sent a total different product
Product was so low quality.	The seller is accountable for the quality of products
Why are you showing here different items and sending some different cheats to customers?	the seller is responsible for advertisement of products in stock or being sold
Ordered 4 pcs of Milton 1800ml glass jars. Received one damaged 1000ml. Absolute cheating by the seller. Requested for replacement with the ordered item. Still I am waiting for it.	This is an honest review about the seller that could help other clients so it shouldn't have been taken down

Table 7. Reviews received by Amazon fulfilled sellers which were struck-through by Amazon. However, majority of the respondents in our survey thought these reviews should not be struck-through due to the mentioned reasons.

been struck-through by Amazon, but majority of our respondents felt that these reviews should *not* have been struck-through under the current Amazon policy.

The agreement of annotators with the Amazon's strike-through policy varied across the different issues that the reviews mention about. Figure 7 shows the distribution of percentage of agreement (that a review should have been struck-through) of annotators for different issues that we observed in the random sample. The average agreement rate, i.e., % of respondents who agreed with the review being struck-through, was higher for 'not delivered' (80%), 'return issue' (72%), and 'delivery issues' (68%). For some of the other issues such as 'product quality' (33%), 'wrong product delivered' (34%), and 'damaged product delivered' (42%) the agreement was lower. We include some of the reasons provided by respondents in Table 7 too. These observations highlight the gap between the way the policy is being implemented by Amazon today, and its interpretation by the respondents.

Fairness of Amazon’s strike-through policy: Our next set of questions to the respondents was to think normatively and answer whether they agree with the strike-through policy in place by Amazon. While 45 respondents disagreed with Amazon’s strike-through policy, 47 respondents agreed to the policy (remaining participants being neutral). Thus, respondents are in a split about whether the policy in place is agreeable or not. We further asked them whether such strike-through policies should be uniform across all sellers irrespective of their fulfillment service. *69 respondents answered in the affirmative.* Thus, majority of the respondents think that if such policies are to be in place, they should be uniformly applicable to all sellers irrespective of their special relationships.

Takeaways from this section: To summarize the major takeaways from this section,

- Amazon’s review strike-through policy has significant impact on the evaluation of performance metrics of *Related Sellers*, resulting in high discrepancy between the shown and the actual metrics.
- 78% of the respondents are not aware of striking-through of negative reviews on Amazon, highlighting the lack of transparency on this matter.
- Given Amazon’s policy, respondents do not agree with reviews discussing certain issues should be struck-through. These observations indicate that Amazon may be trying to favour its *Related Sellers* by brushing many factors under the ‘carpet’ of feedback strike-through.
- 69% respondents think that if at all such policies are in place, they should be in place uniformly for all sellers irrespective of the fulfillment services they opt for.

8 RQ3 (c): Impact of performance metrics on customers’ decisions

In the previous sections, we showed how different seller performance metrics influence the customers’ decision-making while choosing among different offers. Sometimes they prioritise features like #Ratings, which is an indicator of the scale at which a seller operates, or features like percentage of positive feedback, whose accuracy is questionable given Amazon’s review strike-through policy. To understand and quantify the extent of the influence of these metrics on the customers, we conducted some more surveys similar to the survey presented in Section 6, with several changes in the settings (which we shall describe next). Note that although we have discussed about what features influence customers’ decision making according to their self-reported explanations and a regression study, it is important to quantify the exact effects by putting customers in a counterfactual setup e.g., where #Ratings is not shown or where rectified seller performance metrics (including the strike through reviews) are shown etc. Such a setting may bring out the precise quantifiable effect of the changes in the customers’ decision-making. We chose to conduct these additional surveys on the Indian marketplace, mainly because we have access to the performance metrics of all the sellers only for the Indian marketplace. Recall that, in the other countries, Amazon does not show any performance metric for itself, nor does it have any seller profile page, which essentially hinders conducting this study in the other marketplaces.

We now describe the additional surveys that we conducted. Each setting, we discuss below, is shown to 50 Indian nationals with approval rate greater than 98% on Prolific. Again, all the participants were remunerated at a rate of 9 GBP per hour. 57% participants in the survey reported their gender to be male while the remaining 43% reported to be females. 72.5% of the participants self reported to be frequent shoppers on Amazon while 19% reported they moderately shop on Amazon too. The four groups of 50 participants are mutually exclusive to one another.

Control setting (Amazon metrics w #Ratings): This is the exact survey conducted in Section 6 for the Indian marketplace where we had shown 15 products and the top-4 offers to 50 respondents from India. To remind the readers about our questions, we asked the respondents “Suppose you are willing to buy a ‘<product name>’ on Amazon and these are the offers from different sellers for the same product. Which one would you prefer to buy?”. Further to understand about what feature

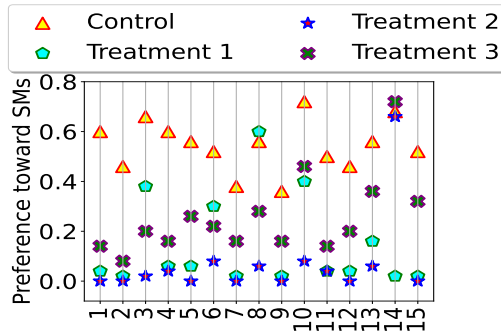


Fig. 8. Fraction of first preference votes to Amazon SMs in control and all the treatment settings as per the different products. Showing the rectified metrics and omitting the #Ratings feature has significant effect on respondents' first preference votes.

is important for the customers to make their decision, we also asked them the following: “Briefly reason about the order of your preferences.”

Treatment 1 (Amazon metrics w/o #Ratings): We pose the exact same questionnaire for the same 15 products, only omitting the #Ratings feature for all offers. Notice the other performance metrics such as average user ratings, percentage of positive feedback were visible to the respondents along with the other features as mentioned in Section 6. Note that this survey will indicate the effectiveness of #Ratings feature on the decision making of customers.

Treatment 2 (Rectified metrics w/o #Ratings): We repeat the survey for the same 15 products, this time with the rectified performance metrics. To this end, we collect data from the seller profile page of the sellers involved in the surveyed competitions and repeat the process mentioned in Section 7.2 to evaluate their performance metrics including all the struck through reviews. Now the respondents will see the newly evaluated performance metrics (which include the struck through reviews). Notice, this setting will help in quantifying the effects of review strike-through policies of Amazon on the decision making of customers.

Treatment 3 (Rectified metrics with #Ratings): Finally, we repeat the survey for the same 15 products, this time with the rectified performance metrics along with the #Ratings shown. During the survey in Section 6, we observed that #Ratings was mentioned to be highly influential in the Amazon India setting. Thus, to conclude the loop in this setting we show the respondents the rectified metrics as evaluated for Treatment 2, but now with an additional #Ratings feature.

Observation: We discuss the effect of the different treatment setting first on the Amazon Special Merchants (Cloudtail and Appario for Amazon.in) and then on the Amazon fulfilled sellers (sellers with Amazon FBA option). In Figure 8 we show the fraction of first preference votes given by respondents to the Amazon Special Merchants for the different settings of the survey explained above. We observe that, except product number 8, in all the other 14 cases, the preference toward SMs have drastically reduced upon omitting the #Rating feature (pentagon markers for Treatment 1 are much below the triangle markers for the Control). When we show rectified metrics without #Ratings (i.e., Treatment 2), the percentage of first preference votes reduces further significantly for Amazon SMs in all cases (the star markers for Treatment 2 are lower than triangles and pentagons) except product 14⁴. Similarly, across all but one case (in product number 14), the fraction of first

⁴Upon closer inspection, we found that the performance metrics of the other Amazon fulfilled sellers for this product also reduced significantly, thereby, keeping the Amazon SM as the seller with best performance metrics.

% of 1st Preference Votes	Amazon Metrics	Rectified Metrics
Special Merchants		
With #Ratings	54.26%	28.53%
Without #Ratings	14.53%	6.93%
Amazon Fulfilled		
With #Ratings	81.20%	64.53%
Without #Ratings	72.13%	54.6%

Table 8. **Percentage of first preference votes aggregated by Amazon Special Merchants and Amazon fulfilled sellers in the different survey settings. We observe that the percentages drop when we show rectified metrics as opposed to metrics shown by Amazon or when we omit the #Ratings feature.**

preference vote reduces when we show the rectified metrics as opposed to the metrics shown by Amazon excluding the struck-through reviews (the cross markers for Treatment 3 are much below the triangle markers for the Control).

In Table 8 (upper panel), we show the percentage of first preference votes aggregated by Amazon SMs in the different settings of our surveys. In the control setting, (as reported earlier) 54% of the first preference votes are given to Amazon SMs. However, in the Treatment 1 setting, we observe that the percentage of first preference votes to Amazon SMs reduces to merely 14.5% (i.e., 109 first preference votes out of 750). In other words, the removal of #Ratings feature massively reduced the number of first preference votes going to Amazon SMs (by 39.5%). In Treatment 2, this percentage further reduces to 6.93% (i.e., 52 first preference votes out of 750). So, with the rectified metrics the percentage of first preference votes got halved as compared to the numbers in Treatment 1. Finally, when we show the respondents the rectified metrics with the #Ratings, the percentage of first preference votes increased to 28.53%, thus again highlighting the effect of #Ratings. However, if we compare the control setting and the Treatment 3 the drop off is still almost nearly 25.47%. Notice the only difference between the two settings is that, in the control setting, we show seller performance metrics excluding the struck-through reviews (as done by Amazon) whereas in Treatment 3, we show the rectified performance metrics.

Now, let us consider the differences in the first preference votes toward Amazon fulfilled sellers (i.e., sellers with FBA options including the two Amazon SMs) as shown in Table 8 (lower panel). In the control setting, 81.2% of the first preference votes went to AF sellers. This number reduces to 72.13% when we simply remove the #Ratings feature. However, when we show the rectified metrics to respondents in Treatment 2 this number reduces to 54.6%. Upon inclusion of #Ratings it further increases to 64.53% but never gets to where it was in the control setting. This may be an artifact of the drop in the Amazon fulfilled sellers' performance metrics (as corroborated in Figure 6).

Moreover, in contrast to the earlier survey, now % of positive feedback, average user ratings of sellers were mentioned by respondents when asked about features that influence their decisions (especially in Treatments 2 and 3– where rectified seller performance metrics were shown). The same is also corroborated by the random forest regression model fit on the new set of first preference votes, wherein the relative importance of % of positive feedback and relative price has increased significantly as indicated in the last column of Table 9.

Takeaways from this section: To summarize the major takeaways from this section,

- Although it is *not* a genuine indication of quality of service (rather indicates the scale of operation of sellers), #Ratings is the feature that largely influences the respondents decision-making. As has been argued earlier, this may nudge customers more toward larger sellers (many of whom are *Related Sellers*), to the disadvantage of smaller third-party sellers.

Rank	Control	Treatment 1	Treatment 2	Treatment 3
1st	#ratings	% of +ve feedback	average user rating	% of +ve feedback
2nd	% of +ve feedback	relative price	% of +ve feedback	average user rating
3rd	average user rating	average user rating	relative price	#ratings
4th	relative price	delivery option	delivery option	relative price
5th	delivery option	–	–	delivery option

Table 9. **Feature importance ranking in respondents' first preference votes as observed in the survey response based on a random forest regression model. Percentage of positive feedback and average user rating's importance have relatively increased in different treatment settings.**

- Showing the rectified performance metrics of the Amazon *Related Sellers* reduces the percentage of first preference votes that they receive during the survey considerably. This suggests that Amazon's review strike-through policy may potentially have considerable effects on the decision making of customers by altering the seller performance metrics on the offer listing pages.

9 Concluding discussion

Summary of the work: To our knowledge, ours is the first systematic attempt of an end-to-end investigation on Amazon e-commerce marketplace, for potential preferential treatment toward its *Related Sellers* due to different design choices and policies in place by Amazon. We studied Amazon's choice architectures deployed to assist customers in cases where multiple sellers compete for selling the same product (over Amazon marketplaces in four different countries). The study was conducted on Amazon's Buy Box algorithm, offer listing page and evaluation of seller performance metrics in the context of its review strike-through policies. Our empirical analyses on 75K+ Buy Box competitions showed that Amazon won the Buy Boxes in more than 80% cases when it competed in our dataset. However, upon surveying competitions of 60 products among 200 participants, we observed significant gap between the algorithmic choices and customer choices. During the survey, customers also mentioned that seller performance metrics influence their decision the most. This led us to understand how these seller performance metrics are being evaluated. To this end, our empirical analyses over 4M seller feedback reviews collected for 4000 top sellers (across the four Amazon marketplaces) suggests that Amazon adopted different policies for evaluating seller performance metrics for sellers who use Amazon's subsidiary services (e.g., FBA). Finally, we also conducted surveys where we show participants the rectified metrics and observed significant differences in preference toward Amazon and its *Related Sellers* among customers.

9.1 Limitations of the work

Limitations of the methodologies adopted: Accessibility to the deployed algorithms and policies would have been the best possible way to approach the investigation at hand. However, a primary limitation of the work is our partial knowledge of the features that are actually used in the algorithms deployed by Amazon. It can be noted that this limitation is present in any *third-party* audit of such deployed proprietary systems. The proprietary nature of the algorithms, in fact, led us to the research pipeline that has been adopted in this work. Nevertheless, we have attempted our best to make the analysis as complete as possible using the publicly known features. Furthermore, we have analysed the Buy Box competitions for a limited set of products for some of the most searched queries which may bring in problems pertaining to selection bias. However, we want to point out that these are products appearing on the first search engine result pages of Amazon's four emerging marketplaces for most popular keywords. Thus, we believe, this strategy enabled

us to check for the competition of some of the most sought after and popular product pages on Amazon platform across the different marketplaces.

Limitations in the survey setting: Another potential limitation pertains to our conducted surveys where customers were asked to order their preferences. We understand that providing preference votes and doing an actual online purchase are two very different settings. Due to lack of other more suitable options, we chose to proceed with the design choices as discussed in the paper. Similarly, with regards to ‘delivery time’ feature, we did not include the exact *delivery duration* during our surveys or analyses. Notice, the delivery duration from the same seller and for the same product can also differ at different timestamps while delivery option (FBA or not) is a more stable indicator. Furthermore, most of our respondents were frequent users of Amazon and are aware of faster shipping being a well-known advantage of Amazon FBA. Their delivery option choices reflected their preference toward faster delivery in general too (see Table 8). Thus, although exact delivery dates were not included in the survey, the usage of the proxy (delivery option) conveys the faster delivery promises. And preference toward the same was also observed during the survey as discussed in the paper.

While the above limitations may affect the findings quantitatively, we believe the qualitative findings will not be significantly different. Moreover, the findings of potential preferential treatment corroborates with similar practices on Amazon ecosystem in other prior studies as well [13, 57].

9.2 Recommendations and takeaways for important stakeholders

Since such choice architectures have such far-reaching effects on multiple stakeholders of the marketplace, next we briefly discuss potential recommendations and takeaways based on the observations in this study.

Potential recommendations for marketplace platforms: Based on the insights of the study, here are few recommendations that may reduce such potential preferential treatments. First, the performance metrics should be evaluated in a uniform fashion for all sellers (irrespective of their special relationship with the marketplace) for a fair competition. This will not only provide more business opportunities to a wider set of sellers but also improve customers’ trust on Amazon’s algorithms. Second, there needs to be more widespread awareness of the different policies that Amazon adopts at different steps, since these policies ultimately affect the decision-making of customers. Although #Ratings is not an indicator of quality, people seem to give more importance to it because it may be considered as a proxy for trust or reliable service (i.e., ‘so many people have bought from this seller’). This was seen in the survey that we conducted in Section 8. Especially when we showed the number of ratings with the rectified metrics the percentage of first preference votes increases from 6.93% to 28.53%. So, complete removal of it might not be a good idea. However, Amazon can instead show how many units of that *specific product* has been sold by the respective sellers. Although, it does not do away with the scale mismatch problem but it may balance the disparity in the scale to some extent. Finally, Amazon should show and reveal its own performance metrics in USA, Germany and France and in all the countries where it can directly sell to customers. Much like third-party sellers, having a seller profile page may improve the transparency which can in fact improve customers’ trust on Amazon or any e-commerce marketplace for that matter.

Takeaways for sellers on online marketplaces: The current work bring to light some of the issues that sellers of different kinds face when they rely on marketplace platforms like Amazon. While most of the issues highlighted in this work are specific to the policies and algorithms deployed on the Amazon marketplaces, similar practices and choice architectures also exist in other e-commerce platforms such as Walmart and Alibaba. So, sellers need to be wary of these practices and policies deployed on e-commerce platforms. Another takeaway point is, irrespective

of which delivery services they adopt (Amazon or third party), they need to make sure that their price and seller performance metrics need to be better for them to remain competitive in the eyes of customers. We hope the intervention of policymakers e.g., the lawsuit from FTC, the DMA regulations [11], etc., into the matter may help the sellers' cause further.

Takeaways for customers on online marketplaces: The choice architecture (deployed algorithms along with the policies) adopted by e-commerce marketplaces not only affect the opportunities of the sellers (by nudging customers toward *Related Sellers*) but also affect the customer satisfaction (e.g., by driving them away from an offer at a lower price by a less 'accomplished' seller). Customers also need to be cautious about the different performance metrics that are shown on the marketplaces and their interpretation. Hopefully, going forward, researchers can play a role in raising such awareness among the general population about the nudges and algorithmic systems deployed on different online socio-technical platforms.

Takeaways for researchers and policymakers: Recently, the EU has agreed on a Digital Markets Act (DMA) [11] to regulate digital marketplaces such as Amazon for making it fair toward different stakeholders. This Act prohibits gatekeepers (such as Amazon) to treat services offered by itself more favorably than similar services of other third-party service providers. However, in order to seamlessly deploy and check the adherence of these regulations, the gap between the laws and their technical operationalization needs to be thought through. This is where Computer Science / AI researchers who have the know-how of the technology and legal scholars who have the know-how of the regulations can come together and bridge such gaps. We believe some of the observations and insights from this work can contribute to such discourse. Given the far reaching economic impact of such apparatus, they should be deployed with the interest of all the stakeholders in mind.

Acknowledgments

This research is supported in part by a European Research Council (ERC) Advanced Grant for the project "Foundations for Fair Social Computing", funded under the European Union's Horizon 2020 Framework Programme (grant agreement no. 789373), and by a grant from the Max Planck Society through a Max Planck Partner Group at Indian Institute of Technology Kharagpur.

References

- [1] Amazon. 2021. Becoming the Featured Offer. <https://amzn.to/3jVq2Vz>.
- [2] Amazon. 2023. Amazon FBA - Get the Prime advantage. <https://bit.ly/3RXqCSm>.
- [3] Solon Barocas and Andrew D Selbst. 2016. Big data's disparate impact. *Calif. L. Rev.* 104 (2016), 671.
- [4] Robin Burke. 2017. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093* (2017).
- [5] Robin Burke, Nasim Sonboli, and Aldo Ordóñez-Gauger. 2018. Balanced neighborhoods for multi-sided fairness in recommendation. In *ACM FAccT*.
- [6] Ryan Calo. 2013. Digital market manipulation. *Geo. Wash. L. Rev.* 82 (2013), 995.
- [7] Amazon Seller central. 2022. Can Amazon remove buyer feedback? <https://bit.ly/3FLGMbu>.
- [8] Amazon Seller central. 2022. Use Buy Shipping services. <https://bit.ly/41Rlnbx>.
- [9] Amazon Seller central. 2022. What happens when I receive a negative feedback on an AFN order? <https://bit.ly/4aOvbqD>.
- [10] Le Chen, Alan Mislove, and Christo Wilson. 2016. An empirical analysis of algorithmic pricing on amazon marketplace. In *WWW*.
- [11] European Commission. 2022. Digital Markets Act (DMA). <https://eur-lex.europa.eu/eli/reg/2022/1925>.
- [12] Mihir Dalal and Shrutika Verma. 2016. Flipkart looks to shift most of its sales to a few sellers. <https://bit.ly/3O4BQmL>.
- [13] Abhisek Dash, Abhijnan Chakraborty, Saptarshi Ghosh, Animesh Mukherjee, and Krishna P. Gummadi. 2021. When the Umpire is also a Player: Bias in Private Label Product Recommendations on E-commerce Marketplaces. In *ACM FAccT*.
- [14] Abhisek Dash, Abhijnan Chakraborty, Saptarshi Ghosh, Animesh Mukherjee, and Krishna P Gummadi. 2022. Alexa, in you, I trust! Fairness and Interpretability Issues in E-commerce Search through Smart Speakers. In *ACM Web Conference*.

- [15] Emily Dayton. 2020. Amazon Statistics You Should Know: Opportunities to Make the Most of America's Top Online Marketplace. <https://bit.ly/2WWEFP8>.
- [16] Anna Edgerton and Leah Nylen. 2023. Lina Khan Is Coming for Amazon, Armed With an FTC Antitrust Suit. <https://www.bloomberg.com/news/articles/2023-06-29/amazon-major-ftc-antitrust-case-expected-in-coming-weeks>.
- [17] USA FTC. 2023. FTC Sues Amazon for Illegally Maintaining Monopoly Power. <https://www.ftc.gov/news-events/news/press-releases/2023/09/ftc-sues-amazon-illegally-maintaining-monopoly-power>.
- [18] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search.
- [19] Anjali Gupta, Shreyans J Nagori, Abhijnan Chakraborty, Rohit Vaish, Sayan Ranu, Prajit Prashant Nadkarni, Narendra Varma Dasararaju, and Muthusamy Chelliah. 2023. Towards Fair Allocation in Social Commerce Platforms. In *ACM Web Conference*.
- [20] Jon D Hanson and Douglas A Kysar. 1999. Taking behavioralism seriously: The problem of market manipulation. *HeinOnline NYUL rev.* 74 (1999).
- [21] Joshua Hardwick. 2021. Top Amazon Searches (2021). <https://ahrefs.com/blog/top-amazon-searches/>.
- [22] Sherry He, Brett Hollenbeck, and Davide Proserpio. 2022. The market for fake reviews. *INFORMS Marketing Science* 41, 5 (2022).
- [23] USA House Committee on the Judiciary. 2020. Judiciary Antitrust Subcommittee Investigation Reveals Digital Economy Highly Concentrated, Impacted By Monopoly Power. <https://bit.ly/3yQmlo3>.
- [24] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on YouTube. *PACM HCI (CSCW)* (2020).
- [25] Eric J Johnson, Suzanne B Shu, Benedict GC Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, David Schkade, et al. 2012. Beyond nudges: Tools of a choice architecture. *Springer Marketing Letters* (2012).
- [26] Prerna Juneja and Tanushree Mitra. 2021. Auditing e-commerce platforms for algorithmically curated vaccine misinformation. In *ACM SIGCHI*. 1–27.
- [27] Daniel Kahneman, Paul Slovic, and Amos Tversky. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.
- [28] Aditya Kalra. 2021. Amazon documents reveal company's secret strategy to dodge India's regulators. <https://reut.rs/2Yt3GIL>.
- [29] Aditya Kalra. 2021. India's draft e-commerce policy calls for equal treatment of sellers. <https://reut.rs/3yNV7P6>.
- [30] J. J Kaziukenas. 2019. Amazon Has Three Million Active Sellers. <https://bit.ly/38JhfiL>.
- [31] Lina M Khan. 2016. Amazon's antitrust paradox. *Yale LJ* (2016).
- [32] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P Gummadi, and Karrie Karahalios. 2017. Quantifying search bias: Investigating sources of bias for political searches in social media. In *ACM CSCW*.
- [33] Eyal Lanxner. 2021. The Amazon Buy Box: How It Works for Sellers, and Why It's So Important. <https://bit.ly/3h5a21a>.
- [34] Sheng Luo, Bin Gu, Xingbiao Wang, and Zhaoquan Zhou. 2018. Online compulsive buying behavior: The mediating role of self-control and negative emotions. In *ICEBI*. 65–69.
- [35] Bismah Malik. 2021. How alpha sellers have put Amazon, Flipkart in dock? <https://bit.ly/3U0xiBB>.
- [36] MarketllacePulse. 2022. Top Amazon.in Marketplace Sellers. <https://www.marketllacepulse.com/amazon/top-amazon-india-sellers>.
- [37] Binny Mathew, Anurag Illendula, Punyajoy Saha, Soumya Sarkar, Pawan Goyal, and Animesh Mukherjee. 2020. Hate begets hate: A temporal study of hate speech. *PACM HCI (CSCW)* (2020).
- [38] Arunesh Mathur, Gunes Acar, Michael J Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark patterns at scale: Findings from a crawl of 11K shopping websites. *PACM HCI 3, CSCW* (2019).
- [39] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *ACM CIKM*.
- [40] Rishabh Mehrotra, Niannan Xue, and Mounia Lalmas. 2020. Bandit based optimization of multiple objectives on a music streaming platform. In *ACM SIGKDD*.
- [41] Carol Moser, Sarita Y Schoenebeck, and Paul Resnick. 2019. Impulse buying: Design practices and consumer needs. In *ACM SIGCHI*.
- [42] Nuno Mota, Abhijnan Chakraborty, Asia J Biega, Krishna P Gummadi, and Hoda Heidari. 2020. On the desiderata for online altruism: Nudging for equitable donations. *ACM PHCI 4, CSCW2* (2020).
- [43] Gourab K Patro, Arpita Biswas, Niloy Ganguly, Krishna P Gummadi, and Abhijnan Chakraborty. 2020. FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms. In *WWW*.

- [44] Gourab K Patro, Abhijnan Chakraborty, Niloy Ganguly, and Krishna P Gummadi. 2020. Incremental Fairness in Two-Sided Market Platforms: On Smoothly Updating Recommendations. *AAAI* (2020).
- [45] PIB. 2018. Review of policy on Foreign Direct Investment (FDI) in e-commerce. <https://bit.ly/3BP8vnO>.
- [46] Prolific. 2023. How much should you pay research participants? <https://bit.ly/47ALRiv>.
- [47] EU Press Release. 2020. Antitrust: Commission sends Statement of Objections to Amazon for the use of non-public independent seller data and opens second investigation into its e-commerce business practices. <https://bit.ly/3h8jAIR>.
- [48] Amazon's Response. 2019. Online Platforms and Market Power, Part 2: Innovation and Entrepreneurship. <https://bit.ly/2VnSS7q>.
- [49] Amazon's Response. 2020. Hearing of the Subcommittee on Antitrust, Commercial, and Administrative Law, Committee on the Judiciary. <https://bit.ly/3n9Um0k>.
- [50] Punyajoy Saha, Binny Mathew, Kiran Garimella, and Animesh Mukherjee. 2021. "Short is the Road that Leads from Fear to Hate": Fear Speech in Indian WhatsApp Groups. In *Proceedings of the Web conference 2021*.
- [51] Nicola Schmitt. 2023. With Amazon Around the World – The Amazon Marketplaces Worldwide in Focus. <https://www.blankspace.eu/blog-posts-en/amazon-marketplaces-worldwide>.
- [52] Christoph Schneider, Markus Weinmann, and Jan vom Brocke. 2018. Digital nudging: guiding online user choices through interface design. *ACM CACM* (2018).
- [53] Daria Sorokina and Erick Cantu-Paz. 2016. Amazon search: The joy of ranking products. In *ACM SIGIR*.
- [54] RH Thaler and CR Sunstein. 2008. *Nudge: improving decisions about health, wealth and happiness Penguin*. Penguin Books, New York.
- [55] Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *science* 211, 4481 (1981).
- [56] Amazon Seller University. 2023. About Feedback Manager. <https://sell.amazon.in/sell-online/fulfillment-by-amazon>.
- [57] Leon Yin and Adrienne Jeffries. 2021. How We Analyzed Amazon's Treatment of Its "Brands" in Search Results. <https://bit.ly/420SiKG>.

A Data collection for Buy Box competitions

One of the primary requirement for our study is to identify a set of products (preferably popular ones) for which there are multiple sellers competing for the Buy Box. In addition, for each product the details of which seller(s) win the Buy Box when the product page is visited a number of times is also required. Note that, visiting the same product pages multiple times is important to check whether the Buy Box winners are rotated over time [10]. Given the popularity of product search on e-commerce platforms [53], we proceeded to collect our dataset by navigating the product pages following the links from the search results. In particular, as a proof of concept, we collected data from all the Amazon marketplaces for a set of 100 distinct most searched keywords (queries) on Amazon [21] across multiple popular categories. The reason behind selecting such queries is twofold. First, since these queries are the most searched ones on Amazon, they enable us to land up on some of the most popular product pages visited by the customers. Second, through these queries we were able to come across multiple products where more than one sellers have offered the same product at different competitive prices.

Crawling methodology: To collect the search results (and subsequently visit the product pages), we designed a crawler performing search operations on Amazon. The crawler was seeded with a query, and the search results on the first SERP for the query were scraped. This process was repeated for all the queries. We used desktop browser automation for data collection. In particular, we used Gecko driver and Mozilla Firefox browser for data collection. We collected the data in different sessions from an Amazon account having prime membership. Therefore our choice is motivated to represent the behavior as observed by a majority of highly active Amazon customers [15].

We took some precautions during the data crawling process to ensure the meaningfulness of the analyses. The Amazon search algorithm is likely to tailor results based on several signals, including the geographical location, browsing history, etc. We also observed that search results and rankings vary over time. To minimize these variations, we collected search results in multiple temporal

Offer Rank	Sellers	Base Price + Delivery Charge (in INR)	Past performance Avg. rating (% of pos. feedback)	% of competitions won
1	Global infotech service	149 + 70 = 219	3.5 (67 %)	9.8 %
2	Bhaskar & Company	158 + 90 = 248	4 (73%)	0.0
3	D-cart	175 + 75 = 250	4.5 (86%)	0.0
4	CloudTradez	265 + Free = 265	No feedback yet	0.0
5	Supersure	265 + Free = 265	3 (47%)	0.0
6	Hindustan kh3	189 + 80 = 269	4 (78%)	0.0
7	Kasnox	229 + 49 = 278	3 (50%)	0.0
8	SUPER DEBUG	285 + Free = 285	4 (82%)	0.0
9	Avega Innovations	245 + 50 = 295	No feedback yet	0.0
10	Clouttail India	301 + Free = 301	4.5 (89%)	90.2 %

Fig. 9. An out of position Buy Box win by an Amazon SM. The offer rank column shows the position in which the corresponding offers were shown as per price and delivery in the offer listing page. The seller marked in red (Clouttail India – an Amazon SM) was the winner of the Buy Box in 90% of the competitions for the product.

snapshots during the *same time frame*, from the *same IP address*, from the same user account, and from the same geographical location.

B Out of position Buy Box win

We show an example in Figure 9. The figure shows, for every offer (for the product), the total price (base price + delivery charges), performance metrics of the seller, and percentage of times the seller won the Buy Box out of all the temporal snapshots we collected for this product. The ‘Offer rank’ column denotes the order in which the offers were shown on the Amazon offer listing page at the time we collected the data (we show the top-10 offers only). Figure 9 shows the set and ranking of offers which we observed in majority of the snapshots for this particular product. The seller highlighted in Figure 9 (Clouttail India, an Amazon SM) is the one that won the Buy Box in more than 90% of the snapshots when we encountered the corresponding products. Importantly, there were a number of sellers who offered the same product at lower prices and had comparable (or even better) performances as the SM. Despite this, a majority of the times the Buy Box went to the SM.

C Data collection for seller feedback

The other important dimension of the work is to investigate the effect of review striking-through on different sets of sellers on Amazon. To this end, we collected the seller feedback collected by 1000 active sellers on Amazon’s Indian, American, German and French marketplaces. We extract the set of sellers from a list curated by Marketplace Pulse [36], where the sellers are sorted as per the number of feedback received in last 30 days.

Crawling methodology: Much like the methodology adopted in the earlier data collection, here also we take help of browser automation. We used Gecko driver and Mozilla Firefox browser for data collection. Amazon arranges 5 seller feedback per page on its seller profile page. In Amazon’s lingo, a feedback is a **<rating, review>** pair i.e., each review is associated with a rating out of 5 stars. For each seller, we go to their Amazon seller profile pages and collect the feedback received by them in last 12 months or until when Amazon allows us to gather the data for the corresponding sellers. For some sellers, our crawler never went beyond a certain number of reviews even after multiple trials. Hence, we resorted to collect either the number of reviews mentioned for last 12 months or until when Amazon allows us to collect the data. The threshold was influenced by the fact that Amazon reports the % of +ve feedback over past 12 months on offer listing pages.

Received July 2023; revised January 2024; accepted March 2024