



“TAMING TECH GIANTS’ ALGORITHMS: THE AMAZON BUY-BOX CASE”

by

Francesco Clavorà Braulin, Ph.D. - Economist

Post-Doctoral Researcher

ZEW – Leibniz Centre for European Economic Research · Digital Economy

L 7, 1 · 68161 Mannheim · Germany

Phone +49 621 1235-318 · francesco.clavorabraulin@zew.de

and

Michela Boldrini, Ph.D. - Economist

Post-Doctoral Fellow

University of Turin · Department of Economics and Statistics “Cognetti de Martiis”

Lungodora Siena 100 · 10153 Turin · Italy

Phone +39 3343786996 · michela.boldrini@unito.it

14 May 2021

1. Motivation

Many digital platforms operate as two-sided markets, facilitating the matching between sellers and buyers. Due to ICT developments, online marketplaces such as Amazon are increasingly relevant in our societies. The functioning of the Amazon platform is of great scientific interest because of its dual role, which poses several policy-relevant questions. On the one hand, Amazon controls buyers' and sellers' interactions through its platform design choices, governed by opaque and not transparent algorithms. On the other hand, Amazon operates as a seller itself, leveraging its gatekeeper power and the information collected on both sides of the market to gain a competitive advantage over other market participants and, ultimately, shape competition.

Regulators fear that algorithms can reduce competition and harm consumers: there is a raising awareness about the risks that algorithms pose and a call for their effective regulation. The European Commission points out that *“the use of seller data allows Amazon to avoid the normal risks of retail competition and to leverage its dominance in the market”*. We will contribute to this debate providing new insights on proprietary algorithms. Our research will also answer regulatory questions about how these algorithms are used, how sellers and consumers' information is exploited on two-sided platforms, and how consumer behavior is shaped by the platform's strategic choices. Ultimately, we plan to provide solutions for antitrust in digital markets both from a competition policy and a consumer protection perspective.

We focus on a very specific concern raised by the design of Amazon's online platform: a customer can directly proceed to checkout by clicking on the “add-to-cart” button, known as the *Buy-Box*. However, it is the Amazon's ranking algorithm that selects the seller to be placed in this prominent position, whereas all the other sellers listing the same product are relegated to subsequent pages. Being in the Buy-Box is crucial for sellers, as available evidence (Chen et al. 2016) suggests that roughly 80% of customer purchases go through this button.

The main issue related to the Buy-Box is that the algorithm selecting the Buy-Box seller is neither observable nor decodable. The criticality of this design choice is recognized also by the European Commission, which plans to investigate *“whether the criteria that Amazon sets to select the winner of the Buy-Box lead to preferential treatment of Amazon's retail business”*.

Research Questions. We analyze Amazon's dual role from the two sides of the market, investigating three interconnected questions. On the sellers' side: (1) Which sellers' features explain the Buy-Box winning dynamics and how does the prominent Buy-Box position granted to a specific seller impact competition? On the consumers' side: (2) How relevant is the Buy-

Box channel for actual sales on the platform? (3) Does Amazon, based on consumers' characteristics, personalize search results or, even, displayed prices on own products?

The first question responds to a fast-growing regulatory demand to investigate the functioning of algorithms and their impact on competition. The second and the third questions pertain to the design of the platform environment and the interface consumers interact with, aiming to quantify the salience of the Buy-Box for buyers and their literacy on its functioning, and to tackle the issue of consumers' data exploitation and discrimination on digital platforms.

The relevance of our findings could go beyond purely scientific interest and contribute to informing policy-makers on how to design a more effective regulation of algorithms. We will also shed light on consumers' behavior and decision-making in online environments through a novel approach that combines observational and experimental data. Consumer protection is a first-order concern, and our findings could guide the implementation of measures to foster transparency and consumer awareness, which would be less intrusive (and more viable for the regulator) than other remedies such as platforms' design changes.

Research Question 1: Unraveling persistence in Amazon's Buy-Box algorithm

To answer our first research question, we plan to rely on a new set of data collected directly from the Amazon website through web scraping techniques. Our goal is to quantify the amount of persistence in the Buy-Box selection unexplained by reviews, pricing, and other relevant sellers' characteristics. A possibly biased algorithm would lead to inefficient matches, lower welfare, and anticompetitive Buy-Box dynamics such as self-preferencing, a practice banned by the recent Digital Markets Act proposal of the European Commission.

Literature Review. Buy-Box's related questions have already been addressed: Chen et al. (2016) and Gómez-Losada & Duch-Brown (2019) investigate the sellers' race for the Buy-Box using a predictive approach based on classification. Instead, Chen and Tsai (2021) study the "Frequently Bought Together" offers showing that products sold by Amazon itself receive substantially more recommendations. Finally, Zhu and Liu (2018) study entry of Amazon as a "seller" on the market. Our approach differs because we dynamically model the probability of winning the Buy-Box using novel data, to overcome the limitation of the data collected by previous studies (e.g. lacking information on all sellers listing the products or on inventory stocks).

Framework & Design. To overcome those limitations, we plan to collect data on a group of Amazon's best-sellers products over time: we will gather full-spectrum information on all sellers listing each product, including information on sellers' pricing, services, overall quality and multi-market presence. This information will allow us to model the Buy-Box winning dynamics

in a high-quality information setting, and to investigate whether this algorithmic-driven mechanism ultimately harms consumers, failing to facilitate an efficient sellers-buyer matching.

Empirical Approach & Hypotheses. To investigate whether relying on an “extreme” form of ranking algorithm could ultimately harm competition and consumers, we focus on unraveling what factors drive the selection - and the persistence - of a certain seller in the Buy-Box position. We assume that the two most relevant factors for Buy-Box eligibility are sellers' (i) ability to price competitively and (ii) "service quality", a comprehensive dimension encompassing features such as customer care, shipping and inventory management. In presence of an unbiased Buy-Box algorithm, we would expect that sellers' pricing and quality characteristics can perfectly predict sellers' Buy-Box winning probability. Should this not be always the case, there may be factors other than those linked to sellers' competitiveness and quality that affect the probability of winning the Buy-Box. Past Buy-Box winners, who get the opportunity to build a strong reputation on the platform, could leverage their competitive advantage - i.e. through non-competitive pricing strategies that ensure higher mark-ups - without compromising their chance to be re-selected as Buy-Box winners in the future. This would lead to some form of *unexplained persistence* in the Buy-Box selection, with possibly detrimental effects for consumers.

$$Y_{\text{BUY-BOX}_{i,t}} = I(\beta_0 + \beta_1 Y_{\text{BUY-BOX}_{i,t-1}} + \beta_2 Z_{i,t} + c_i + u_{i,t} > 0)$$

To this extent, we model the Buy-Box winning probability as to capture the role of observable sellers' characteristics ($Z_{i,t}$), while disentangling the direct effect of the past visibility competitive advantage from that of sellers' unobserved heterogeneity (c_i).

Research Question 2: What do consumers know (and want)? An analysis of the Buy-Box

To answer our second research question, we will collect novel experimental data that would allow us to study consumers' experience with the Buy-Box. We aim to quantify how relevant is the Buy-Box channel for actual sales on the platform and to measure consumers' literacy on its functioning; second, we aim to investigate if – and to what extent - platforms' design choices, which determine the quality and the quantity of information provided on sellers, affects consumers' choice to rely on the Buy-Box for their purchases on Amazon. Ultimately, we aim to understand whether the Buy-Box is utility-enhancing for consumers.

Literature Review. Most of the empirical literature focusing on consumers' behavior on online marketplaces relied on ad hoc field experiments to investigate how consumers respond to the information provided by platforms on sellers' quality and reputation (Jin & Kato, 2006; Resnick et al., 2006). Instead, less attention has been devoted to issues related to platform design (Dinerstein et al., 2018).

Framework & Design. From a theoretical standpoint, to study how consumers interact with Amazon’s Buy-Box and react to information disclosed about the Buy-Box functioning, we model the consumers’ purchasing choice environment in a stylized way based on Jolivet and Turon (2018). Consider n sellers, listing a homogeneous product: each seller posts a price p_i and is endowed with a quality index q_i . Consumers have homogenous valuations for sellers’ quality but incur in a search cost when exploring the second page showing all other sellers listing the product. We allow for two levels of heterogeneity, so that consumers differ in: (1) their marginal cost of search (i.e. effort, time); (2) their level of “trust” in the Buy-Box algorithm, having different beliefs on how the algorithm is effective in selecting the best deal for them.

We plan to collect novel individual-level data on consumers’ purchasing behavior on the Amazon marketplace through an incentivized experiment where we manipulate the amount of information disclosed to consumers on the functioning of the Buy-Box mechanism. We plan to conduct the experiment online in a controlled setting, recruiting a heterogeneous and representative pool of participants, so to have a higher variability in terms of individual search propensity. Participants, on top of receiving a fixed fee for their participation in the experiment, will be given an exogenously set budget and let free to make their purchasing choices in the way that best fits their needs without any constraint but the budget: they will be asked to “add to their cart” all the products they would wish to buy, knowing that we will proceed finalizing the desired purchases only for a subset of participants to be randomly extracted at the end of the experiment.

The treatment manipulation relates to the quantity and the content of the information provided to participants on the functioning of the Buy-Box assignment mechanism, prior to the execution of the main task. The design will include three treatment conditions (between-subjects):

- *T0 - Baseline:* No information/mention on the functioning and the role of the Buy-Box.
- *T1 - Partial information:* We will provide information on Amazon’s platform design and on the functioning of the Buy-Box mechanism. Subjects will be informed of the possibility to learn – with an additional click - about the offers made by all the other sellers.
- *T2 - Full information:* We will provide the same information provided in T1. Additionally, we will inform participants of the lack of transparency of Amazon’s algorithm.

Empirical Approach & Hypotheses. Comparing the share of consumers opting in and out of the Buy-Box under the three different treatment conditions, we want to test whether:

- *H0:* most of the consumers (at least 50%) go through the Buy-Box for their purchases.
- *H1:* the “partial information” provision in T1 reduces the share of Buy-Box users (T1 vs. T0)
- *H2:* the “full information” provision in T2 reduces the share of Buy-Box users (T2 vs. T0)

- *H3: the “full information” treatment T2 has a stronger effect than the “partial information” treatment T1, with a stronger reduction in the share of Buy-Box users (T2 vs. T0 > T1 vs. T0)*

Research Question 3: Discrimination and personalization on the Amazon marketplace

To answer our third research question, through an online incentivized survey, we plan to collect information on the degree of “personalization” in what Amazon shows to its consumers, for example, in terms of shown products, sellers’ ranking in a given product page or products’ prices. To this aim, we will collect data on what Amazon shows to new/unregistered users and registered users with different purchasing histories, so to test whether, and eventually to what extent, Amazon employs discriminatory practices exploiting personal information on consumers, quantifying the welfare effects of such practices.

Literature Review. Access to consumer data increases the ability of firms to categorize consumers and reach them with personalized offers. There is evidence, although limited, of personalization in digital markets (Hindermann, 2018; Hupperich et al., 2018; Hannak et al., 2014). The punchline of these studies is that consumer targeting revolves around user-based, technical and location-based features. On this note, Amazon is in an extremely privileged position, knowing consumers’ order history, prospective purchases and payment card details.

Framework & Design. Within our framework, we will be able to detect personalization along the following dimensions: (i) Buy-Box seller identity, (ii) Buy-Box price when Amazon itself is the default option, (iii) displayed ranking of third-party sellers, and (iv) further product suggestions. We plan to select a subset of approximately 20 representative products across different categories and price ranges. Subjects will be invited to visit these product pages simultaneously and in the same sequential order, so to gather individual-level data on all the features shown to consumers on the marketplace scraping the entire product page. We will complement our data collection strategy scraping what is shown to a “fake” unregistered user without any previous order history. At the end of the task, will ask subjects to make one *real* purchase decision by placing their preferred item in their cart. Randomly, on top of the participation fee, some participants will have their order placed and will receive their purchases.

Empirical Approach & Hypothesis.

- *H1: Verify whether there is some form of personalization over the dimensions considered.*
- *H2: In presence of discriminatory practices involving the Buy-Box button, does this form of personalization affect the share of users opting for the Buy-Box offer?*

References

- Chen, L. (2017). Measuring algorithms in online marketplaces. *Ph.D. Dissertation, College of Computer and Information Science, Northeastern University.*
- Chen, L., Mislove, A., and Wilson, C. (2016). An empirical analysis of algorithmic pricing on Amazon marketplace. In *International World Wide Web Conference Committee (IW3C2).*
- Chen, N. and Tsai, H-T. (2021). Steering via algorithmic recommendations. *Mimeo.*
- Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. (2018). Consumer price search and platform design in internet commerce. *American Economic Review*, 108(7): 1820-1859.
- Ezrachi, A. and Stucke, M. E. (2016). Virtual Competition. The promise and perils of the algorithm-driven economy. *Harvard University Press.*
- European Commission. 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, *Press release (Brussels, 10 November 2020).*
- Gomez-Losada, A. and Duch-Brown, N. (2019). Competing for the Amazon's Buy-Box: A Machine-Learning approach. In *W. Abramowicz and R. Corchuelo: BIS 2019 Workshops.*
- Hannak, A., G. Soeller, D. Lazer, A. Mislove, and C. Wilson (2014). Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference, ACM*, 305-318.
- Hindermann, C. M. (2018). Price discrimination in online retail. ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg.
- Hupperich, T., D. Tatang, N. Wilkop, and T. Holz (2018). An empirical study on online price differentiation. In *Proceedings of the Eight ACM Conference on Data and Application Security and Privacy*, 76-83.
- Jin G. Z. and Kato A. (2006). Price, Quality, and Reputation: Evidence from an Online Field Experiment. *The RAND Journal of Economics*, 37(4), 983-1004.
- Jolivet, G. and Turon, H. (2019). Consumer search costs and preferences on the Internet. *Review of Economic Studies*, 86, 1258–1300.
- Resnick P., Zeckhauser R., Swanson J., and Lockwood K. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics*, 9: 79-101.
- Zhu, F. and Liu, Q. (2018). Competing with complementors: An empirical look at Amazon.com. *Strategic Management Journal*, 39, 2618-2642.