

Engel, Julia F.; Nüß, Patrick; Rudolph, Meike; Schwarz, Julia

Working Paper

A Comment on "Taste-Based Gender Favouritism in High-Strike Decisions: Evidence from the Price is Right"

I4R Discussion Paper Series, No. 151

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Engel, Julia F.; Nüß, Patrick; Rudolph, Meike; Schwarz, Julia (2024) : A Comment on "Taste-Based Gender Favouritism in High-Strike Decisions: Evidence from the Price is Right", I4R Discussion Paper Series, No. 151, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/302895>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

No. 151

I4R DISCUSSION PAPER SERIES

A Comment on “Taste-Based Gender Favouritism in High-Strike Decisions: Evidence from the Price is Right”

Julia F. Engel

Patrick Nüß

Meike Rudolph

Julia Schwarz

September 2024

I4R DISCUSSION PAPER SERIES

I4R DP No. 151

A Comment on “Taste-Based Gender Favouritism in High-Strike Decisions: Evidence from the Price is Right”

Julia F. Engel¹, Patrick Nüß¹, Meike Rudolph¹, Julia Schwarz¹

¹University of Kiel/Germany

SEPTEMBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

A comment on “Taste-based Gender Favouritism in High-Stake Decisions: Evidence from the Price is Right”*

Julia F. Engel
U Kiel

Patrick Nüß
U Kiel

Meike Rudolph
U Kiel, IAB

Julia Schwarz
U Kiel

June 30, 2024

Abstract

We conduct a computational replication of [Atanasov et al. \(2023\)](#). In total, our analysis covers three variations: we use the cleaned dataset provided in the replication package, we clean the original data ourselves, and finally we extend the dataset to encompass an additional three years of data using the webscraper provided by the authors. The additional data boosts the final observation count by approximately one-quarter. We find that the results are robust; the data in the replication package results in nearly the same estimates and an extension of the data and specifications reduces the effect size and statistical significance, but does not change the conclusions. We further conduct a wide range of robustness checks. While some estimates have smaller effect sizes and lower statistical significance, all results support the original findings.

KEYWORDS: Replication, Analysis of Collective Decision-Making, Wages, Compensation, and Labor Costs, Gender Discrimination

JEL CODES: D7, J3, J7

*Authors: Corresponding author: Julia Engel, E-mail: j.engel@economics.uni-kiel.de, Patrick Nüß E-mail: nuess@economics.uni-kiel.de, Meike Rudolph, E-mail: rudolph@economics.uni-kiel.de, Julia Schwarz, E-mail: j.schwarz@economics.uni-kiel.de.

1 Introduction

[Atanasov et al. \(2023\)](#) present evidence for own-gender favoritism with data from the popular US television series *The Price is Right*. Thereby, they make an important contribution to the literature by presenting a framework that allows for a clean disentanglement of gender favoritism from beliefs about gender differences in skill.

The Price is Right is an American game show whose premise involves contestants guessing the price of various items. The particular focus of the analysis is on the *One Bid* minigame and, more specifically, gender differences in bidding behavior. The minigame is played with four contestants, picked from the audience. The contestants must then guess the price of a given item in their selection order. The contestant whose guess is closest to the price without exceeding it wins. As winners of this minigame become contestants in further games with expected prizes of \$12,000, the *One Bid* game therefore has high stakes.

In the *One Bid* game, the sequential order of the bids gives rise to a weakly dominant strategy for the fourth player, who is the last to bid. A fourth player has the incentive to overbid the player they currently believe is leading by \$1, employing a cutoff strategy. Overbidding the (perceived) leader by a marginal amount ensures that they are closer to the price, but reduces the margin of error as much as possible.¹ This strategy is established: in the original dataset 48% of contestants employ a cutoff strategy at some point. This property is exploited by the authors.

[Atanasov et al. \(2023\)](#) study how the cutoff behavior varies by gender, in particular whether men/women are less likely to cut off a same gender opponent. They analyse the bidding behavior in 11,016 *One Bid* games, aired between 1972 and 2021. The main findings of the authors are as follows. In general, men are more likely to employ a cutoff strategy than women. However, both men and women are more likely to overbid an opposite-gender opponent by exactly \$1 than a same gender opponent. The authors find no evidence that this behavior is driven by rational, i.e. profit-maximizing, reasons. Instead, this behavior is consistent with own-gender favoritism.

We conduct a computational replication of [Atanasov et al. \(2023\)](#). While the original publication conducted the analysis in R, we provide estimates of a computational replication in Stata. Our replication consists of three sub-replications, as well as a

¹For example: Assume an item costs \$260, the current leader bid \$250, submitting a too high bid of \$270 (\$20 more) would lose the game. However, submitting a bid that is \$1 higher places the bidder closer to the true price than the previous leader (\$251 vs. \$250).

handful of robustness checks with respect to fixed effect selection and model choice. The first replication uses the analysis dataset provided in the authors' replication package as a .csv file. Using this file, we are able to fully reproduce the results. Because the original raw and scraped data is provided, we conduct a second replication including the data cleaning process. Due to a minor coding error in the provided R script for data cleaning, extracting data from the HTML files was initially unsuccessful. After correcting this error, we obtained a similar cleaned dataset. Yet, in eight observations, the latter differed from the dataset provided by the authors. This discrepancy was caused by the "gender" package, which assigns the most likely gender based on the participants' names and produces different results depending on the version of R used (Blevins and Mullen, 2015).² Despite these issues, we were able to replicate the same results with our own cleaned dataset. Finally, our third replication extends the analysis to episodes from 2021 (end of sample) to April 2024. We collect data on the most recent episodes of the *Price is Right*, increasing the number of observations from 11,016 to 14,832. We run the same models as before and find that even though the point estimates change slightly, the patterns and directions of the coefficients are consistent with those initially reported.

In addition to the results already mentioned, the authors also note that own-gender favoritism is stable over time, while the gender composition varies. We test this channel in two ways. First, we add a mediator analysis with time fixed effects. Second, we analyze subsamples to account for potential changes over time. Overall, our results strongly support the findings and conclusions in Atanasov et al. (2023).

2 Computational Reproducibility

We used the replication package [linked in the originally published article](#). The replication package included the cleaning codes, as well as both the raw data and the cleaned analysis data. We successfully computationally reproduced all the main results (*i.e.*, Tables 1, 2 and 3) from the raw data. See Table 1 for details.

In accordance with the replication package provided by Atanasov et al. (2023), we divided the data cleaning procedure into two parts. The first step involved downloading HTML files scraped from *The Price Is Right Episode Guide* forum and cleaning them. Given that Stata is not optimal for unstructured web scraping exercises, we opted to recycle the R code provided by the authors. Although the initial web scraping process was successful, a minor coding error in the data cleaning script

²In addition, some team members had difficulty accessing the "gender" package.

Table 1: Replication Package Contents and Reproducibility

Replication Package Item	Fully	Partial	No
Webscraper provided	✓		
Raw data provided	✓		
Analysis data provided	✓		
Cleaning code provided	✓		
Analysis code provided	✓		
Reproducible from raw data	✓		
Reproducible from analysis data	✓		

Notes: This table summarizes the replication package contents contained in [Atanasov et al. \(2023\)](#).

resulted in the inability to extract data from the previously scraped HTML files. We were able to identify and fix the coding error, which consisted of a misspecified path to the scraped files, causing subsequent commands to address a vector of episode names instead of a directory of files. Once corrected, the R script worked as intended. Based on the revised R script, we constructed two datasets.

The first dataset is based on the data originally scraped by the authors in 2021 and uploaded in the replication package. We clean the provided raw data ourselves, we do not scrape the data anew. This process yielded a dataset that was largely similar to the original one. In fact, only eight observations differed from the version provided directly by the authors. The discrepancies can be attributed to the differing results obtained when the “gender” package ([Blevins and Mullen, 2015](#)) is used with different versions of R (while the authors utilized R 4.2.1, our script was executed on R 4.2.2.), which assigns the most likely gender based on participants’ names. This results in 11,018 observations for the main data set employed in the analysis, in comparison to 11,016 observations in the original version utilized by the authors.

The second data set is an extended version of the original data set with newly scraped data. All data was re-scraped from the forum, while the original time span was extended by three years, until April 2024, using the web scraping code provided by [Atanasov et al. \(2023\)](#). The retrieved HTML files were then cleaned using the corrected version of the R script provided in the replication package. The extended dataset contains 14,832 observations, representing an increase of approx-

imately one quarter.³ In summary, our analysis is thus based on three distinctive cleaned datasets: the original version provided by the authors, the newly cleaned version, and the updated and cleaned data.

In a second step, for each of the three datasets, the main variables were constructed and sample restrictions were implemented.⁴ We translated these steps from R to Stata. The majority of the procedures carried out by the authors were reproducible. Nevertheless, we consider three aspects worth mentioning. First, ([Atanasov et al., 2023](#), p. 867) state that each prize included in the *One Bid* game sample is ranked according to whether males or females are more likely to know the price of the item. However, while this is indeed the case when examining the raw data, the process of merging the rating data into the main dataset is defective in the replication package.

For 718 out of 10,886 items rated (approx. 7%), the rating score is not matched. This discrepancy is due to the fact that these items contain symbols such as " or ', which are coded differently in both datasets. This phenomenon is most prevalent in items such as laptops, for which display size is reported in inches. Therefore, it is unlikely that the number of misses is random. We corrected the coding error, resulting in a mean distance of the item to the gender of the participant of 1.0022 instead of 1.0007, an increase of approximately 1.5%. The revised data are then subjected to a robustness analysis, the results of which are presented in [Section 3](#). Second, the authors state that two ratings are considered per prize ([Atanasov et al., 2023](#), p. 868). It should be noted, however, that the rating data includes up to four ratings for each item, and all of these ratings are incorporated into the construction of the corresponding variable. While we ultimately decided not to change this specification as it did not seem to be critical to the setup, it is still worth noting. Third, [Atanasov et al. \(2023, p. 866\)](#) state that the distance of the leading bidder's bid from the target value is measured in relative terms, namely as a fraction of the target value. Yet, the authors actually employ the absolute distance. We address this issue in our analysis below in [Section 3](#). In addition to these points, there were

³In order to compute the inflation-adjusted target consistently, the CPI data was updated using the same Fred time series the authors used.

⁴Rounds with missing data and participants whose gender could not be inferred were excluded. Only decisions of fourth bidders considered.

some minor irregularities, such as inconsistent naming of variables, conflicting variable definitions throughout the script, or sample restrictions that were not precisely specified in the paper and thus required checking the R code provided. However, overall, the data set obtained by the authors was reproducible using the R script for data cleaning and translating the second step into Stata.

3 Replication Results and Replication with New Data

This section briefly describes the econometric models used for the main tables in [Atanasov et al. \(2023\)](#). The models presented are all linear probability models, which simplifies the interpretation of the coefficients. The authors state that their results are robust to an alternative logit specification ([Atanasov et al., 2023](#), p. 863) which we verify in [Appendix C](#).

As mentioned above, our replication is based on three separate datasets: (1) the .csv provided in the replication package with re-coding in Stata, (2) the raw data we prepared and cleaned ourselves with re-coding in Stata, and (3) an updated dataset with more recent episodes. The replications with datasets (1) and (2) should be considered computational replication. The extension that dataset (3) may be considered a direct replication, since we use the same methods as before but introduce 25% more observations.

3.1 Computational Replication Results

The paper has one primary outcome $cutoff_{ir}$, indicating whether the fourth contestant i overbids any of the previous contestants by exactly \$1 in a given round r . The main specification is in the form of a linear probability model (Table 1 in [Atanasov et al. \(2023\)](#), p. 863) and given as follows.

$$\begin{aligned} cutoff_{ir} = & \beta_0 + \beta_1 female_i + \beta_2 gendermatch_{ir} + \beta_3 itemvalue_r \\ & + \beta_4 priorcutoffs_{ir} + \beta_5 female_i \times gendermatch_{ir} + \varepsilon_{ir} \end{aligned} \quad (1)$$

Here $female_i$ represents a dummy variable which equals one if the fourth partic-

ipant is female, while $gendermatch_{ir}$ indicates whether the perceived leader, i.e. the one who is overbid, is of the same gender as the fourth participant. Moreover, $itemvalue_{ir}$ denotes the inflation adjusted price of the items currently at stake, and $priorcuttoffs_{ir}$ gives the number of cutoffs during the preceding rounds. Finally, ε_{ir} represents the error term.

Additional stricter variations of Equation (1) include round fixed effects, since each episode of *The Price is Right* encompasses up to 6 *One Bid* games, and gender composition effects to control for the composition of the other three bidders. The composition controls account for how many of the four players are men and women. Our computational replications produce similar point estimates and statistical significance as the original (see Table 2).

Table 2: Replication Table 1. Likelihood of Cutting off the Perceived Lead Based on Gender Match.

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Original (1)	Replication (2)	Original (3)	Replication (4)	Original (5)	Replication (6)	Original (7)	Replication (8)	Original (9)	Replication (10)	Original (11)	Replication (12)
Male	0.051*** (0.010)	0.051*** (0.010)	0.051*** (0.010)	0.052*** (0.010)	0.052*** (0.011)	0.053*** (0.011)	0.051*** (0.011)	0.052*** (0.011)	0.049*** (0.011)	0.049*** (0.011)	0.046*** (0.015)	0.046*** (0.015)
Gender Match	-0.045*** (0.010)	-0.045*** (0.010)	-0.049*** (0.010)	-0.049*** (0.010)	-0.047*** (0.010)	-0.047*** (0.010)	-0.047*** (0.010)	-0.047*** (0.010)	-0.045*** (0.010)	-0.044*** (0.010)	-0.047*** (0.014)	-0.047*** (0.014)
Target value							-0.000*** (0.000)	-0.000*** (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)
Prior Cutoff									0.052*** (0.005)	0.052*** (0.005)	0.052*** (0.005)	0.052*** (0.005)
Male x Gender match											0.006 (0.024)	0.006 (0.024)
Observations	11.016	11.018	11.016	11.018	10.133	10.134	10.133	10.134	10.133	10.134	10.133	10.134
Adjusted R ²	0.005	0.006	0.019	0.019	0.021	0.021	0.022	0.022	0.033	0.033	0.033	0.033
Round fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the replication using our own cleaning and original results of [Atanasov et al. \(2023\)](#) Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To further test the robustness of their results, the authors exploit the fact that some individuals play the *One Bid* game for several rounds. This specification, while highly restrictive in terms of sample size, allows to control for individual fixed effects (Table 2 in [Atanasov et al. \(2023\)](#), p. 865). The latter are represented as $individual_i$ in the equation below, which is an adapted version of Equation (1).

$$\begin{aligned}
 cutoff_{ir} = & \beta_0 + \beta_1 gendermatch_{ir} + \beta_2 itemvalue_r + \beta_3 priorcuttoffs_{ir} \\
 & + individual_i + \varepsilon_{ir}
 \end{aligned}
 \tag{2}$$

Again, stricter variations include round and gender composition fixed effects. We are able to replicate the results of the original study both in terms of the estimated coefficients and statistical significance (see Table 3).

Table 3: Replication Table 2. Likelihood of Cutting off Based on Gender Match, Controlling for Individual Fixed.

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Original (1)	Replication (2)	Original (3)	Replication (4)	Original (5)	Replication (6)	Original (7)	Replication (8)	Original (9)	Replication (10)
Gender Match	-0.040* (0.021)	-0.040* (0.021)	-0.046** (0.021)	-0.046** (0.021)	-0.041* (0.022)	-0.041* (0.022)	-0.041* (0.022)	-0.041* (0.022)	-0.045** (0.021)	-0.045** (0.021)
Target Value							-0.00003** (0.00001)	-0.000** (0.000)	-0.00002 (0.00001)	-0.000 (0.000)
Prior Cutoff									-0.238*** (0.015)	-0.238*** (0.015)
Observations	3,980	3,982	3,980	3,982	3,683	3,684	3,683	3,684	3,683	3,684
Adjusted R^2	0.234	0.234	0.270	0.270	0.266	0.266	0.268	0.268	0.352	0.352
Round fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the replication using our own cleaning and original results of [Atanasov et al. \(2023\)](#) Table 2, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior accounting for individual fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The final part of the analysis addresses the question of whether individuals are better at identifying a leading bid if they share the same gender as the leader (Table 3 in [Atanasov et al. \(2023\)](#), p. 866). Correctly identifying a leading bid is defined as the fourth bidder submitting a higher bid than the actual current leader, regardless of whether this bid exceeds the actual value of the item. Thus, the dependent variable is binary, taking on a value of 1 if the leader was correctly identified and 0 otherwise. The corresponding equation is given as follows.

$$\begin{aligned}
 LeaderIDD_{irt} = & \beta_0 + \beta_1 female_i + \beta_{2g} endermatch_{ir} + \beta_3 itemdistance_{ir} \\
 & + \beta_4 itemvalue_{ir} + \beta_5 priorcutoffs_{ir} + \varepsilon_{ir}
 \end{aligned}
 \tag{3}$$

All variables are defined as above, except for one new addition. $itemdistance_{ir}$ describes the distance between the leading bid (prior to the fourth bidder) and the item value as a fraction of the item value. As already mentioned, we find a minor coding error here. While the authors describe the variable in the dataset as a fraction of item price, it is just the absolute difference rather than the difference as a fraction of the price. We repeat the analysis with the difference as a fraction of the price. Our results reassuringly have the same sign as those initially reported, although the magnitude is different (see Table 4).

Table 4: Coding Error Table 3. Likelihood of Correctly Identifying the Lead Based on Gender Match.

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Original (1)	Corrected (2)	Original (3)	Corrected (4)	Original (5)	Corrected (6)	Original (7)	Corrected (8)	Original (9)	Corrected (10)
Male	0.001 (0.010)	0.001 (0.010)	0.0003 (0.010)	0.000 (0.010)	0.002 (0.011)	0.003 (0.011)	-0.001 (0.011)	0.002 (0.011)	-0.002 (0.011)	0.001 (0.011)
Gender Match	-0.014 (0.010)	-0.015 (0.010)	-0.015 (0.010)	-0.016 (0.010)	-0.014 (0.010)	-0.015 (0.010)	-0.012 (0.010)	-0.014 (0.010)	-0.012 (0.010)	-0.013 (0.010)
Distance lead to actual price	0.0001*** (0.00001)	0.496*** (0.025)	0.0001*** (0.00001)	0.500*** (0.025)	0.0001*** (0.00001)	0.496*** (0.026)	0.0002*** (0.00001)	0.504*** (0.027)	0.0002*** (0.00001)	0.497*** (0.027)
Actual price							-0.00004*** (0.00001)	-0.000 (0.000)	-0.00004*** (0.00001)	-0.000 (0.000)
Prior Cutoff									0.025*** (0.005)	0.025*** (0.005)
Observations	10,991	10,991	10,991	10,991	10,133	10,133	10,133	10,133	10,133	10,133
Adjusted R^2	0.015	0.035	0.015	0.035	0.014	0.034	0.016	0.034	0.019	0.036
Round FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the replication using our own cleaning and original results of [Atanasov et al. \(2023\)](#) Table 3, about the likelihood of correctly identifying the lead based on gender match. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The difference in magnitude primarily stems from the scaling of the distance variable. When expressing the distance as a fraction of the original price, the variable is bounded between 0 and 1. Hence, the coding error does not change any conclusions of the paper.

3.2 Updated Data Results

The replication package for [Atanasov et al. \(2023\)](#) included the webscraper and cleaning files that the authors used for the data collection. This enabled us to update the dataset to include data from 2021 onwards. We use the same webscraper and (fixed) cleaning file to ensure compatibility between the original dataset and the updated version. Following the data update, the observation count increases from 11,018 to 14,832 (a 34% increase).

Note that we are not able to replicate an additional variable of the “genderedness” of the products for logistical reasons. The original paper used *MTurk* workers to assess whether they believed that men or women were more likely to know the price of a given item. The responses were converted to a variable to capture expectations of whether a specific gender is more likely to submit an accurate bid. This variable was used in an additional robustness check and is not needed for the main tables.

The analysis of the updated results shows some changes in comparison with the

original results. We still find that contestants are less likely to cut-off a same-sex leader. While the new analysis of the likelihood of cutting the perceived leader based on gender gives similar results to the original study (see Table 5), the update affects the estimates and statistical significance of the original results in Tables 2 and 3. If controlling for individual fixed effects as in Table 2 of the original study, for the gender match we find a coefficient smaller than the original article and only statistically significant on the 10% significance level (see Table 6). Given the highly restrictive nature of individual fixed effects and the overall consistency of the results in comparison with the original study, this does not affect the interpretation or conclusion of the original study. Yet, to better understand these results, Appendix B) further explores changes over time.

Table 5: Replication Table 1. Likelihood of Cutting off the Perceived Lead Based on Gender Match with Extended Data.

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Original (1)	Extended Dataset (2)	Original (3)	Extended Dataset (4)	Original (5)	Extended Dataset (6)	Original (7)	Extended Dataset (8)	Original (9)	Extended Dataset (10)	Original (11)	Extended Dataset (12)
Male	0.051*** (0.010)	0.060*** (0.009)	0.051*** (0.010)	0.059*** (0.009)	0.052*** (0.011)	0.063*** (0.009)	0.051*** (0.011)	0.062*** (0.009)	0.049*** (0.011)	0.059*** (0.009)	0.046*** (0.015)	0.054*** (0.013)
Gender Match	-0.045*** (0.010)	-0.040*** (0.009)	-0.049*** (0.010)	-0.044*** (0.008)	-0.047*** (0.010)	-0.038*** (0.009)	-0.047*** (0.010)	-0.038*** (0.009)	-0.045*** (0.010)	-0.035*** (0.009)	-0.047*** (0.014)	-0.039*** (0.012)
Target value							-0.000*** (0.000)	-0.000*** (0.000)	-0.000** (0.000)	-0.000*** (0.000)	-0.000** (0.000)	-0.000*** (0.000)
Prior Cutoff									0.052*** (0.005)	0.057*** (0.004)	0.052*** (0.005)	0.057*** (0.004)
Male x Gender match											0.006 (0.024)	0.011 (0.021)
Observations	11,016	14,832	11,016	14,832	10,133	13,666	10,133	13,666	10,133	13,666	10,133	13,666
Adjusted R ²	0.005	0.006	0.019	0.018	0.021	0.020	0.022	0.020	0.033	0.033	0.033	0.033
Round fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the replication using our extended dataset and original results of [Atanasov et al. \(2023\)](#) Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Replication Table 2. Likelihood of Cutting off Based on Gender Match, Controlling for Individual Fixed with Extended Data.

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Original (1)	Extended Dataset (2)	Original (3)	Extended Dataset (4)	Original (5)	Extended Dataset (6)	Original (7)	Extended Dataset (8)	Original (9)	Extended Dataset (10)
Gender Match	-0.040* (0.021)	-0.027 (0.019)	-0.046** (0.021)	-0.033* (0.018)	-0.041* (0.022)	-0.034* (0.019)	-0.041* (0.022)	-0.034* (0.019)	-0.045** (0.021)	-0.032* (0.018)
Target Value							-0.00003** (0.00001)	-0.000*** (0.000)	-0.00002 (0.00001)	-0.000* (0.000)
Prior Cutoff									-0.238*** (0.015)	-0.236*** (0.013)
Observations	3,980	5,286	3,980	5,286	3,683	4,902	3,683	4,902	3,683	4,902
Adjusted R ²	0.234	0.233	0.270	0.266	0.266	0.265	0.268	0.267	0.352	0.347
Round fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the replication using our extended dataset and original results of [Atanasov et al. \(2023\)](#) Table 2, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior accounting for individual fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 examines whether participants’ ability to identify the leading player dif-

fers by gender. The original study reports two main findings: men and women are equally capable of correctly identifying the leading bidder and are neither more nor less likely to identify the leader if they are of the same gender. However, repeating this analysis with the extended dataset shows that having the same gender as the current leading bidder makes identification less likely. Our point estimate has the same sign and is comparable in magnitude to [Atanasov et al. \(2023\)](#), but is now significant at the 10% level, which was not the case before. With the updated data, contestants are marginally more likely to misidentify the leader if the current leader is of the same gender as themselves.

Table 7: Replication Table 3. Likelihood of Correctly Identifying the Lead Based on Gender Match with Extended Data.

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Original (1)	Extended Dataset (2)	Original (3)	Extended Dataset (4)	Original (5)	Extended Dataset (6)	Original (7)	Extended Dataset (8)	Original (9)	Extended Dataset (10)
Male	0.001 (0.010)	0.002 (0.009)	0.0003 (0.010)	0.002 (0.009)	0.002 (0.011)	0.000 (0.009)	-0.001 (0.011)	-0.002 (0.009)	-0.002 (0.011)	-0.004 (0.009)
Gender Match	-0.014 (0.010)	-0.015* (0.008)	-0.015 (0.010)	-0.015* (0.008)	-0.014 (0.010)	-0.018* (0.009)	-0.012 (0.010)	-0.016* (0.009)	-0.012 (0.010)	-0.015* (0.009)
Distance lead to actual price	0.0001*** (0.00001)	0.000*** (0.000)	0.0001*** (0.00001)	0.000*** (0.000)	0.0001*** (0.00001)	0.000*** (0.000)	0.0002*** (0.00001)	0.000*** (0.000)	0.0002*** (0.00001)	0.000*** (0.000)
Actual price							-0.00004*** (0.00001)	-0.000*** (0.000)	-0.00004*** (0.00001)	-0.000*** (0.000)
Prior Cutoff									0.025*** (0.005)	0.026*** (0.004)
Observations	10,991	14,792	10,991	14,792	10,133	13,666	10,133	13,666	10,133	13,666
Adjusted R ²	0.015	0.015	0.015	0.015	0.014	0.014	0.016	0.016	0.019	0.019
Round FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the replication using our extended dataset and original results of [Atanasov et al. \(2023\)](#) Table 3, about the likelihood of correctly identifying the lead based on gender match. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The main analysis found that men and women were both more likely to overbid an opposite-gender original contestant by \$1 if they believed them to be leading. In our re-analysis of the correctness of beliefs, see [Table 7](#), we find suggestive evidence that men and women are both equally less likely to accurately identify same-gender leaders. Given the similarities of the point estimates to the original article, this change in significance is likely caused by the increase in sample size. We can cautiously interpret this as suggestive evidence of incorrect beliefs about gender skills in the *One Bid* game.

Still, the main results in [Tables 2](#) and [5](#) are not affected by these results, as these analyses focused on whether or not contestants cut off the *perceived* lead - regardless of whether these beliefs were accurate. By submitting their bid above or below

those of the other players, the fourth player signals who they believe to be leading or not. In these results and in our replication, we find robust evidence of own-gender favoritism for a given belief, even if additional analysis reveals that players are slightly inaccurate in their assessments of own-gender competitors.

3.3 Robustness Analysis Results

To further test the robustness of the results reported in the original study, we conduct additional analyses in four dimensions. First, due to the time frame of 50 years, we re-estimate all tables with time fixed effects (see Appendix A). Second, to account for potential changes in the cutoff behavior over time, we re-analyse subsamples for the periods 1972 to 2011 and 2012 to 2024 (see Appendix B). Third, since the authors point out that the results are robust to the use of a logit model instead of a linear probability model, but not provide any results obtained when using logit estimations, we re-estimate all tables with logit models (see Appendix C). Fourth, as there are good arguments for using robust standard errors or standard errors clustered at the individual level, instead of the homoscedastic standard errors, we re-estimate all tables with robust standard errors or standard errors clustered at the individual level.

We find that the inclusion of time fixed effects reduces the magnitude of estimated coefficients in size and levels of statistical significance. We stress that these changes do not affect the interpretation of the results or the conclusion of the paper. The re-analysis of subsamples for the periods 1972 to 2011 and 2012 to 2024, indicates changes in the cutoff behavior over time. While men were more likely to cutoff perceived leaders, compared to women in the period of 1972 to 2011, there are no gender differences observable from 2012 to 2024. For all other results in the subsamples, we do not find any changes over time, again supporting the overall results of the original study. For the re-analysis using logit models, robust standard errors, as well as standard errors clustered on the individual level, we find no differences in the results compared to the original study.

In light of a wide range of robustness checks and minor changes in coefficients and statistical significance, our results confirm the robustness of the interpretation and

conclusion of (Atanasov et al., 2023).

4 Conclusion

We presented the results of a computational replication of Atanasov et al. (2023), producing the estimates with another statistical software language (Stata) rather than the language of the original replication package (R). The authors use fifty years of data from the American TV show *The Price is Right* to determine that men and women both show own-gender favoritism in a high-stakes competition taking place in a setting that cleanly separates beliefs and preferences. While we uncover some minor coding errors, they have no substantial effect on the results which we are able to fully reproduce. This was also the case for the estimates obtained using an extension of the dataset that includes a further 3 years of *One Bid* minigames. Additional robustness checks conducted cover the inclusion of time fixed effects, subsample analysis, the use of robust standard errors, as well as results obtained when employing logit models. While some of these robustness checks lead to smaller effect sizes and lower statistical significance, all results support the robustness of the original article.

References

- Atanasov, P., Dana, J. D., and Teeselink, B. K. (2023). Taste-Based Gender Favoritism in High-Stake Decisions: Evidence from the Price is Right. *The Economic Journal*, 134(February):856–883.
- Blevins, C. and Mullen, L. (2015). Jane, John . . . Leslie? A Historical Method for Algorithmic Gender Prediction. *Digital Humanities Quarterly*, 9(3):2–2.
- Brodeur, A., Mikola, D., Cook, N., and al. (2024). Mass reproducibility and replicability: A new hope. I4R Discussion Paper Series 107, Institute for Replication (I4R), Ottawa, Canada.

Appendix of
A comment on “Taste-based Gender Favouritism in
High-Stake Decisions: Evidence from the Price is
Right”

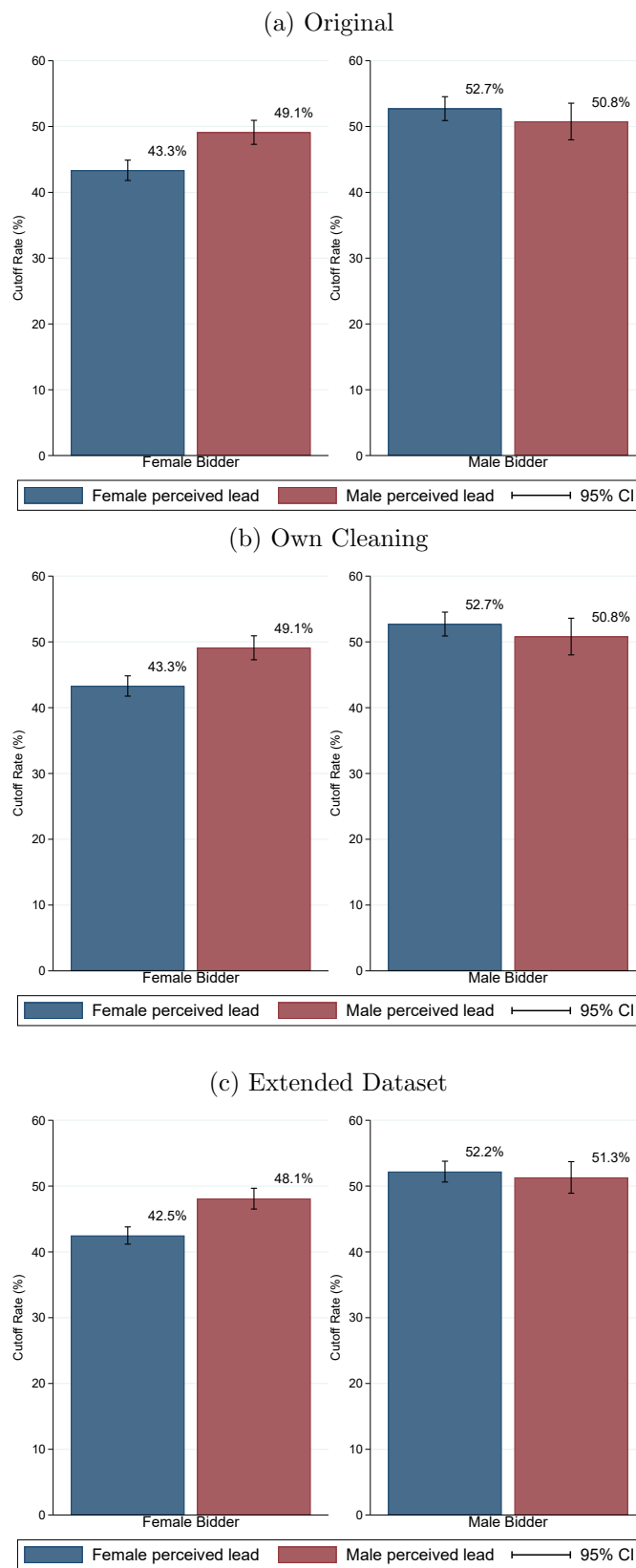
Julia F. Engel
U Kiel

Patrick Nüß
U Kiel

Meike Rudolph
U Kiel, IAB

Julia Schwarz
U Kiel

Figure A1: Cutoff Rates for Gendered Items



Note: This figure presents the cutoff rates for gendered items, comparing the original sample of Atanasov et al. (2023), our own cleaned dataset and the extended dataset.

A Time Fixed Effects Re-estimation

The dataset in [Atanasov et al. \(2023\)](#) covers a timespan of nearly 50 years, from the 1970s to the early 2020s. Figure A1 shows that there is evidence of learning as aggressive cutoff strategies are increasingly adopted. In the 1970s, under 20% of fourth bids used a cutoff strategy while in 2021 over 50% of all bids did so. At the same time, norms around women’s roles in society changed and may have given rise to different group dynamics. In [Atanasov et al. \(2023\)](#) Figure A6, the authors already illustrate that the gender composition of the four bidders changed over time. From 1970-1990 the bidders were over 70% female in the available episodes, before settling around 60% in the 2000s.

We repeat the main analysis in Equation 1 with year fixed effects to see whether changes over time have an impact on gender favoritism behavior. Because the data was scraped from a forum, not all episodes are covered, especially missing episodes from before internet forums were commonplace. This results in an over-representation of episodes recorded between 2000 and 2021. Years with fewer documented episodes may face precision issues when using time fixed effects. We attempt to mitigate these problems by running an additional analysis with decade fixed effects as a coarser time measure.

Table A1: Replication Table 1. Likelihood of Cutting off the Perceived Lead Based on Gender Match with Year Fixed Effects.

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Original (1)	Time FE (2)	Original (3)	Time FE (4)	Original (5)	Time FE (6)	Original (7)	Time FE (8)	Original (9)	Time FE (10)	Original (11)	Time FE (12)
Male	0.051*** (0.010)	0.032*** (0.010)	0.051*** (0.010)	0.033*** (0.010)	0.052*** (0.011)	0.026** (0.011)	0.051*** (0.011)	0.026** (0.011)	0.049*** (0.011)	0.027** (0.011)	0.046*** (0.015)	0.023 (0.015)
Gender Match	-0.045*** (0.010)	-0.027*** (0.010)	-0.049*** (0.010)	-0.032*** (0.010)	-0.047*** (0.010)	-0.034*** (0.010)	-0.047*** (0.010)	-0.034*** (0.010)	-0.045*** (0.010)	-0.034*** (0.010)	-0.047*** (0.014)	-0.037*** (0.014)
Target value							-0.00002*** (0.00001)	0.000 (0.000)	-0.00002** (0.00001)	0.000 (0.000)	-0.00002** (0.00001)	0.000 (0.000)
Prior cutoffs									0.052*** (0.005)	0.037*** (0.005)	0.052*** (0.005)	0.037*** (0.005)
Male x Gender Match											0.006 (0.024)	0.008 (0.024)
Observations	11,016	11,015	11,016	11,015	10,133	10,132	10,133	10,132	10,133	10,132	10,133	10,132
Adjusted R ²	0.005	0.032	0.019	0.042	0.021	0.044	0.022	0.044	0.033	0.050	0.033	0.050
Round fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Note: This table reports the replication using time fixed effects and original results of [Atanasov et al. \(2023\)](#) Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We find that the inclusion of time fixed effects reduces the magnitude of the gender match coefficient by around a fifth. The effect size change we report with our stricter reanalysis is comparable to other replications ([Brodeur et al., 2024](#)). In our most restrictive estimation, we estimate a coefficient of -0.037 compared to the

reported estimate of -0.047 . These findings hold when switching to decade-based fixed effects in an attempt to better pool episodes at the beginning of the sample period (see Table A2).⁵ We conclude that time fixed effects primarily account for the trend towards increased use of a cutoff strategy in later episodes, but do not change fundamental conclusions of own-gender favoritism.

Table A2: Original Table 1 - Robustness Check with Decade FE

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Male	0.034*** (0.010)	0.035*** (0.010)	0.027** (0.011)	0.027** (0.011)	0.027** (0.011)	0.025 (0.015)
Gender Match	-0.029*** (0.010)	-0.034*** (0.010)	-0.036*** (0.010)	-0.036*** (0.010)	-0.034*** (0.010)	-0.038*** (0.014)
Target value				0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Prior cutoffs					0.037*** (0.005)	0.040*** (0.005)
Male x Gender Match						0.007 (0.024)
Observations	11,016	11,016	10,133	10,133	10,133	10,133
Adjusted R^2	0.031	0.041	0.043	0.043	0.050	0.049
Round fixed effects	No	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the replication using decade time fixed effects and original results of [Atanasov et al. \(2023\)](#) Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

⁵The fixed effects were grouped in years by decades (1970-1979, 1980-1989, 1990-1999, 2000-2009, 2010-2019, 2020-2021).

B Subsample Re-estimation

The analysis of the extended dataset in Section 3.2 and the consideration of time fixed effects in Section A both indicate the sensitivity of the results to time. To understand the roots of this sensitivity, this section focuses on the analysis of subsamples of the fully updated dataset from 1972 to April 2024. For this reason, the following analysis splits the sample into two datasets with approximately equal sample sizes. The first sample consists of the period 1972 to 2004. The second sample consists of the remaining observations from 2005 to 2024.

Table A3 replicates the original Table 1 containing results about the likelihood of cutting off by perceived lead based on the same gender. In line with Figure A2, the results show that gender differences in cutoff behavior decreased over time. During the period from 1972 to 2011 men were about 8 percentage points more likely to cutoff compared to women. However, these gender differences in cutoff decreased over time and became statistically insignificant in the most recent subsample of 2012 to 2024, providing no longer any evidence for gender differences in the cutoff behavior. The estimates for the gender differences in the cutoff behaviour, based on the gender match between the two contestants over time, are not conclusive. While some models indicate an increase in the gender match, others indicate a reduction in the gender match over time. On the contrary, the point estimate of the gender match variable stays negative and statistically significant in all specifications. These patterns are therefore robust over the sample period, independent of changes in the cutoff behavior of men and women over time. The same patterns are observable in the case of a sample split using three equally sized sample periods instead of two (see Table A6). Overall, these results support the original findings by (Atanasov et al., 2023).

Given the reduced sample size and small effect sizes in the original study, the analysis of the subsamples for the original Tables 2 and 3 is less reliable. The subsample reanalysis in Table A4 and Table A5 both weakly indicate that the gender match increases in the more recent subsample from 2012 to 2024 in comparison to the one from 1972 to 2011. While these results all together indicate that the gender differences in cutoff behavior have changed over time, the results are still in line with the original study and therefore support the general conclusions of Atanasov et al. (2023).

Table A3: Period Subsample Replication for Table 1 (1972 to 2011, 2012 to 2024)

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	1972 to 2011 (1)	2012 to 2024 (2)	1972 to 2011 (3)	2012 to 2024 (4)	1972 to 2011 (5)	2012 to 2024 (6)	1972 to 2011 (7)	2012 to 2024 (8)	1972 to 2011 (9)	2012 to 2024 (10)	1972 to 2011 (11)	2012 to 2024 (12)
Male	0.081*** (0.013)	0.029** (0.012)	0.075*** (0.013)	0.033*** (0.012)	0.087*** (0.014)	0.016 (0.013)	0.087*** (0.014)	0.016 (0.013)	0.078*** (0.014)	0.018 (0.013)	0.071*** (0.019)	0.013 (0.018)
Gender match	-0.035*** (0.012)	-0.024** (0.012)	-0.038*** (0.012)	-0.029** (0.012)	-0.024* (0.013)	-0.033*** (0.013)	-0.024* (0.013)	-0.033*** (0.013)	-0.023* (0.013)	-0.031** (0.012)	-0.029* (0.017)	-0.036** (0.017)
Target value					-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Prior cutoffs									0.061*** (0.007)	0.039*** (0.006)	0.061*** (0.007)	0.039*** (0.006)
Male x Gender Match											0.015 (0.031)	0.012 (0.028)
Observations	7,128	7,704	7,128	7,704	6,678	6,988	6,678	6,988	6,678	6,988	6,678	6,988
Adjusted R^2	0.009	0.001	0.020	0.011	0.023	0.011	0.023	0.011	0.035	0.017	0.035	0.017
Round fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results using subsamples of our extended dataset replicating original Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior. The subsamples cover the periods 1972 to 2011 and 2012 to 2024. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Period Subsample Replication for Table 2 (1972 to 2011, 2012 to 2024)

	Model 1		Model 2		Model 3		Model 4		Model 5	
	1972 to 2011 (1)	2012 to 2024 (2)	1972 to 2011 (3)	2012 to 2024 (4)	1972 to 2011 (5)	2012 to 2024 (6)	1972 to 2011 (7)	2012 to 2024 (8)	1972 to 2011 (9)	2012 to 2024 (10)
Gender match	-0.016 (0.028)	-0.036 (0.025)	-0.030 (0.027)	-0.036 (0.025)	-0.026 (0.029)	-0.043 (0.027)	-0.027 (0.029)	-0.042 (0.027)	-0.017 (0.027)	-0.040 (0.025)
Target value					-0.000** (0.000)	-0.000 (0.000)			-0.000** (0.000)	-0.000 (0.000)
Prior cutoffs									-0.237*** (0.020)	-0.252*** (0.019)
Observations	2,442	2,844	2,442	2,844	2,297	2,605	2,297	2,605	2,297	2,605
Adjusted R^2	0.228	0.208	0.264	0.239	0.252	0.243	0.255	0.243	0.332	0.335
Round fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results using subsamples of our extended dataset replicating original Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior accounting for individual fixed effects. The subsamples cover the periods 1972 to 2011 and 2012 to 2024. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Period Subsample Replication for Table 3 (1972 to 2011, 2012 to 2024)

	Model 1		Model 2		Model 3		Model 4		Model 5	
	1972 to 2011 (1)	2012 to 2024 (2)	1972 to 2011 (3)	2012 to 2024 (4)	1972 to 2011 (5)	2012 to 2024 (6)	1972 to 2011 (7)	2012 to 2024 (8)	1972 to 2011 (9)	2012 to 2024 (10)
Male	0.003 (0.013)	0.002 (0.009)	0.002 (0.013)	0.001 (0.012)	0.002 (0.014)	-0.006 (0.013)	0.000 (0.014)	-0.005 (0.013)	-0.003 (0.014)	-0.004 (0.013)
Gender match	-0.012 (0.012)	-0.015* (0.008)	-0.013 (0.012)	-0.017 (0.012)	-0.012 (0.013)	-0.023* (0.013)	-0.012 (0.013)	-0.022* (0.012)	-0.011 (0.013)	-0.022* (0.012)
Distance lead to actual price	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Actual price										
Prior cutoffs										
Observations	7,123	7,669	7,123	7,669	6,678	6,988	6,678	6,988	6,678	6,988
Adjusted R^2	0.017	0.015	0.017	0.014	0.015	0.014	0.016	0.017	0.018	0.021
Round fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

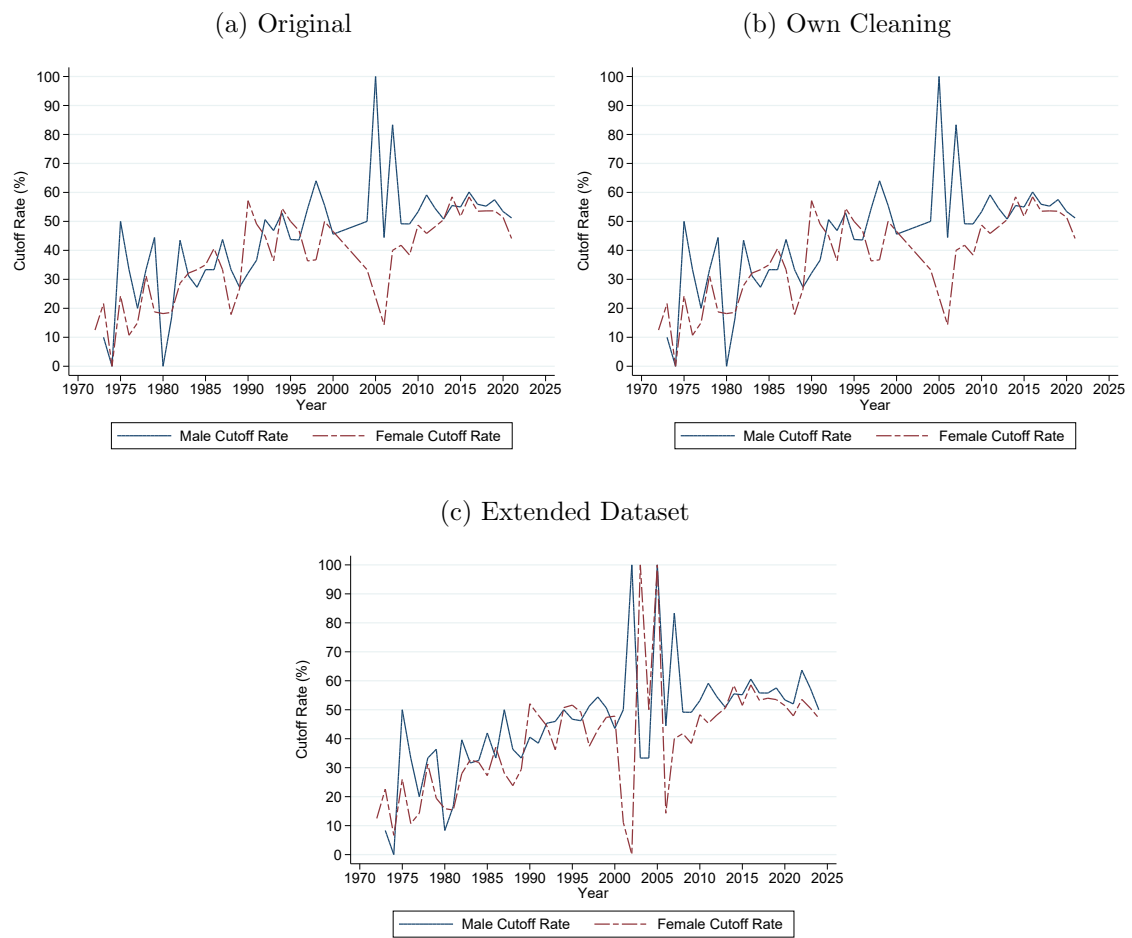
Note: This table reports the results using subsamples of our extended dataset replicating original Table 3, about the likelihood of correctly identifying the lead based on gender match. Standard errors in parentheses. The subsamples cover the periods 1972 to 2011 and 2012 to 2024. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Period Subsample Replication for Table 1 (1972 to 2004, 2005 to 2016, 2017 to 2024)

	Model 1			Model 2			Model 3			Model 4			Model 5					
	1972 to 2004 (1)	2005 to 2016 (2)	2017 to 2024 (3)	1972 to 2004 (4)	2005 to 2016 (5)	2017 to 2024 (6)	1972 to 2004 (7)	2005 to 2016 (8)	2017 to 2024 (9)	1972 to 2004 (10)	2005 to 2016 (11)	2017 to 2024 (12)	1972 to 2004 (13)	2005 to 2016 (14)	2017 to 2024 (15)	1972 to 2004 (16)	2005 to 2016 (17)	2017 to 2024 (18)
Male	0.066*** (0.016)	0.048*** (0.014)	0.038*** (0.015)	0.062*** (0.016)	0.047*** (0.014)	0.042*** (0.015)	0.075*** (0.017)	0.038*** (0.015)	0.028 (0.017)	0.074*** (0.017)	0.038*** (0.015)	0.028 (0.017)	0.064*** (0.017)	0.039*** (0.015)	0.030* (0.017)	0.064*** (0.024)	0.031 (0.022)	0.024 (0.024)
Gender match	-0.040*** (0.015)	-0.015 (0.014)	-0.029* (0.015)	-0.044*** (0.015)	-0.021 (0.014)	-0.034*** (0.015)	-0.027* (0.016)	-0.021 (0.015)	-0.039** (0.016)	-0.027* (0.016)	-0.021 (0.015)	-0.039** (0.016)	-0.028* (0.016)	-0.020 (0.015)	-0.037** (0.016)	-0.029 (0.021)	-0.026 (0.019)	-0.042* (0.022)
Target value																		
Prior cutoffs																		
Male x Gender Match																		
Observations	4,871	5,426	4,535	4,871	5,426	4,535	4,600	4,902	4,074	4,600	4,902	4,074	4,600	4,902	4,074	4,600	4,902	4,074
Adjusted R ²	0.007	0.002	0.002	0.007	0.014	0.010	0.022	0.016	0.010	0.022	0.016	0.010	0.032	0.025	0.017	0.031	0.025	0.017
Round fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results using subsamples of our extended dataset replicating original Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior. The subsamples cover the periods 1972 to 2004, 2005 to 2016, and 2017 to 2024. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A2: Cutoff Rates over Time



Note: This figure presents the development in the cutoff behavior by gender over time, comparing the original sample of [Atanasov et al. \(2023\)](#), our own cleaned dataset and the extended dataset.

Table A7: Logit Specification Table 1 on Probability of Cutting Off Perceived Leader

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Male	0.205*** (0.041)	0.208*** (0.041)	0.214*** (0.045)	0.210*** (0.045)	0.201*** (0.045)	0.190*** (0.064)
Gender match	-0.180*** (0.040)	-0.200*** (0.040)	-0.193*** (0.043)	-0.192*** (0.043)	-0.185*** (0.043)	-0.194*** (0.057)
Target value				-0.000*** (0.000)	-0.000** (0.000)	-0.000** (0.000)
Prior cutoffs					0.217*** (0.021)	0.217*** (0.021)
Male x Gender Match						0.023 (0.100)
Observations	11,016	11,016	10,133	10,133	10,133	10,133
Round fixed effects	No	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	Yes	Yes	Yes	Yes

Note: This table reports the replication using logit models and the original linear probability model results of [Atanasov et al. \(2023\)](#) Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Logit Re-estimation

We re-estimate the main tables with an alternative logit specification rather than a linear probability model. Using a logit model has the advantage that the model design bounds the predicted outcomes between 0 and 1. Yet, it comes at the expense of an easy interpretation of the coefficients. We use the same control variables as in the main tables.

The first table studied how likely contestants were to overbid the perceived leading player by exactly \$1 and whether this varied with being of the same gender as the leader. The original results (p. 863 and replicated in [Table 2](#)) found that men were approximately 5 percentage points more likely than women to employ a cutoff strategy. Moreover, the probability of a \$1 cutoff increased if the strategy had been used in an episode's prior round. Finally, cutoffs were less common for higher value items. If a fourth bidder had the same gender as the perceived leader, they were 4.5 percentage points less likely to use a cutoff on average.

Table 2 in [Atanasov et al. \(2023\)](#) continues to study how a gender match with the perceived leader affects the fourth bidder's probability of overbidding by exactly \$1. However, the authors employ a more restrictive specification here by using individual fixed effects. The *One Bid* game starts with calling four members from the

audience and playing the game for one round. The round’s winner then leaves the quartet and a replacement is called (except in the last round). The remaining three members keep their position in the bidding order. The game is played for 6 rounds total. The average fourth round player in our dataset has 1.5 games on record.

To use individual fixed effects, the sample is restricted to contestants who played in more than one round. This brings down the sample to 3,980 fourth bids from 11,016 that are used in the main analysis. When conducting a logit analysis with individual fixed effects, an additional difficulty emerges, as a significant share (57%) of players do not vary in their choice to cutoff or not. These contestants have no within-variation in outcomes and thus an additional 2,278 observations are dropped to make a logit estimation feasible. This procedure reduces the sample to 18% of the initial size, which affects the statistical power.

Table A8: Logit Specification Table 2 with Individual Fixed Effects

	(1)	(2)	(3)	(4)	(5)
	Model 1	Model 2	Model 3	Model 4	Model 5
Gender match	-0.206*	-0.259**	-0.220*	-0.224*	-0.302*
	(0.111)	(0.119)	(0.126)	(0.126)	(0.159)
Target value				-0.000**	-0.000**
				(0.000)	(0.000)
Prior cutoffs					-2.468***
					(0.185)
Observations	1,702	1,702	1,501	1,501	1,501
Round fixed effects	No	Yes	Yes	Yes	Yes
Gender composition FE	No	No	Yes	Yes	Yes

Note: This table reports the replication using logit models and the original linear probability model results of [Atanasov et al. \(2023\)](#) Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior accounting for individual fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The original second table as well as our replication (see [Table 3](#)) confirmed the patterns initially described in [Table 2](#), though the significance levels were affected by the heavy sample size reduction and limited power. Sharing the same gender as the perceived leading bidder had an average 4-4.6 percentage point lower likelihood of using a cutoff strategy. Our logit results in [Table A8](#) with a smaller and more restrictive sample broadly confirm this pattern based on the sign of the coefficient and are still statistically significant on at least a 10% level.

The third table addresses the concern whether the lower incidence of same-gender cutoffs are influenced by difficulties in identifying the leading bidder or stereotypes.

The dependent variable is whether the fourth bidder found the correct person to overbid.⁶ The results indicate that this is not the case: male and female fourth bidders are equally likely to correctly identify the current leader. Similarly, sharing the same gender as the leader does not make a successful identification more likely. Moreover, higher priced items make it more difficult, and larger gaps between the leader's guess and the true item value are positively associated with correctly identifying the leader.

When running the additional analysis as a logit regression, the fourth variation of Equation 3 does not converge with Stata's log-likelihood approach. This specific regression includes controls for male fourth bidders, whether the bidder shares the same gender as the current leader, the gap between the leader and the item value, the value of the target. Additional controls include round and gender composition fixed effects. The model does not converge, despite the next variation with an additional variable of how many cutoffs had previously been issued converging again without any problems.

Table A9: Logit Specification Table 3 on Probability of Correctly Identifying the Leader

	(1)	(2)	(3)	(4)	(5)
	Model 1	Model 2	Model 3	Model 4	Model 5
Male	-0.003 (0.041)	-0.005 (0.041)	-0.002 (0.045)		-0.018 (0.045)
Gender match	-0.055 (0.040)	-0.059 (0.040)	-0.058 (0.043)		-0.046 (0.043)
Distance lead to actual price	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)		0.001*** (0.000)
Target value					-0.000*** (0.000)
Prior cutoffs					0.096*** (0.020)
Observations	10,991	10,991	10,133		10,133
Round fixed effects	No	Yes	Yes		Yes
Gender composition FE	No	No	Yes		Yes

Note: This table reports the replication using logit models and the original linear probability model results of [Atanasov et al. \(2023\)](#) Table 3, about the likelihood of correctly identifying the lead based on gender match. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

⁶For example: Assume an item value is \$100. The bids, in order are 40, 80, 110. The second bidder is leading with 80, since the third bidder overbid. Any bid the fourth bidder submits in the interval [81;109] implies they correctly inferred that the second bidder is leading and the third overbid.

D Standard Errors

The standard errors reported in [Atanasov et al. \(2023\)](#) are homoscedastic standard errors. We run the same regressions with the original data provided in the .csv to verify the reliability and validity of the original findings, we distinguish between the general choice of robust standard errors and standard errors clustered at the individual level. The choice between the two clustering options arises from the fact that approximately a third of observations play more than one *One Bid* game. Regular robust clustering argues that the observations are independent of one another, but some observations have up to four games on record.

We show that the standard errors are nearly identical to the ones reported in [Atanasov et al. \(2023\)](#). The results are listed in a three-way comparison in Tables [A10](#), [A11](#), and [A12](#). In some models, the third decimal place changes, but this has no bearing on the significance levels. The results in all three main tables are robust to an alternative choice of standard errors.

Table A10: Sensitivity Table 1 to Choice of Standard Errors

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6		
	Original (1)	Ind. Clustering (2)	Robust (3)	Original (4)	Ind. Clustering (5)	Robust (6)	Original (7)	Ind. Clustering (8)	Robust (9)	Original (10)	Ind. Clustering (11)	Robust (12)	Original (13)	Ind. Clustering (14)	Robust (15)	Original (16)	Ind. Clustering (17)	Robust (18)
Male	0.051*** (0.010)	0.051*** (0.011)	0.051*** (0.010)	0.051*** (0.010)	0.051*** (0.011)	0.051*** (0.010)	0.052*** (0.011)	0.052*** (0.011)	0.052*** (0.011)	0.051*** (0.011)	0.051*** (0.011)	0.051*** (0.011)	0.052*** (0.011)	0.052*** (0.011)	0.052*** (0.011)	0.052*** (0.015)	0.052*** (0.014)	0.052*** (0.016)
Gender match	-0.045*** (0.010)	-0.045*** (0.010)	-0.045*** (0.010)	-0.045*** (0.010)	-0.045*** (0.010)	-0.045*** (0.010)	-0.047*** (0.011)	-0.047*** (0.011)	-0.047*** (0.010)	-0.047*** (0.011)	-0.047*** (0.011)	-0.047*** (0.010)	-0.045*** (0.011)	-0.045*** (0.011)	-0.045*** (0.010)	-0.047*** (0.014)	-0.047*** (0.014)	-0.047*** (0.014)
Target value									-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
Prior cutoffs													0.052*** (0.005)	0.052*** (0.005)	0.052*** (0.005)	0.052*** (0.005)	0.052*** (0.005)	0.052*** (0.005)
Male x Gender match																	0.006 (0.024)	0.006 (0.024)
Observations	11,016	11,016	11,016	11,016	11,016	11,016	10,133	10,133	10,133	10,133	10,133	10,133	10,133	10,133	10,133	10,133	10,133	10,133
Adjusted R ²	0.005	0.005	0.005	0.019	0.019	0.019	0.021	0.021	0.021	0.022	0.022	0.022	0.033	0.033	0.033	0.033	0.033	0.033
Round fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the replication using robust standard errors and standard errors clustered at the individual level and homoscedastic standard errors as in the original results of [Atanasov et al. \(2023\)](#) Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Sensitivity Table 2 to Choice of Standard Errors

	Model 1			Model 2			Model 3			Model 4			Model 5		
	Original (1)	Ind. Clustering (2)	Robust (3)	Original (4)	Ind. Clustering (5)	Robust (6)	Original (7)	Ind. Clustering (8)	Robust (9)	Original (10)	Ind. Clustering (11)	Robust (12)	Original (13)	Ind. Clustering (14)	Robust (15)
Male	-0.040* (0.021)	-0.040* (0.021)	-0.040* (0.022)	-0.046** (0.021)	-0.046** (0.021)	-0.046** (0.021)	-0.041* (0.022)	-0.041* (0.022)	-0.041* (0.023)	-0.041* (0.022)	-0.041* (0.022)	-0.041* (0.023)	-0.045** (0.021)	-0.045** (0.021)	-0.045** (0.021)
Target value									-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
Prior cutoffs													-0.238*** (0.015)	-0.238*** (0.016)	-0.238*** (0.016)
Observations	3,980	3,980	3,980	3,980	3,980	3,980	3,683	3,683	3,683	3,683	3,683	3,683	3,683	3,683	3,683
Adjusted R ²	0.234	0.234	0.234	0.270	0.270	0.270	0.266	0.266	0.266	0.268	0.268	0.268	0.352	0.352	0.352
Round fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the replication using robust standard errors and standard errors clustered at the individual level and homoscedastic standard errors as in the original results of [Atanasov et al. \(2023\)](#) Table 1, about the estimated effect of the gender match of the fourth bidder with the perceived leader and their cutoff behavior accounting for individual fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Sensitivity Table 3 to Choice of Standard Errors

	Model 1			Model 2			Model 3			Model 4			Model 5		
	Original (1)	Ind. Clustering (2)	Robust (3)	Original (4)	Ind. Clustering (5)	Robust (6)	Original (7)	Ind. Clustering (8)	Robust (9)	Original (10)	Ind. Clustering (11)	Robust (12)	Original (13)	Ind. Clustering (14)	Robust (15)
Male	0.001 (0.010)	0.001 (0.010)	0.001 (0.010)	0.000 (0.010)	0.000 (0.010)	0.000 (0.010)	0.001 (0.011)	0.001 (0.011)	0.001 (0.011)	-0.001 (0.011)	-0.001 (0.011)	-0.001 (0.011)	-0.002 (0.011)	-0.002 (0.011)	-0.002 (0.011)
Gender match	-0.014 (0.010)	-0.014 (0.010)	-0.014 (0.010)	-0.015 (0.010)	-0.015 (0.010)	-0.015 (0.010)	-0.014 (0.010)	-0.014 (0.010)	-0.014 (0.010)	-0.012 (0.010)	-0.012 (0.011)	-0.012 (0.010)	-0.011 (0.010)	-0.011 (0.010)	-0.011 (0.010)
Distance lead to actual price	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Actual price										-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
Prior cutoffs													0.025*** (0.005)	0.025*** (0.005)	0.025*** (0.005)
Observations	10,992	10,992	10,992	10,992	10,992	10,992	10,134	10,134	10,134	10,134	10,134	10,134	10,134	10,134	10,134
Adjusted R ²	0.015	0.015	0.015	0.015	0.015	0.015	0.014	0.014	0.014	0.016	0.016	0.016	0.019	0.019	0.019
Round fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender composition FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table reports the replication using robust standard errors and standard errors clustered at the individual level and homoscedastic standard errors as in the original results of [Atanasov et al. \(2023\)](#) Table 3, about the likelihood of correctly identifying the lead based on gender match. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.