How Well Do Women Sell?
New Evidence for Non-Professionals

Carina Fleischer[a]   Holger Kraft[b]   Farina Weiss[c]

This version: January 12, 2024

Abstract: This paper is the first to analyze the performance of non-professional individuals selling personal belongings. We study a novel hand-collected data set from a popular German TV show and find that women obtain on average about 7.3% less than men. This gap cannot fully be explained by known moderators for professionals. We document a novel relationship between age, education and negotiation outcomes of females. In particular, we find that midlife women working in the industry perform as well as men, whereas most other women obtain much less than men. Remarkably, female teams perform significantly better than single females.

Keywords: gender gap, auction, negotiation, performance, teams

JEL subject codes: D12, D44, D91, J16

[a] Faculty of Economics and Business, Goethe University, Theodor-W.-Adorno-Platz 3, 60323 Frankfurt am Main, Germany.
E-mail: `c.fleischer@finance.uni-frankfurt.de`

[b] Faculty of Economics and Business, Goethe University, Theodor-W.-Adorno-Platz 3, 60323 Frankfurt am Main, Germany. Phone: +49 (0) 69 798 33 699.
E-mail: `holgerkraft@finance.uni-frankfurt.de`

[c] Faculty of Economics and Business, Goethe University, Theodor-W.-Adorno-Platz 3, 60323 Frankfurt am Main, Germany.
E-mail: `farina.weiss@hof.uni-frankfurt.de`.

1 Introduction

The literature has documented several gender gaps with the gender wage gap being a prominent example.[1] Furthermore, there is a large body of research showing gender differences in negotiation outcomes.[2] In this context, many studies focus on situations from the business world in which the subjects are professionals (e.g., performance of salesman) or in which individuals negotiate work-related topics (e.g., labor market outcomes).[3] Few studies analyze the performance of non-professionals when they have to deal with daily-life decisions.[4] To the best of our knowledge, the performance of non-professional individuals selling personal belongings has not been studied yet. This is an important research gap since the majority of economic transactions is carried out by non-professionals. Further, there is no reason to believe that the results documented in the literature for the selling performance of professionals also hold for non-professionals. In fact, we find that some moderators explaining negotiating outcomes of professionals also apply to non-professionals, whereas others do not.

We use a novel data set that is hand-collected from the popular German television show "Bares für Rares" (BfR).[5] The selling process is organized in form of an English auction. However, the seller is present and communication between dealers and seller during the auction is possible. The seller can thus influence the flow of bidding by making comments about the currently highest bid.

First, we document a significant *gender revenue gap*. After controlling for a host of covariates, single females receive 7.3% less than single males. Since the show "Bares für Rares" has a high regular market share of more than 20%, this gender gap is witnessed by more than two million viewers.

We show that not all women lose to the same extent, but seller characteristics matter. In particular, we document a U-shaped pattern for age, i.e., younger and elderly women obtain significantly less than men. This result is new and can be attributed to the competing effects of experience vs. gender-role perception in different cohorts. To support this finding, we provide

---

[1]See, e.g., Blau and Kahn (2017).

[2]See, e.g., Bowles and McGinn (2008); Kray and Thompson (2005); Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015); Stuhlmacher and Walters (1999); Walters, Stuhlmacher, and Meyer (1998).

[3]See, e.g., Day (1993); Levy and Sharma (1994); Stevens, Bavetta, and Gist (1993); Bowles, Babcock, and McGinn (2005).

[4]Ayres (1991) and Ayres and Siegelman (1995) study car purchases and Goldsmith-Pinkham and Shue (2023) analyze investments in real estate.

[5]The English translation of the title of the show would be "cash for rarities" or "cash for rare items".

evidence that the level of education alone cannot explain the performance of women. Instead, negotiation experience can be an important person-based moderator of gender differences in negotiation outcomes.[6] We show that midlife women having this experience perform as well as men, whereas most other women obtain less than men. This result has important policy implications since it suggests that better negotiation results might be achieved by making women better acquainted with negotiation situations through training courses.

Also, attractive women perform better than less attractive women, while there is no such effect for men. This is a new finding since for professionals attractiveness matters for both sexes.[7] For two seller characteristics, optimism and sentiment, we do not find gender differences. Men and women both benefit to the same extent from a high level of optimism and sentiment.

Regarding buyer characteristics, women lose significantly when confronted with pushy or inactive dealers. Interestingly, a significant gender gap also arises if women are assigned to generous dealers, which suggests that women are not making the most of their opportunities.[8] This finding points towards our earlier conclusion that negotiation training courses might empower women to achieve better performances.

Focusing on item characteristics, there is a significant gender gap if women sell items that are incongruent with the female gender role. Analogously, women obtain significantly less when selling an item from a category for which the dealers' valuation has a large dispersion.[9] Controlling for both dimensions (gender-role congruence and dispersion of valuation), we find that the gender gap is greatest if women sell female-incongruent items with a high disperse valuation.[10]

Finally, we compare the relative performance of single women with teams consisting of two females or couples where the female is the seller. We show that teaming up is beneficial for females. In particular, teams of two females perform as well as single males, which is in line with an advocacy explanation as in Bowles, Babcock, and McGinn (2005) or Amanatullah and Morris (2010). These authors show that women perform better if they negotiate on behalf of

---

[6]See, e.g., Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015).

[7]See, e.g., Mobius and Rosenblat (2006) and Solnick and Schweitzer (1999).

[8]This finding is in line with results by Daubman, Heatherington, and Ahn (1992) and Gould and Slone (1982) who document that negotiating women behave more modestly than men.

[9]We measure dispersion by the standard deviation of our performance measure across a particular item category. See Section 8 for details.

[10]These results complement and extend findings by Bear and Babcock (2012) and Hüffmeier, Freund, Zerres, Backhaus, and Hertel (2014), among others, showing that the topic of a negotiation and ambiguity concerning negotiation opportunities can be an important moderator of gender differences in negotiation outcomes.

others.

Our data set covers about 1,700 selling processes. The total revenue collected by the participants in our sample is about 1,490,000 euros. The median value of the items to be sold is 500 euros whereby the most expensive item is a painting with a sales price of 30,500 euros. Our data is unique in the sense that it is non-experimental and heterogeneous along several dimensions: The participants of the show offer a great variety of items such as china, jewelry, furniture, paintings, art, vases, tableware, clothes, and toys to a five-person podium of professional dealers. The participants are between 18 to 93 years old and have varying educational and occupational backgrounds (e.g., worker, manager, school teacher, public servant, retiree, farmer, housewife).

The five-person podium of professional dealers is of changing composition and we are able to measure their characteristics such as generosity, activeness, pushiness, and average dealer age. We thus have a data set with a lot of heterogeneity across items, buyers and sellers.

The remainder of the paper is organized as follows: Section 2 summarizes the related literature. Section 3 describes the concept of the show. Section 4 explains our data set. In Section 5, we develop our hypotheses. Section 6 presents our benchmark regressions. Section 7 analyzes the effects of seller characteristics. Section 8 provides results if we focus on different item characteristics. Section 9 studies the effects of dealer characteristics. Section 10 compares the performance of single females to female teams. Section 11 concludes.

2 Related Literature

Our paper contributes to the strand of literature examining gender differences in negotiation behavior and outcomes. Past research has repeatedly shown that women tend to achieve less favorable negotiation agreements than men (e.g., Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015); Stuhlmacher and Walters (1999); Walters, Stuhlmacher, and Meyer (1998)). The gender revenue gap documented in our paper is in line with these findings. Researchers attribute those gender differences in negotiation outcomes to differences in the overall level of assertiveness and competitiveness in men's and women's negotiation behavior (e.g., Walters, Stuhlmacher, and Meyer (1998)). As an example of this, negotiating women have been found to behave more modestly than men (e.g., Daubman, Heatherington, and Ahn (1992); Gould and Slone (1982)) and to set lower goals than their male counterparts (e.g., Kray, Thompson, and Galinsky (2001); Major and Konar (1984)). We do not find evidence for the latter, but

3

document that women perform worse if they are assigned to generous dealers, which could be attributed to modesty.

To explain this different negotiation behavior, early studies focused on inherent personality differences between men and women (Maccoby and Jacklin (1974)). Those studies find that men are more prone to risk-taking (e.g., Croson and Gneezy (2009); Eckel and Grossman (2008)) and display a greater tendency to overconfidence than women (e.g., Lundeberg, Fox, and Punccohar (1994); Beyer (1990)), while women tend to feel a lack of entitlement (Callahan-Levy and Messé (1979); Major, McFarlin, and Gagnon (1984)).

Recent studies bypassed the issue of whether women and men are inherently different but instead focused on social role theory (Eagly (1987); Eagly and Karau (2002)) as a theoretical basis to explain gender differences in negotiation behavior. Social roles are socially shared expectations regarding the appropriate behavior of men and women in society (Eagly and Wood (2012)). They are not only descriptive, but also prescriptive. Hence, violating a given social role leads to negative evaluations and treatment by others, i.e., *social backlash* (Rudman (1998); Rudman and Glick (1999); Rudman and Phelan (2008); Damnals, Zlatev, Halevy, and Neale (2021)). Negotiating is generally consistent with the masculine stereotype of the agentic, assertive and dominant man, but contradicts normative expectations of women as being other-oriented, passive, caring and supportive (Amanatullah and Morris (2010); Bowles, Babcock, and Lai (2007); Eagly (1987); Wade (2001)). Consequently, when compared to men, women face a mismatch between their gender role and effective negotiation behavior. As shown by Kray and Thompson (2005) and Kulik and Olekalns (2012), women behaving in accordance with their gender role and avoiding potential backlash are likely to show ineffective negotiation behavior and achieve poor economic outcomes. This is in line with our finding that women lose more if they exhibit high sentiment scores.

However, as documented by Bowles, Babcock, and Lai (2007) and Tinsley, Cheldelin, Schneider, and Amanatullah (2009), women departing from their gender role and negotiating assertively are perceived to be less likable and socially unskilled compared to men showing the same behavior. Further, Amanatullah and Morris (2010) show that women are concerned that negotiating like men will lead to backlash and hence are reticent to negotiate assertively in the first place. Likewise, others suggest that women's poor negotiation performance stems from expectations of the negotiating counterpart of how women have to behave (e.g., Kray and Thompson (2005)). In a controlled set of field studies Ayres (1991) and Ayres and Siegelman (1995) find that female car buyers were offered worse deals than male car buyers. We complement these results and

show that women perform worse if they sell items that are incongruent with the female gender role.

Another strand of literature examines situational factors that moderate gender differences in negotiation outcomes (e.g., Amanatullah and Morris (2010); Amanatullah and Tinsley (2013); Bowles, Babcock, and McGinn (2005); Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015); Stuhlmacher and Linnabery (2013)). Bowles, Babcock, and McGinn (2005) and Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015) find that gender differences are most likely to arise in situations with a high degree of ambiguity regarding the appropriate negotiation behavior. According to Bowles, Babcock, and McGinn (2005), in strong or unambiguous situations (Mischel (1977)), there are clearly known protocols of expected behavior that people are willing to follow such that the influence of individual differences is constrained by the situational cues. Weak situations, however, do not constrain behavior by protocols so that individual differences like gender are more likely to affect behavior. In an empirical study, Bowles, Babcock, and McGinn (2005) compared men's and women's negotiation outcomes when the bargaining range was told to the negotiators (low ambiguity) or when it was ambiguous (high ambiguity). They find that gender differences vanished in the low ambiguity situation, while men outperformed women in the high ambiguity situation. Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015) obtain similar results in a meta-analysis. In line with these findings, we show that women perform worse during the auction (high ambiguity) and as well as men in the post-auction renegotiations (low ambiguity).

Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015) further document that negotiation experience reduces ambiguity in a negotiation. Consequently, they find that gender differences in negotiation outcomes were reduced when negotiators had negotiation experience. Complementing these results, we show that women who are used to negotiate due to their profession perform relatively better. Furthermore, we find that young women do significantly worse than midlife women who have more experience in negotiating.

Bowles, Babcock, and McGinn (2005) and Amanatullah and Morris (2010) propose that advocacy, i.e., negotiating on behalf of others, reduces gender differences in negotiation outcomes. While negotiating for one's own self-interest is stereotypically agentic, negotiating on behalf of others fits communal behavior and is therefore congruent with the female gender role of caring. In an empirical study, Bowles, Babcock, and McGinn (2005) and Amanatullah and Morris (2010) find that women negotiating on behalf of others secured significantly higher negotiation outcomes and acted more assertively than women negotiating on their own behalf.

5

Amanatullah and Tinsley (2013) further show that women negotiating for others fear less risk of experiencing backlash and are punished instead if they do not negotiate assertively enough. In line with these results, we document that female teams perform significantly better than single females.

Meta-analytic evidence further shows that gender differences favoring men are greater when the subject of negotiation is less gender-role congruent for women (Reif, Kugler, and Brodbeck (2019); Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015)). Bear and Babcock (2012) compare men's and women's performances when negotiating about the price of either motorcycle headlamps (masculine stereotyped) or lamp beads for jewelry (feminine stereotyped). When assigned to negotiate about motorcycle parts, men obtained significantly better agreements than women, while there were no significant gender differences in negotiations over jewelry beads. We contribute to this literature by showing that women perform better when they sell items being congruent with the female gender role.

Further, Stuhlmacher and Walters (1999) show that men outperform women in single-issue negotiations, while differences in negotiation agreements between men and women are mitigated in multi-issue negotiations where mutual gains are possible. Bowles, Babcock, and Lai (2007) find that the gender composition of the negotiation dyad also affects gender differences in negotiation outcomes. They show that women were more reluctant to negotiate when the negotiation counterpart was male. Further, women negotiating with a male counterpart were more likely to experience backlash than women negotiating with a female opponent. Kugler, Reif, Kaschner, and Brodbeck (2018) document in a meta-analysis that gender differences in negotiations have diminished over time. They argue that the male and female gender role have become more aligned over time due to ongoing emancipation thus reducing the inconsistency between the female gender role and the negotiator role. We provide evidence that complements these findings: First, women lose relatively more if the average age of the male dealers is high. Second, elderly women do significantly worse than midlife women.

3   Concept of the Show

BfR is a show in which candidates can sell their personal belongings to a group of professional dealers who buy for their own accounts. Typically, in every episode five to six potential single sellers or seller teams of two persons are presented. We refer to these potential sellers as
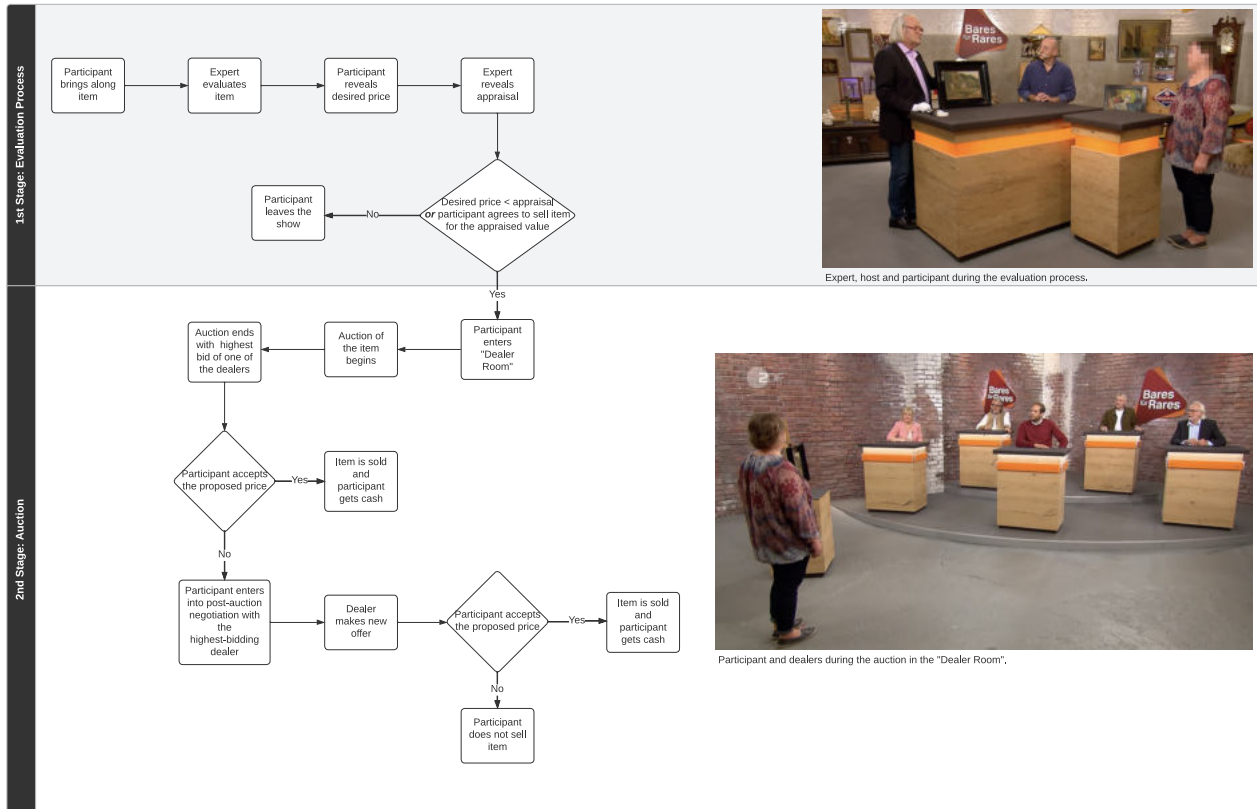
Figure 1: **Selling Process.** This figure depicts the selling process during which the item is appraised and the participant offers the item to a five-person podium of competing professional dealers.

participants. We now describe the show using a *single female participant as an example.*

The show consists of two stages that are shown in Figure 1: the evaluation of the item by a neutral expert and the sale of the item in an auction. The evaluation process starts with the participant coming to the show and bringing along her item. After some small talk between the participant and the host, the expert starts to examine the item. The expert describes the item in detail and provides the participant with in-depth historical, craft, or artistic information.[11] Before the expert states an approximate monetary valuation of the item, the host asks the participant for the desired price she wants to realize. Afterwards, the expert reveals the appraisal. Only if either the desired price is lower than the appraisal or the participant explicitly states that she would be willing to sell her item for the (lower) appraised value, she is allowed to enter the second stage of the show, i.e., the auction. Otherwise the participant must leave the show. However, almost all participants pass this first stage. In particular, in more than 70% of the

---

[11]There are in total nine different experts, each being specialized on certain kinds of items.

cases, the desired price turns out to be below the appraised value.

In the second stage of the show, the participant enters the so-called "dealer room", where the auction of the item takes place. The participant faces a five-person podium of competing professional dealers ("group of dealers"). The composition of these groups of dealers varies from show to show and the participant does not know a priori which group of dealers she is assigned to. The dealers can inspect the item about 20 minutes before the start of the auction for the first time and have no information about the appraisal. The auction itself is conducted in the form of an open-outcry English auction with the special feature that the participant is present during the auction. There is no separate auctioneer who is leading the auction but one of the five dealers is typically taking over this duty. All of the dealers interested in the item can cry out their bids and overbid each other. There are no special rules regarding the stepsize between bids, number of bids etc. As the participant is present during the auction, she can interact with the dealers and hence eventually influence the flow of the bidding. For instance, she can comment on the bids and indicate whether she is already satisfied or not, she can push the dealers to bid further or she can reveal the appraisal. Similarly, also the dealers have the chance to talk to the participant during the auction and, for instance, ask for the appraisal or the desired price. However, the participant is not required to reveal these pieces of information.

The auction ends with the highest bid of one of the dealers. If none of the other dealers is willing to challenge the standing bid, the participant is asked whether she wants to sell the item for the proposed price of the highest bidder. In case she accepts the offer, she receives the amount offered in cash from the highest bidding dealer and the item is sold. The participant then leaves the show. However, sometimes the participant is not satisfied with the proposed price and tells the dealers that she wants more. This induces some post-auction negotiation between the participant and the highest-bidding dealer. The dealer can then make new (higher) offers to the participant. In case she accepts the highest offer, the item is sold and the participant receives the corresponding amount in cash. Otherwise the participant leaves the show without selling the item. However, in more than 95.6% of the cases the participant accepts the offer and a sale happens. Therefore, the impasse rate is below 4.4%.

4 Data

We have watched and analyzed more than 305 episodes of the German television series *Bares für Rares*, which amounts to about 265 hours of video footage. The term "Bares für Rares" means "cash for rarities" or "cash for rare items" in English. This series is broadcasted by Zweites Deutsches Fernsehen (ZDF) which is one of two major public television channels in Germany.[12] The series has received prestigious awards and is very popular also among younger television viewers according to the audience ratings. Its regular market share is about 20%. Although the majority of the watchers is above 50 years old (probably also because of the time 3pm when the show is broadcasted[13]) and the average participant has an age of 55 years, the approval ratings for the age group of 14 to 49 years is close to 10 percent, which is a high value for this age group.[14] Our sample contains all episodes of the year 2021 as well as older episodes that were available between 24 June 2021 and 24 August 2021 in the video-on-demand system called *Mediathek*.[15]

Our data set consists of 1,693 selling processes. We have 552 selling processes with a single female participant and 627 selling processes with a single male participant. Furthermore, there are 514 selling processes with teams of two individuals where the majority are couples with 302 observations. In the case of seller teams, the owner of the item is usually revealed during the introduction. For single sellers, we have almost always information about the age of the participants. This is not the case for teams where the ages are almost never revealed (for reasons that we do not know). Other than that, the available pieces of information for single sellers and teams are the same.

We have 74 selling processes that do not end with a sale. In the case of single females (males), only 23 (29) participants do not sell. Therefore, there is no systematic difference in impasse rates between females and males. The small impasse rates are not surprising since participants do not have strong outside options. Furthermore, every participant obtains a fair and objective appraisal of her/his item at the beginning of the selling process.

We group the items presented by the participants into 22 categories. Table 1 reports the

---

[12]The major public television channels are financed by a mandatory fee and advertisement. Every household in Germany owning a TV must pay this fee.

[13]Notice that the show is replayed after 6pm and is also available as video on demand for free.

[14]https://www.quotenmeter.de/n/115319/bares-fuer-rares-bleibt-mit-neuem-2020-quotenrekord-auf-erfolgskurs.

[15]At a certain time point during that period not all episodes were available, since episodes have an expiration date. On average, there were about 170 available per day during that period.

No	Item	Wom	Man	Total	Prop Wom	Adjusted	Item_fem	Item_dis
1	Art	6	10	16	0.375	0.404	0	0
2	Bronze	18	24	42	0.429	0.459	0	0
3	China	24	27	51	0.471	0.501	1	0
4	Clothes	5	5	10	0.500	0.531	1	1
5	Coin	5	4	9	0.556	0.586	1	1
6	Collect	6	34	40	0.15	0.166	0	1
7	Device	13	34	47	0.277	0.302	0	1
8	Femitem	8	7	15	0.533	0.564	1	1
9	Flatware	4	3	7	0.571	0.601	1	1
10	Furniture	56	99	155	0.361	0.390	0	0
11	Glass	12	7	19	0.632	0.660	1	0
12	Jewelry	179	73	252	0.710	0.735	1	0
13	Not	23	29	52	0.442	0.473	0	1
14	Painting	39	69	108	0.361	0.390	0	1
15	Promotion	4	16	20	0.200	0.220	0	1
16	Silvergold	37	25	62	0.597	0.626	1	0
17	Smoke	7	8	15	0.467	0.497	0	1
18	Tableware	16	12	28	0.571	0.601	1	1
19	Toy	31	67	98	0.316	0.343	0	1
20	Travel	7	7	14	0.500	0.531	1	1
21	Vase	13	10	23	0.565	0.595	1	1
22	Watch	16	28	44	0.364	0.392	0	0
	Total	529	598	1127			490	509

Table 1: **Summary Statistics for Item Categories**. This table reports summary statistics concerning the frequency of the different item categories. The columns labeled "Wom" and "Man" report the number of selling processes where the participant is female or male. The column labeled "Prop Wom" reports the proportion of women selling an item from a particular category. The last two columns show a classification into items that are congruent with the female gender role and another classification into items where the valuation on the side of the dealers is disperse. More details on these classifications can be found in Section 8.

frequencies of these groups in our sample. The first column gives the number that we assign to a particular category. The most common items are furniture, jewelry, and paintings (number 10, 12, and 14). Table 2 describes these item categories in more detail.

To evaluate the outcome of an auction, we use a relative performance measure as our main dependent variable. We refer to it as Perform and define

$$\text{Perform} = \frac{\text{Final}}{\text{Appraisal}}, \tag{4.1}$$

where "Final" is the final price and "Appraisal" is the appraisal value provided by an expert before the participant enters the dealer room. Perform can be interpreted as the relative revenue of a seller. If it is bigger (smaller) than one, then the seller receives more (less) than the appraised value.

Table 3 reports summary statistics concerning the participants and the selling processes. We see that there is no systematic difference in the average age of men and women coming to the show. The items brought along by men are slightly more expensive than those brought in by women as measured by the appraised value. The variable "Desired" stands for the desired price that the participants must reveal before the appraisal is revealed by the expert. If a participant gives a range, then we take the average of the bounds. The variable "Appraisal" is the appraised value. If the expert gives a range, then we take the average of the bounds.[16] The participant is only allowed to enter the dealer room if either Desired is smaller than Appraisal or the participant agrees to lower her/his desired price to the appraised value. Therefore, we calculate Trunc_des_app, which is the ratio between the potentially reduced desired price divided by the appraised value. Notice that Trunc_des_app is almost the same for men and women. Furthermore, we see that the sample average of Ln_perform is 0.014 higher for men than for women. Later on, we will show that this gap widens if we control for various covariates. The variable "Hammer" is the hammer price of the auction. More precisely, it is the highest price reached at the end of the auction before any potential renegotiation between the participant and the highest bidding dealer takes place. The variable "Ham_app" is the ratio between the hammer price and the appraised value. It can be seen that its logarithm denoted by Ln_ham_app is on average 0.021 higher for men than for women, i.e., the gap is slightly higher than for Ln_perform.

In our sample, we have a pool of 18 different dealers. They specialize in different kinds of item

---

[16]We have performed robustness checks showing that this assumption is not crucial.

Item Category	Description
Art	All kinds of art works attributed to famous artists, e.g., sculptures, busts, art posters
Bronze	All kinds of bronzes in different forms and designs
China	Items made of china by famous brands like Meissen or KPM, e.g., figurines, plates
Clothes	All kinds of textile items and accessories, e.g., costumes, sunglasses, uniforms,walking sticks
Coin	All kinds of collector coins (numismatic), often of gold
Collect	All kinds of collector's items with an intrinsic value, e.g., autographs, medals, trophies
Device	All kinds of technical and mechanical items, e.g., cameras, automats, telescopes
Femitem	Items that are typically associated with females, e.g., handbags, perfume bottles, powder compact
Flatware	All kinds of cutlery, e.g., spoons
Furniture	All kinds of interior furnishing and decoration, e.g., chairs, tables, lamps, sculptures, figurines
Glass	Items made of glass, e.g., pots, decanters
Jewelry	All kinds of jewelry, e.g., necklaces, rings, earrings, broochs
Not	Odd objects or very special items that do not fit in with all the other categories mentioned
Painting	All kinds of paintings, e.g., oil paintings, aquarelles, lithographs
Promotion	All items related to advertisement, e.g., enamel plates, placards
Silvergold	Items made of silver or gold, e.g., spoons, bowls, pots
Smoke	Items related to smoking, e.g., cigarette cases, cigarette lighters
Tableware	All kinds of dishes, e.g., tea or coffee set, teapot
Toy	All kinds of toys (mainly for males), e.g., Matchbox cars, parlor games, model railway
Travel	All kinds of items associated with traveling, e.g., suitcases
Vase	All kinds of vases of different brands and materials. Often famous designers like Galle
Watch	All kinds of wristwatches and fob watches. Often famous brands like Rolex

Table 2: **Item Descriptions.** This table reports descriptions of the kinds of items that the respective category contains.

	Wom	Man	Overall
Age	57.13	56.72	56.91
Desired	683.01	734.16	710.13
Appraisal	956.80	983.06	970.74
Final	881.82	914.98	899.41
Hammer	854.44	891.83	874.28
Trunc_des_app	0.671	0.673	0.672
Ln_perform	-0.012	0.002	-0.004
Ln_ham_app	-0.078	-0.057	-0.066
Obs.	529	598	1127

Table 3: **Summary Statistics for Participants and Selling Processes**. This table reports the averages of relevant variables. Age is the age of a participant. Desired stands for the desired price that a participant reveals at the beginning of the selling process. Appraisal is the appraisal value of an item which is provided by a neutral expert. Final stands for the sales price of an item. Hammer is the hammer price of an item at the end of the auction (before potential renegotiations). Trunc_des_app is the ratio between Desired and Appraisal truncated at one. Ln_perform is defined as the logarithm of (4.1). Ln_ham_app is the logarithm of the ratio between Hammer and Appraisal.

categories, but also buy items outside of their main fields. Table 4 reports summary statistics about the dealers. A group of dealers consists of five different dealers from the pool. In total, we observe 62 different groups of dealers.[17] Those groups are certainly male dominated, i.e., in every group there are at most two female dealers and often even only one female dealer.

5   Hypotheses

In the following, we build on the insights of the previous literature to derive testable hypotheses regarding gender differences and potential moderators in competitive negotiations. As documented in this literature, women tend to achieve less favorable negotiation agreements in competitive single-issue negotiations (e.g., Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015); Stuhlmacher and Walters (1999); Walters, Stuhlmacher, and Meyer (1998)). The participants of BfR are confronted with a competitive, single-issue bargaining situation. We therefore hypothesize that women will achieve poorer economic outcomes than men. We quantify these outcomes by the performance measure (4.1).

Hypothesis 1: There is a performance gap between women and men.

---

[17]Section 9 provides more details about the groups of dealers.

Name	Short	Avg Ask	Avg Bid	Age	Appear
Christian Vechtel	Chri	0.021	0.612	46	420
Daniel Meyer	Dani	0.030	0.820	48	434
David Suppes	Supp	0.027	0.692	33	448
Elke Velten	Elke	0.013	0.587	68	623
Esther Ollick	Esth	0.019	0.422	41	206
Fabian Kahl	Fabi	0.005	0.609	30	1032
Friedrich Häusser	Frie	0.000	0.568	68	88
Jan Cizek	JanC	0.044	0.540	45	272
Julian Schmitz-Avila	Juli	0.045	0.645	35	770
Lisa Nüdling	Lisa	0.026	0.396	41	429
Ludwig Hofmaier	Ludw	0.006	0.686	79	156
Markus Wildhagen	Mark	0.007	0.763	55	410
Roman Runkel	Roma	0.004	0.418	60	467
Steve Mandel	Stev	0.025	0.586	67	237
Susanne Steiger	Susi	0.004	0.458	38	684
Thorsden Schlößner	Thor	0.011	0.508	59	266
Walter Lehnertz	Wald	0.050	0.574	54	846
Wolfgang Pauritsch	Wolf	0.004	0.960	49	672

Table 4: **Dealer Characteristics**. This table reports summary statistics of all dealers. It provides the number of appearances (Appear) and the relative frequency that a particular dealer asks for the desired price or the appraisal (Avg Ask). Besides, Avg Bid measures the relative frequency that a dealer is active and bids. Finally, Age is a dealer's age as of October 2021. The table also reports the shortcut of the dealer's name (Short).

Furthermore, we want to test whether gender differences can be reduced by certain moderators. Bowles, Babcock, and McGinn (2005) and Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015) show that the degree of structural ambiguity mitigates or amplifies gender differences. In the setup of BfR, there is some room for ambiguity. In general, we hypothesize that all forms of ambiguity lead to worse outcomes for women.

Hypothesis 2: The performance gap between men and women increases if ambiguity increases.

BfR contains at least three types of ambiguity. We thus formulate three hypotheses targeting these forms of ambiguity.

First, as depicted in Figure 1, the second stage of a selling process can be further subdivided into *two steps*: The first step is the auction, the second step involves the potential renegotiation with the highest bidder. It is clear that the first step is more complex since it involves multiple dealers and advances at a faster pace. By contrast, during the second step the participant is only interacting with one dealer and has more control over the flow of actions. This leads to different *levels of ambiguity* in both steps. According to Bowles, Babcock, and McGinn (2005)

14

and Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015), gender differences are most likely to arise in situations with a high degree of ambiguity regarding the appropriate negotiation behavior. We thus hypothesize that the relative outcome is worse in the first step than in the second step. We can measure both outcomes separately since we have information about the hammer price of the auction and the final price (after renegotiating). The difference between the two prices measures the outcome of the renegotiation.

Hypothesis 2a: The performance gap between men and women opens up during the auction and is less pronounced during the renegotiation with the highest bidder.

The second form of ambiguity is *item-related* and arises from the type of item to be sold. We expect to observe different sizes of variation in the dealers' assessments of the items depending on the item category. One reason might be that items of some categories are easier to value since there are objective criteria. The best examples are jewelry or silverware. In both cases, the assessment is typically determined by the value of the raw materials. On the other hand, collectibles like autographs are harder to evaluate. This leaves room for negotiation that participants must realize and exploit. Since women are more modest than men (Daubman, Heatherington, and Ahn (1992); Gould and Slone (1982)) and tend to premature-concession making (Curhan, Neale, Ross, and Rosencranz-Engelmann (2008); Hüffmeier, Freund, Zerres, Backhaus, and Hertel (2014)), we hypothesize that the gender gap is more pronounced for item categories for which such opportunities exist.

Hypothesis 2b: The performance gap between men and women is larger for item categories for which the assessments of the dealers are more disperse.

Finally, although the economic structure of the selling process is known beforehand and participants know the appraisal value, they do not know how generous the group of dealers is they are assigned to. This generates uncertainty about the dealers' negotiation limits and which price is achievable given the particular group of dealers. Hence, we hypothesize that the performance gap between men and women is more pronounced for generous groups of dealers. We measure generosity by the average outcome of all auctions observed for the particular group of dealers. This is a form of ambiguity that is driven by the *environment* as for instance analyzed by Bowles, Babcock, and McGinn (2005). It is also related to the question whether women are able to exploit opportunities or whether they are more modest than men. In other words, women might not exploit their opportunities if there is ambiguity about the outcome.

Hypothesis 2c: The performance gap between men and women increases if women are assigned

to generous groups of dealers.

Following Bear (2011) and Bear and Babcock (2012), we further hypothesize that the negotiation topic can function as a moderator. In the BfR context, the negotiation topic is characterized by the specific item at sale. There are some item categories that are more congruent with the female gender role (i.e., jewelry or tableware), whereas others are less congruent with the female gender role (i.e., paintings, technical devices or toys).[18] We thus hypothesize that women perform relatively better if the item at sale is congruent with the female gender role.

Hypothesis 3: The performance gap between men and women is smaller if the item at sale is congruent with the female gender role.

Notice that Hypothesis 2 and 3 are connected. This can best be explained by using the examples of jewelry and paintings. Jewelry generates little ambiguity since the value of the raw material is known. At the same time, this is a category that is congruent with the female gender role. Both facts should mitigate the gender gap. On the other hand, paintings are less congruent with the female gender role, but still have rather objective values given by past auction results of the particular artist. Therefore, we expect women to perform worse if they sell paintings as opposed to jewelry.

Age proxies for at least two dimensions: Elderly people might be more experienced with negotiations since they might have done that multiple times in their (work) lives. As documented by Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015), Zerres, Hüffmeier, Freund, Backhaus, and Hertel (2013) and Eagly and Wood (1999), negotiation experience enhances the understanding of the negotiation tasks and thus helps to reduce the gap in negotiation performance between men and women. Hence, the gender gap in negotiation performance could decrease for elderly experienced women. On the other hand, older participants have been more exposed to gender-role issues during their life time. As shown by Diekman and Eagly (2000) and Twenge (1997), gender roles are dynamic, with the female gender role becoming more agentic over time due to ongoing emancipation. Hence, men and women of different generations are exposed to different gender role norms. With respect to BfR, we therefore hypothesize that the performance gap is mitigated for young women, but is more pronounced for elderly women. Since both dimensions work in opposite directions, the joint effect of experience and gender-role issues is not clear. The net effect might also be driven by the type of education, since people working in certain professions might be more used to negotiate. Notice however that a higher

---

[18]For instance, the art market (in particular concerning paintings) is male-dominated. This is documented by Adams, Kräussl, Navone, and Verwijmeren (2021), who show that female artists receive lower prices.

level of education does not always lead to more negotiation experience (e.g., medical doctor vs. sales person). To summarize, depending on the particular importance of age, education, and negotiation experience multiple performance patterns could arise.

Hypothesis 4: The performance gap between men and women is age-dependent and education-dependent, but not necessarily monotonic.

Furthermore, Mobius and Rosenblat (2006) and Solnick and Schweitzer (1999) provide empirical evidence that attractiveness matters for outcomes of negotiations. They find that attractive people reach better negotiation agreements than unattractive people. A priori, it is not clear whether there is a difference between males and females.

Hypothesis 5: Attractiveness matters for the performance of participants.

In tough environments, women are forced to actively negotiate, which is conflicting with the gender role assigned to women (e.g., Bowles, Babcock, and Lai (2007); Eagly (1987); Wade (2001)). Therefore, we hypothesize that women perform relatively worse than men if the selling environment is "tough". We measure the "toughness" of a selling environment by dealer characteristics such as inactiveness or pushiness.

Hypothesis 6: Women perform relatively worse in tough selling environments.

Advocacy has also been shown to mitigate gender differences in negotiation outcomes (e.g., Amanatullah and Morris (2010); Amanatullah and Tinsley (2013); Bowles, Babcock, and McGinn (2005)). In BfR, women can team up with another woman. This can be considered as an example of an advocacy setting since the second woman clearly negotiates on behalf of the first woman. We hypothesize that the performance gap between men and women decreases when a woman teams up with another woman.

Hypothesis 7: The performance gap between men and women is smaller for female teams.

6   Benchmark Results

We first present our benchmark results to test *Hypothesis 1*. Table 5 reports the estimates of two regressions.[19] The dependent variable is given by the natural logarithm of the performance measure "Perform" as defined in (4.1). We refer to it as "Ln_perform".[20] Both regressions

---

[19]All our regression results report heteroscedasticity consistent standard errors in the hc3 sense as derived by MacKinnon and White (1985). See Long and Ervin (2000) for an overview and additional references.

[20]We winsorize Ln_perform at the 1% level, but our results are robust to this specification.

include controls for the appraisal values and for the item categories. For the latter, we use the categorization given in Table 1 and include 22 dummies, one dummy for each category. Regression (2) further includes dummies for the different groups of dealers to control for the environment in which participants sell their items. As participants are randomly assigned to those groups of dealers, this is an exogenous variable.

Regression (1) documents a revenue gap of about 6.4% for females, i.e., the average revenue of a single female is more than 6% smaller than for a single male. The results are significant at the 5% level. When including dummies for the groups of dealers, the revenue gap becomes even more pronounced with women receiving 7.3% less than men. The results are significant at the 1% level. As groups of dealers constitute an important part of the exogenous selling environment, it is reasonable to control for them. We thus regard regression (2) as our benchmark specification given by:[21]

$$
\begin{aligned}
\text{Ln\_perform} \quad = \quad & \beta_0 + \beta_1 \text{Single\_woman} + \beta_2 \text{Ln\_appraisal} \\
& + \sum_{n=1}^{22} \beta_{3n} \text{Item}_n + \sum_{i=1}^{62} \beta_{4i} \text{DealerGroup}_i + \sum_{t=2016}^{2021} \beta_{5t} \text{Year}_t + \epsilon,
\end{aligned}
\tag{6.1}
$$

where Ln_appraisal is the natural logarithm of the appraised value. The other explanatory variables, Single_woman, Item, DealerGroup, and Year, are dummies for sex, item, group of dealers, and the year when the particular episode is produced. The error term is denoted by ϵ. Our results confirm *Hypothesis 1* and are in line with findings in the literature showing that women obtain worse negotiation outcomes than men (e.g., Mazei, Hüffmeier, Freund, Stuhlmacher, Bilke, and Hertel (2015); Stuhlmacher and Walters (1999)). As documented in the literature, in general women dislike negotiations more than men. Since potential participants actively apply, only women who are comfortable with competitive single-issue negotiations participate in the show. The gender gap for single females in our paper is thus a conservative estimate of the gap that arises for the whole population.[22]

---

[21]An alternative specification that is equivalent to (6.1) uses the logarithm of the final price as dependent variable. We have however decided to use (6.1), since this makes it easier to test whether the performance of a seller depends on the value of the item. More precisely, if we find that β_2 is not significantly different from zero, then the outcome of a selling process is independent of the value of the item. Otherwise, the performance of a seller depends on the value.

[22]There could also be a concern that the selection process of the TV production company affects our results. However, it is difficult to explain why the selection process leads to a gender gap for single women on the one hand, but at the same time the same selection process does not cause such a gap for female teams as documented in Section 10.

	(1)	(2)
Single_woman	-0.064**	-0.073***
	(-2.58)	(-2.88)
Ln_appraisal	-0.184***	-0.186***
	(-14.00)	(-13.61)
R^2	0.246	0.317
Dealer group dum.	no	yes

Table 5: **Benchmark Regressions**. The table reports the results of OLS regressions of Ln_perform on selected variables. Ln_perform is the natural logarithm of the ratio between the final price and the appraisal value. Both regressions are based on 1,127 observations of selling processes. There are 598 male participants and 529 female participants. Both regressions involve year dummies and item dummies. Regression (2) also involves dummies for the groups of dealers. Both regressions involve intercepts that are not reported. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

As a next step, we wish to understand whether performance differences can be attributed to the auction or to the renegotiation that might occur after the auction has ended. We thus focus on the men's and women's auction results before potential renegotiations. This addresses *Hypothesis 2a*. Hence, we use the natural logarithm of the ratio between the hammer price of the auction and the appraised value as dependent variable ("Ln Hammer Price over Appraisal"). The hammer price is the highest price reached at the end of the auction before any renegotiation between the highest bidding dealer and the participant takes place.[23] Using this ratio as dependent variable, we rerun our two benchmark regressions. The results are shown in Table 6. The gender gap arising during the auction is even more pronounced than the gender revenue gap documented by the regressions of Table 5. Regression (4) shows that the auction results for women are on average 8.4% worse than the outcomes for men (compared to 7.3% in (2)). This suggests that the gender gap is not further widening during post-auction renegotiations[24] and confirms *Hypothesis 2a* saying that women do worse during the auction than during the renegotiation.

7 Seller Characteristics

We now examine how certain seller characteristics affect negotiation performance and potentially moderate gender differences. Precisely, we focus on age, education, attractiveness, optimism, and sentiment.

---

[23]We winsorize "Ln Hammer Price over Appraisal" at the 1% level.

[24]Notice that slightly more women in our sample renegotiate than men.

	(3)	(4)
Single_woman	-0.075***	-0.084***
	(-2.78)	(-3.10)
Ln_appraisal	-0.197***	-0.200***
	(-14.59)	(-14.37)
R^2	0.229	0.302
Dealer group dum.	no	yes

Table 6: **Regressions with "Hammer over Appraisal"**. The table reports the results of OLS regressions of the natural logarithm of the ratio between the hammer price and the appraisal value on selected variables. The hammer price is the highest price reached in the auction before any potential renegotiations between the highest bidding dealer and the participant. Both regressions are based on 1,178 observations of selling processes where 51 observations do not end with a sale. There are 626 male participants and 552 female participants. Both regressions involve year dummies and item dummies. They also involve intercepts that are not reported. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.05, **p < 0.01, ***p < 0.001$.

7.1 Age and Education

First, we focus on the effects of age and education on female performance. We start by analyzing both dimensions separately. We sort participants into three age groups and six educational levels. However, since age and education might be systematically related, we also group the participants along both dimensions (age and education) and then study the relative performances of women in the resulting 18 groups.

For the age groups, we have chosen the ranges 18-39 (young), 40-59 (midlife), and 60-99 (elderly).[25] The rational is that the group 18-39 contains young adults getting an education or being at an early stage of their professional career. The group 40-59 consists of participants that have already longer work experiences and typically have reached their career peaks. The group 60-99 contains individuals close to or in retirement. Table 7 provides summary statistics concerning age and education.

First, if we only add age-group dummies to our benchmark regression (2), then the female dummy hardly changes and is 0.071 being significant at all levels. Therefore, after additionally controlling for age, women still lose about 7.1% compared to men.[26] Second, we compare the performance of women in a particular age group to the performance of men. We can do this in two ways: On the one hand, we can compare women in a certain age group to men in the same

---

[25]A finer classification would lead to small group sizes below the age of 40 since the average age of the participants is about 55 years.

[26]Notice that we do not have information about the age of 12 participants. So our regressions involving age are based on 1,116 observations.

Educ	18-39	40-59	59-99	Total	18-39	40-59	59-99	Total
	All				Proportion Wom			
1	11	29	15	55	0.18	0.52	0.33	
2	80	208	286	574	0.59	0.49	0.57	
3	37	80	68	185	0.49	0.30	0.31	
4	23	50	102	175	0.39	0.34	0.35	
5	0	4	38	42		0.25	0.50	
6	25	1	0	26	0.44			
Total	176	372	509	1,057				
	Wom				Man			
1	2	15	5	22	9	14	10	33
2	47	102	164	313	33	106	122	261
3	18	24	21	63	19	56	47	122
4	9	17	36	62	14	33	66	113
5	0	1	19	20	0	3	19	22
6	11	0	0	11	14	1	0	15
Total	87	159	245	491	89	213	264	566

Table 7: **Summary Statistics for Age and Education**. This table reports summary statistics concerning the frequency of participants with a certain educational level and in a certain age group. The upper left panel gives the numbers for all participants that we can classify according to the ISCO-08 classification. The upper right panel gives the relative frequency of women in a certain category. There are few empty spots since we do not observe old students or young retirees. The lower panels report the numbers for women and men separately. We were able to classify 1,057 males or females. For 122 single participants, we have too little or inconclusive information.

age group or we can compare these women to all men (independent of the men's ages).

To compare the performance of men and women in the same age group, we extend our benchmark regression (2) by replacing the female dummy "Single_woman" by six dummies (one for every age group and sex). The first two columns of Table 8 show the respective estimates of the dummy variables, while the third column shows the differences between men's and women's performance in a particular age group. Here elderly males are used as the reference group. Hence, their dummy is zero. Furthermore, to obtain the results where all men are the reference group, we extend our benchmark regression (2) replacing the female dummy by three female categories for the three age groups plus a fourth dummy for all men who constitute the reference group. The last column of Table 8 labeled "All Men" reports the corresponding results.

It can be seen that in both cases a U-shaped pattern arises. Younger and elderly women lose, whereas women in the midlife group do not lose significantly when compared to men. Apparently, they do as well as men of the same age group and also do not lose when compared to all men. These findings are in line with the first part of *Hypothesis 4* that age matters for

Age	Wom	Man	W-M	All Men
18-39	-0.139***	-0.086*	-0.054	-0.082*
40-59	-0.097**	-0.118***	0.022	-0.038
60-99	-0.145***	0.000	-0.145***	-0.087***

Table 8: **Performance per Age Group**. This table reports the performance of women and men in different age groups. We extend our benchmark regression (2) by replacing the female dummy by six group dummies (three age groups and sex). The estimates of the respective dummies are reported in the columns "Wom" and "Man". The column "W-M" reports the performance differences between all women and men in the three different age groups. The column labeled "All Men" shows the relative performance if women in a particular age group are compared to all men. The corresponding regressions are based on 1,116 observations of selling processes. There are 597 male participants and 519 female participants. The regressions involve dummies for the groups of dealers, year dummies, and item dummies as well as Ln_appraisal. All regressions involve intercepts. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

the performance of females. Surprisingly, young women lose significantly if compared to all men.[27] They do better if they are compared to men of their age group. This is evidence that experience matters (also for men).

Group	Description
1	Unskilled worker, up to one year of vocational training and education
2	Skilled employee, two to three years of vocational training and education
3	Complex specialization, e.g., bachelor degree
4	High complexity, e.g., master degree
5	Retiree
6	Student
7	Not classified

Table 9: **Description of Education Groups**. This table summarizes the six education groups used to classify the participants. We use a German version of of the International Standard Classification of Occupations ISCO-08 provided by the German Federal Employment Agency.

Our analysis so far does not control for education. To group the participants according to their educational levels, we exploit the pieces of information about the education and occupation that are revealed during the show. Typically, at the beginning of the selling process, the host asks the participants what educational background or job they have.

We can thus apply a German version[28] of the International Standard Classification of Occupations ISCO-08 and assign all participants to four main educational categories (1: low, up to one

---

[27]Notice that overall men in the age group 60-99 perform the best in our sample.

[28]This is published by the German federal entity called Bundesagentur für Arbeit (engl. Federal Employment Agency).

year of vocational training and education, 2: skilled employee, which typically involves a German apprenticeship[29] of about 2 to 3 years, 3: complex specialization. This typically involves a Bachelor degree or a master in craftsmanship[30], 4: high complexity. This typically involves a Master degree of a university or similar skills). In few cases, the information provided in the show is incomplete. For instance, it might be only revealed that a participant is retired without mentioning her/his job before retirement. Therefore, we include a fifth category consisting of retired participants for which we do not know their pre-retirement profession.[31] Furthermore, we also include a sixth class of participants that are currently getting an education (school, college, university, or apprenticeship). We end up with 1,067 single participants being classified in one of the six groups. Table 9 summarizes the six groups. For 122 single participants we have too little or inconclusive information to classify them. They are in the seventh group.

Educ	Wom	Man	W-M
1	-0.150*	-0.069	-0.081
2	-0.067*	0.000	-0.067*
3	-0.100*	0.046	-0.146**
4	-0.146**	0.004	-0.149**
5	-0.057	0.002	-0.059
6	-0.028	-0.159	0.131

Table 10: **Performance per Educational Level**. This table reports the performance of women and men in different education groups. We extend our benchmark regression (2) by replacing the female dummy by 12 group dummies (six education groups and sex). The estimates of the respective dummies are reported in the columns "Wom" and "Man". The column "W-M" reports the performance differences between women and men in the different education groups. The corresponding regression is based on 1,057 observations of selling processes. There are 563 male participants and 494 female participants. The regression involves dummies for the groups of dealers, year dummies and item dummies as well as Ln_appraisal. The regression involves an intercept. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

First, we extend our benchmark regression (2) by adding six dummies for the education groups.[32] The female dummy increases to 0.088 being significant at all levels. Therefore,

---

[29]This only exists in few other countries. It is a dual education of two to three years where individuals work part time and go to some specialized school the rest of their work time. It is not quite a bachelor degree, but leads to a distinct degree.

[30]This only exists in few other countries. The typical craftsman in Germany completes an apprenticeship putting the individual in the second category. After this, the craftsman can further advance by completing another degree called "Meister" (not to be confused with university master), which has similar requirements as a Bachelor degree, but focuses on a particular field in craftsmanship, e.g., carpenter.

[31]If a participant is retired, but reveals her/his pre-retirement profession, we use this information since we believe that it is more informative.

[32]The corresponding regression results are available upon request.

Educ	18-30				40-59				60-99			
	Wom	Man	W-M	All Men	Wom	Man	W-M	All Men	Wom	Man	W-M	All Men
1	-0.269*	-0.028	-0.241	-0.219	-0.153	-0.101	-0.052	-0.094	-0.048	-0.010	-0.037	0.013
2	-0.066	0.065	-0.131	-0.079	0.012	-0.036	0.048	0.004	-0.081	0.039	-0.119**	-0.094**
3	-0.080	0.000	-0.080	-0.142	-0.012	-0.015	0.003	-0.075	-0.165	0.162	-0.327***	-0.227***
4	-0.229*	-0.070	-0.159	-0.248***	-0.198	-0.084	-0.114	-0.222**	-0.084	0.089	-0.172*	-0.104
5									-0.064	0.052	-0.116	-0.070
6	-0.019	-0.115	0.095	0.123								

Table 11: **Performance per Educational Level and Age**. This table reports extensions of our benchmark regression (2). The female dummy in our benchmark regression (2) is replaced by 36 group dummies (female interacted with six dummies for education interacted with three dummies for age group). The columns "Wom" and "Man" report the estimates of the respective dummy variables. The columns "W-M" report the performance differences between women and men in the different education groups per age group. In the columns labeled "All Men", the female dummy in our benchmark regression (2) is replaced by 18 group dummies for women (six dummies for education interacted with three dummies for age group) and six dummies for all men with a certain educational level. Therefore, the columns report the relative performances of women in a certain age group with a certain educational level compared to all men of the same educational level (independent of the ages of the men). The corresponding regressions are based on 1,057 observations of selling processes. There are 563 male participants and 494 female participants. The regressions involve dummies for the groups of dealers, year dummies, and item dummies as well as Ln_appraisal. All regressions involve intercepts. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$. All regressions involve intercepts.

after additionally controlling for education, women lose about 8.8% compared to men. Second, we modify our benchmark regression (2) and interact the female dummy with the six dummies for education groups leading to 12 groups. The columns "Wom" and "Man" of Table 10 report the estimates of the respective dummy variables, while the column labeled "W-M" reports the differences between men's and women's performances per educational level. There are some interesting findings: First, women with a good education (3 and 4) and women with a low education (1) lose the most (within the group of women and when compared to men in the same education group). Second, female students perform best.

Finally, we run regressions controlling for both education and age. Again, we first compare women in a certain age group with a certain educational level to men in the same age group and with the same educational level. We thus interact sex, educational level, and age group and obtain 36 groups (18 for woman, 18 for men). The corresponding estimates are reported in the columns "Wom" and "Man" of Table 11. The performance differences are reported in the columns labeled "W-M" of Table 11. It can be seen that all younger and elderly women in the education groups 1 to 4 perform worse than their male counterparts. For midlife women, the performances are better. In fact, the point estimates are even positive for females in the

second and third education group.[33]

Further, we compare women in a certain age group with a certain educational level to all men of the same educational level (independent of the ages of the men). The columns labeled "All Men" of Table 11 report the corresponding results. Apparently, almost all groups of women lose. If we focus on the education groups 2 to 4, which represent about 90% of our sample, then in 8 out of 9 cases the losses are about 7% or bigger, which is close to or even bigger than the loss in our benchmark regression (2). As before, the losses are the biggest for younger and elderly women. The best-performing group consists of midlife women in the second educational category.

At first, it might be puzzling that well-educated female participants perform worse than average. It however makes sense that women in the second and third educational category perform better since they typically work in the industry and are used to negotiations due to their professions (e.g., saleswomen). Many women in the fourth group are for instance school teachers or medical doctors who are presumably not so much exposed to selling processes in their professional lives. In fact, more than a third of these women in our sample are teachers or medical doctors. To summarize, the selling performance of women is driven by age and occupation and not necessarily by the educational level. This confirms *Hypothesis 4.*

7.2   Attractiveness

We now investigate how a participant's attractiveness affects her/his performance. One might suppose that attractive people reach better negotiation results because they are for instance more confident (Mobius and Rosenblat (2006)) or because they are treated differently than unattractive people by others (Solnick and Schweitzer (1999)). We use the Chicago Face Database to create a Deep Learning model with Convolutional Neural Network and transfer learning with the VGG face model. The Chicago Face Database includes about 900 pictures of people (also showing various emotions) and their respective attractiveness scores. We use these pictures and a dense convolution with the VGG face model for transfer learning to predict so-called beauty scores of the BfR participants.

We then sort women and men into three beauty-score terciles based on the ordered beauty score

---

[33]Notice that the education groups 5 and 6 do not contain many participants. Furthermore, having the information that someone is a student or a retiree is not very informative, since we do not know the field of specialization or the former professions. For completeness, we show the results in Table 11, but we do not think that the results for these groups can be reasonably interpreted.

distribution for women and men. The first tercile contains the less attractive participants, while the third tercile includes the most attractive ones. This yields 6 groups of participants along the dimension sex and attractiveness. We then extend our benchmark regression (2) replacing the female dummy "Single_woman" with dummies for these 6 groups, where we use women in the third attractiveness tercile as reference group. The regression results are reported in Table 12.

Attractive	Wom	Man	W-M
Low	-0.083*	0.076**	-0.158***
Mid	0.000	0.045	-0.045
High	0.000	0.027	-0.027
H-L	0.083*	-0.048	0.131**

Table 12: **Performance for Attractiveness Groups**. This table reports an extension of our benchmark regression (2). We replace the female dummy by six group dummies (sex and three attractiveness groups). The columns "Wom" and "Man" report the estimates of the respective dummy variables. The column "W-M" reports the performance differences between all women and men in the three different attractiveness groups. Females in the third attractiveness group are the reference group. The row labeled "H-L" reports differences between High and Low. The regression is based on 1,116 observations of selling processes. There are 597 male participants and 519 female participants. The regression involves dummies for the groups of dealers, year dummies and item dummies as well as Ln_appraisal. The regression involves an intercept. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

We find that women in the first attractiveness tercile lose significantly compared to women in the third attractiveness tercile. The difference is as high as 8.3%. Thus, there is a significant effect of attractiveness for women. We do not find systematic attractiveness effects for male participants. There is no evidence in our sample that beauty drives the results for midlife women that we found in Section 7.1, since beauty scores are similarly distributed within age groups.[34]

To summarize, we find evidence for *Hypothesis 5* that attractiveness can significantly matter in selling processes. We document strong systematic attractiveness effects for females, whereas there are no systematic effects for male participants.

7.3 Optimism

Before the professional appraisal of her/his item is revealed to the participant, the host forces every participant to reveal her/his desired price. This price could be considered as the par-

---

[34]Results are available upon request.

ticipant's prior about the item's value. We now group all participants into three categories according to whether their relative prior desired price is low, medium or high. The relative prior desired price is defined as the ratio between the prior desired price (Desired) and the appraisal value (Appraisal). The size of the relative prior desired price could be interpreted as a proxy for the (a priori) optimism of the participant.[35] Participants whose relative prior desired price is smaller than 0.8 are in the first group (Low) and participants with a relative prior desired price between 0.8 and 1.2 are in the second group (Mid). All participants with a relative desired price bigger than 1.2 are in the third group (High).

We first test whether the relative desired price systematically differs for females and males: Running the regression

$$
\ln\left(\frac{\text{Desired}}{\text{Appraisal}}\right) = \beta_0 + \beta_1\text{Single\_woman} + \beta_2\text{Ln\_appraisal}
$$
$$
+ \sum_{n=1}^{22}\beta_{3n}\text{Item}_n + \sum_{t=2016}^{2021}\beta_{4t}\text{Year}_t + \epsilon
$$

yields an insignificant female dummy.[36] Therefore, we do not have significant evidence that the women of our sample set lower goals than their male counterparts.[37]

To analyze the effect of optimism, we extend our benchmark regression (2) replacing the female dummy by six group dummies (sex, relative desired price). Table 13 reports our results. There are several interesting findings: First, participants with high prior expectations outperform participants with low prior expectations (0.167). This is independent of sex. Thus, women and men benefit in the same way from optimistic priors. Further, women and men seem to update their priors in the same fashion, i.e., women lose across all optimism groups about 6.5% compared to men, which is close to the loss we report in our benchmark regression (2). To summarize, we find that the effect of optimism on participant's performance is independent of sex.

---

[35]We expect the participant to update her/his beliefs about the valuation of the item taking into account the appraisal value provided by the expert. Notice that this is actually enforced by the host when the appraisal value is below the prior desired price. In this case, the participant is only allowed to continue if she/he commits to sell for the appraisal value. If the prior desired price is below the appraisal, we do not have clear pieces of information about how the participants update their beliefs.

[36]The point estimate is 0.002 with a p-value of 0.955. Additional evidence that there is no systematic difference between females and males derives from our sorting into the three categories for the relative desired price. Females and males are represented almost equally in all three categories.

[37]This finding is different from the results in Kray, Thompson, and Galinsky (2001) and Major and Konar (1984).

Optimism	Wom	Man	W-M
Low	-0.144***	-0.075	-0.069**
Mid	0.000	0.065	-0.065
High	0.023	0.092	-0.069
H-L	0.167***	0.167***	0.000

Table 13: **Regression with Optimism**. This table reports an extension of our benchmark regression (2). We replace the female dummy by six group dummies for sex and the relative prior desired price of a participant. We group participants in three groups based on their relative prior desired price. In the group "Low" are participants whose relative prior desired price is smaller than 0.8, while participants whose relative prior desired price is bigger than 1.2 are in the group "High". All other participants are in the group "Mid". The columns "Wom" and "Man" report the estimates of the respective dummy variables. The row labeled "H-L" reports differences between High and Low per sex. The column labeled "W-M" reports differences between females and males holding the optimism classification fixed. The regression is based on 1,127 observations. There are 598 male participants and 529 female participants. The regression involves dummies for the groups of dealers, year dummies and item dummies as well as Ln_appraisal. The regression involves an intercept. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

7.4   Sentiment

We now turn to the question of how the relative performance of women is affected when we control for what is said during the auction and the post-auction negotiation. For each selling process, we calculate a sentiment score for the particular participant based on the words she/he says after entering the dealer room. We use an unsupervised rule-based sentiment analysis model that relies on a pre-defined dictionary classifying positive and negative words. Our dictionary is the German Polarity Lexicon by Clematide and Klenner (2010). It contains a list of German words (adjectives, nouns and verbs) associated with positive and negative sentiment and the respective sentiment scores for those words. The sentiment score for each word lies in the interval [-1,1] with -1 defining the most negative sentiment, 0 defining a neutral sentiment and 1 defining the most positive sentiment.

We first preprocess and normalize the text data by applying tokenization and lemmatization at the sentence level.[38] We then calculate the sentiment of each sentence by first identifying the sentiment score for each word in the sentence using the chosen dictionary and then taking an arithmetic average over the sentiment scores to arrive at the overall sentiment of the sentence. Those words in the sentence that are not part of the lexicon are neglected. We further take possible negations into account using a customized list of negation words. In case a word is preceded by a negation, we multiply its sentiment score with -0.5. Finally, to compute the

---

[38]Tokenization is the process of splitting a text object into smaller units known as tokens. Lemmatization is the process of reducing the different forms of a word to one single form.

sentiment score of the whole text, we average over all sentence-level sentiment scores.

One might be concerned about endogeneity issues since positive (negative) outcomes might lead to positive (negative) sentiment at the end of a selling process. To mitigate this potential problem, we measure the sentiment using the first 75% of the words only. We disregard the final part, which is also biased by farewell phrases.[39]

	Wom	Man	W-M
Low	-0.077*	-0.027	-0.050
Mid	-0.068	0.048	-0.116***
High	0.000	0.060	-0.060
H-L	0.077*	0.088*	-0.011

Table 14: **Regression with Seller Sentiment Scores**. This table reports an extension of our benchmark regression (2). We replace the female dummy by six group dummies for sex and sentiment tercile (Low, Mid, and High). The columns "Wom" and "Man" report the estimates of the respective dummy variables. The row labeled "H-L" reports differences between High and Low per sex. The column labeled "W-M" reports differences between females and males holding the sentiment tercile fixed. The regression is based on 1,127 observations. There are 598 male participants and 529 female participants. The regression involves dummies for the groups of dealers, year dummies and item dummies as well as Ln_appraisal. The regression involves an intercept. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

We group females and males into terciles of sentiment scores (low, medium, high). This yields six groups of participants along the dimension sex and sentiment. We extend our benchmark regression (2) replacing the female dummy by dummies for these six groups, where we use women with high sentiment scores as reference group. Table 14 shows that the effect of sentiment is significant across both females and males (0.077 and 0.088), i.e. participants benefit from being friendly. We also find that women in the medium tercile lose significantly in relation to men, while there is no significant difference between men and women ind the upper and lower tercile.

8 Item Characteristics

We now turn to the question whether there are differences in men's and women's performances depending on item characteristics. This question is motivated by evidence in the literature that the context of a negotiation matters for a woman's performance. First, as stated in *Hypothesis 3*, women are known to perform better in negotiations where the subject is in line with the *female gender role* (e.g., Bear and Babcock (2012); Bear (2011)). We can address this question by

---

[39]Our main results do not fundamentally change if we focus on 50% or 100% of the words.

analyzing their performance for items that are more or less congruent with the female gender role. Therefore, we classify all items as being either female-congruent or female-incongruent.[40]

Second, as summarized in *Hypothesis 2*, women perform relatively worse if some form of ambiguity is present. In our data, there is heterogeneity in the precision with which dealers evaluate different categories of items. One reason might be that the valuations of some items can be based on objective criteria like the value of the raw material (e.g., jewelry and sterling cutlery). For other items this is not the case (e.g., collectibles like autographs). This form of ambiguity leaves room for negotiation that participants can exploit. To summarize, item categories differ according to the precision of the dealers' evaluation.

The split into items that are more/less in line with the female gender role is obvious for most item categories. For instance, jewelry, tableware, or china are clearly female-congruent items, whereas technical devices, watches, or toys are not female-congruent items.[41] Given the results of Adams, Kräussl, Navone, and Verwijmeren (2021) about the art market, also paintings are supposed to be in the group of female-incongruent items. Nevertheless, for some categories the classification is less obvious (e.g., furniture).

We have thus decided to use an objective criterion namely the relative frequency of an item. If 50% or more items in a particular category are brought in by a woman, then we label the item as female-congruent item. Table 1 reports the resulting classification. In the column labeled "Item_fem" there is a one if the relative frequency reported in the column "Adjusted" is equal or larger than 0.5. Notice that this column reports the relative frequency adjusted for the fact that we observe slightly fewer females than males. According to this classification, there are 310 selling processes in which a women brings in a female-congruent item. The resulting classification seems to make sense. For instance, jewelry, china, clothes, tableware, and flatware are in the group of female-congruent items, whereas toys, devices, watches, and paintings are not classified as being female-congruent items.[42]

We now turn to the classification into items with more or less disperse valuation. We measure the dispersion in the dealers' valuation of an item category by the variation in the performance measure for this category. Therefore, for each category we separately compute the standard

---

[40]As mentioned in the introduction, we use "female-congruent item" as abbreviation for items that are congruent with the female gender role.

[41]Notice that watches are watches for males, whereas watches for females are classified as jewelry. Furthermore, the category "toys" is dominated by male toys such as Matchbox cars etc.

[42]As robustness check, we have for instance classified art, bronze, furniture, or watches as female-congruent items, but our main results still hold.

deviation of our performance measure across all selling processes. We then sort the item categories into two groups. The first group (labeled "0") contains the item categories with a low standard deviation in the performance measure, whereas the second group (labeled "1") includes the item categories with a high standard deviation in the performance measure. We refer to these groups as items with low disperse valuation (short: non-disperse valued items) and items with high disperse valuation (short: disperse valued items). The resulting classification can be found in the column labeled "Item_dis" of Table 1.

	Wom	Man	W-M
		Item_fem	
0	-0.156***	-0.064**	-0.092***
1	0.000	0.018	-0.018
		Item_dis	
0	0.000	-0.015	0.015
1	-0.096**	-0.000	-0.096**

Table 15: **Separate Regressions with Dummy for Female-Congruent Items or Items with Disperse Valuations**. This table reports two variants of our benchmark regression (2). In the first regression, we replace the female dummy and the item dummies by two group dummies for sex and female-congruent vs. female-incongruent items. The results are reported in the row with the headline "Item_fem". In the second regression, we replace the female dummy and the item dummies by two group dummies for sex and disperse-valued vs. non-disperse-valued items. The results are reported in the row with the headline "Item_dis". The column labeled "W-M" reports differences between females and males. The regressions are based on 1,127 observations. There are 598 male participants and 529 female participants. The regressions involve year dummies and dummies for the groups of dealers as well as Ln_appraisal. Both regressions involve intercepts. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

Having established both classifications, we modify our benchmark regression (6.1) and run two alternative regressions. In the first specification, we replace the female dummy and the item dummies by two group dummies for sex and female-congruent vs. female-incongruent items. The results are reported in the row with the headline "Item_fem" of Table 15. It can be seen that women lose significantly if they sell female-incongruent items, but there is no gap for female-congruent items. This is in line with *Hypothesis 3*. In the second specification, we replace the female dummy and the item dummies by two group dummies for sex and disperse-valued vs. non-disperse-valued items. The results are reported in the row with the headline "Item_dis" of Table 15. It can be seen that women lose significantly if they sell items of a category with high disperse valuation, but there is no gap for items of a category with low disperse valuation. This supports *Hypothesis 2b*.

To analyze the joint effect of both moderators, we now group all selling processes according to

Case	Item_fem	Item_dis	Wom	Man	W-M
a	0	0	-0.134***	-0.102**	-0.032
b	0	1	-0.182***	-0.047	-0.135***
c	1	0	0.000	0.008	-0.008
d	1	1	-0.023	0.030	-0.052
d-c			-0.023	0.022	-0.044
b-a			-0.048	0.055	-0.103
d-b			0.159**	0.077	0.083
c-a			0.134***	0.110**	0.024
b-c			-0.182***	-0.055	-0.127**

Table 16: **Regression with Dummies for Female-Congruent Items and Items with Disperse Valuations**. The table reports results of a regression replacing the female dummy in (2) by eight dummies for eight groups along the following three dimensions: woman vs. man, female-congruent item vs. female-incongruent item, and disperse-valued item vs. non-disperse valued item. The estimates of the respective dummies are reported in the columns "Wom" and "Man". The column "W-M" reports the performance differences between women and men for the different item characteristics. Women selling female-congruent items with low disperse valuations are the reference group. The last five rows report results for differences between the cases and diff-and-diff results in the last column The corresponding regression is based on 1,127 observations of selling processes. There are 598 male participants and 529 female participants. The regression involves year dummies and dummies for the groups of dealers as well as Ln_appraisal. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

three dimensions: female vs. male participant, female-congruent item vs. female-incongruent item and disperse-valued item vs. non-disperse valued item. This leads to eight cases and thus eight dummies. Table 16 shows the estimates of these dummies in the first four rows of the columns labeled "Wom" and "Man". The reference case is a women selling a female-congruent item with low disperse valuation. Therefore, the corresponding dummy has a value of zero.

Looking at the column "W-M", it is obvious that again both classifications matter. However, the most pronounced gap arises if women sell female-incongruent items with more disperse valuation (case "b"). The joint effect of dispersion in valuation and female congruence is reported in the row "b-c" and amounts to 12.7%. Case "c", i.e., a female-congruent item with low disperse valuation, is essentially the best combination for women, whereas case "b", i.e., a female-incongruent item with a high disperse valuation, is the worst one. We find that women lose about 18.2% when selling a female-incongruent item with high disperse valuation as compared to a female-congruent item with less disperse valuation. This difference is significant. Notice that the gaps for female-incongruent items with less disperse valuation and for female-congruent items with more disperse valuation are negative, but not significant. This documents that it is the combination of both moderators, which drives our findings. Another piece of evidence for this result is that all cases where only one dimension is varied ("d-c", "b-a", "d-b", and "c-a")

lead to insignificant gaps.

9 Dealer Characteristics

Our data sample contains 62 different compositions of groups of dealers. We can thus exploit the heterogeneity across dealer characteristics (e.g., generosity or activeness) and study how women and men perform given certain dealer characteristics. We focus on 47 groups of dealers for which we have at least 10 observations. This leaves us with 1,060 selling processes of 567 single males and 493 single females. The relevant groups are reported in Table 17.

We characterize groups of dealers along four dimensions: the generosity of a group, the average number of bidders, the relative frequency that the group of dealers asks for the appraisal or the desired value, and the average age of male dealers. All quantities are calculated across all selling processes observed for a particular group of dealers.[43]

We measure the *generosity* of a group of dealers by the average value of Perform for this group: The higher the average, the more generous is this group. Furthermore, we interpret the average number of bidders as a group's *activeness*: The more group members bid for items on average, the more active is this group. Activeness could facilitate selling items to a group of dealers, since the seller does not have to be particularly active her/himself. The relative frequency that a group of dealers asks for the appraisal or the desired price proxies for the group's *pushiness*. If a group asks more often for these quantities, the group is interpreted as being more pushy, which might intimidate participants. Finally, the average age of male dealers can be a proxy for the presence of backlash and gender biases in a certain group of dealers. The older the male dealers are, the more likely is the presence of backlash or gender biases.

To examine how these dealer characteristics affect the participant's performance, we replace the dummies for the groups of dealers in our benchmark regression (2) by variables capturing the characteristics. We are running the regression:

$$
\begin{aligned}
\text{Ln\_perform} \;=\; & \beta_0 + \beta_1 \text{Single\_woman} + \beta_2 \text{Ln\_appraisal} + \beta_3 \text{Avg\_perform} \qquad\qquad (9.1) \\
& + \beta_4 \text{Avg\_bid} + + \beta_5 \text{Avg\_ask} + \sum_{n=1}^{22} \beta_{6n} \text{Item}_n + \sum_{t=2016}^{2021} \beta_{7t} \text{Year}_t + \epsilon,
\end{aligned}
$$

---

[43]We have included all participants, singles and teams. For every participant, we calculate the relevant averages across all observations of the corresponding group of dealers disregarding the values of the particular selling process.

Composition	Avg Perform	Avg Bid	Avg Ask	AvgAMD	Appear
Wolf_Juli_Wald_Fabi_Lisa	1.153	3.065	0.176	42.0	108
Susi_Roma_Supp_Dani_Esth	1.232	2.743	0.171	47.0	70
Roma_Elke_Juli_Chri_Mark	1.035	2.970	0.212	49.0	66
Susi_Wolf_Juli_Wald_JanC	0.917	3.206	0.302	45.8	63
Susi_Wolf_Wald_Fabi_Ludw	0.917	2.860	0.228	53.0	57
Roma_Dani_Elke_Fabi_Mark	1.163	3.179	0.196	48.3	56
Supp_Juli_Fabi_Lisa_Stev	1.060	3.036	0.321	41.3	56
Susi_Dani_Wolf_Wald_Fabi	1.228	3.296	0.167	45.3	54
Susi_Supp_Dani_Wolf_Wald	1.155	3.500	0.241	46.0	54
Wald_Fabi_Lisa_JanC_Frie	1.268	2.778	0.333	43.0	54
Susi_Juli_Wald_Chri_Thor	1.059	2.478	0.109	48.5	46
Supp_Wolf_Elke_Fabi_Stev	1.308	3.511	0.156	44.8	45
Roma_Supp_Wolf_Elke_Juli	1.014	3.091	0.114	44.3	44
Susi_Wolf_Elke_Mark_Stev	1.259	3.791	0.209	57.0	43
Roma_Supp_Juli_Lisa_Thor	0.995	2.771	0.343	46.8	35
Susi_Roma_Esth_Chri_Thor	1.263	2.886	0.229	55.0	35
Roma_Elke_Juli_Fabi_JanC	0.963	3.000	0.324	42.5	34
Roma_Elke_Juli_Frie_Chri	1.155	3.029	0.147	47.0	34
Supp_Elke_Wald_JanC_Chri	0.963	3.088	0.265	44.5	34
Wolf_Elke_Wald_Fabi_Chri	1.146	3.294	0.206	44.8	34
Wolf_Wald_Fabi_Lisa_Chri	0.953	3.088	0.294	44.8	34
Elke_Juli_Wald_Fabi_Mark	1.035	3.394	0.242	43.5	33
Juli_Wald_Fabi_Lisa_Mark	1.126	3.152	0.182	43.5	33
Roma_Dani_Fabi_Lisa_Thor	1.076	3.273	0.091	49.3	33
Susi_Dani_Fabi_Chri_Stev	1.178	3.242	0.121	47.8	33
Esth_Elke_Juli_Chri_Thor	1.234	2.969	0.188	46.7	32
Susi_Roma_Wolf_Fabi_Thor	1.287	2.750	0.156	49.5	32
Supp_Elke_Wald_Mark_Stev	1.317	3.774	0.097	52.3	31
Susi_Wolf_Wald_Fabi_Chri	1.021	3.519	0.111	44.8	27
Dani_Elke_Fabi_Mark_Thor	1.164	2.792	0.167	48.0	24
Juli_Fabi_Lisa_Thor_Ludw	0.967	2.957	0.391	50.8	23
Supp_Fabi_Lisa_JanC_Mark	1.409	3.000	0.174	40.8	23
Susi_Juli_Wald_Fabi_Mark	0.895	2.609	0.261	43.5	23
Dani_Wald_Fabi_Lisa_Ludw	0.991	2.889	0.222	52.8	18
Susi_Dani_Esth_Juli_Mark	0.973	2.833	0.167	46.0	18
Susi_Dani_Juli_Wald_JanC	1.086	2.778	0.222	45.5	18
Roma_Supp_Elke_Fabi_Mark	1.205	3.059	0.000	44.5	17
Susi_Wolf_Juli_Wald_Fabi	0.933	3.000	0.235	42.0	17
Esth_Elke_Fabi_JanC_Chri	0.992	2.688	0.313	40.3	16
Susi_Supp_Juli_Wald_Fabi	1.243	2.375	0.188	38.0	16
Wolf_Elke_Wald_Fabi_Ludw	0.866	3.333	0.067	53.0	15
Dani_Elke_Fabi_JanC_Stev	1.013	3.083	0.333	47.5	12
Dani_Elke_Wald_JanC_Ludw	1.111	3.583	0.250	56.5	12
Susi_Esth_Wolf_Fabi_Mark	1.388	3.250	0.000	44.7	12
Susi_Juli_Wald_Fabi_Chri	0.868	2.500	0.250	41.3	12
Susi_Wolf_Juli_Fabi_Mark	1.144	3.167	0.167	42.3	12
Susi_Roma_Juli_Fabi_Chri	1.672	2.909	0.091	42.8	11

Table 17: **Dealer Group Characteristics**. This table reports summary statistics concerning the average of Perform (Avg Perform), the average number of bidders (Avg Bid), the probability to ask for the desired or appraised price (Avg Ask), and the age of the male dealers in a particular group of dealers (AvgAMD). Appear is the number of appearances of a group of dealers. All quantities are based on all available observations (single participants and teams). The column labeled "Composition" provides the compositions of the group of dealers. They derive from the shortcuts of the dealer names reported in Table 4. There are 15 groups of dealers that appear less than eight times and that are disregarded in this table.

where Avg_perform is the average of Perform, Avg_ask is the average number of bidders and Avg_bid is the probability to ask for the desired or appraised price for a specific group of dealers. For every participant we calculate the relevant averages across all observations of the corresponding group of dealers disregarding the values of the particular selling process to avoid issues with endogeneity. Table 18 reports the corresponding regression results.

Ln_appraisal	-0.188***
	(-13.87)
Single_woman	-0.084***
	(-3.26)
Avg_perform	0.246**
	(2.07)
Avg_bid	0.087*
	(1.94)
Avg_ask	-0.311*
	(-1.73)
R^2	0.266

Table 18: **Regression with Dealer Characteristics**. This table reports our results if we substitute the dummies for the groups of dealers in our benchmark regression (2) by the three dealer characteristics (generous, active and pushy). The corresponding regression is based on 1,060 observations of selling processes. There are 567 male participants and 493 female participants. The regression involves year dummies and item dummies. The regression also involves an intercept. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

We find that women still lose significantly compared to men. The loss is as high as 8.4% and thus even higher as in our benchmark regression (2). Further, participants benefit from generous and active groups of dealers. They lose when assigned to a group of dealers that is on average pushy. Those results are all significant.

We further want to examine how those dealer characteristics affect the performance of women relative to men. Therefore, for all four dimensions (generous, active, pushy, average age male dealers), we sort the selling processes into terciles: being assigned to a group of dealers with a low, medium, or high realization of the particular characteristic. We thus obtain six groups per characteristic (sex and dimension terciles). For instance, we have three terciles for generosity interacted with the sex of the participant. To examine the effect of these dealer characteristics, we extend our benchmark regression (2) by taking out the female dummy and the dummies for the groups of dealers and replacing them with six dummies for the particular dealer characteristic. Table 19 reports our results. It consists of four panels, one for each dealer characteristic. The rows labeled "Low", "Mid", and "High" show our findings for groups of dealers with low, medium, and high values of the particular characteristic. The rows labeled "H-L" report dif-

ferences between the estimates for high and low values per sex. The columns labeled "W-M" report differences between the estimates for females and males holding fixed the particular level of a dealer characteristic. The entries in the lower right corners (intersection between W-M and H-L) are the diff-and-diff estimates.

	Wom	Man	W-M	Wom	Man	W-M
		Generous			Active	
Low	-0.066	-0.016	-0.050	-0.100**	0.025	-0.125***
Mid	-0.008	0.050	-0.058	-0.085*	-0.031	-0.054
High	0.000	0.133***	-0.133***	0.000	0.062	-0.062
H-L	0.066	0.149***	-0.083	0.100**	0.037	0.063
		Pushy			Age Male Dealer	
Low	0.000	0.051	-0.051	0.000	0.054	-0.054
Mid	-0.008	0.070	-0.078*	0.017	0.117***	-0.100**
High	-0.122***	-0.002	-0.120***	-0.015	0.073	-0.087**
H-L	-0.122***	-0.053	-0.069	-0.015	0.019	-0.034

Table 19: **Performance for Dealer Characteristics**. This table reports an extension of our benchmark regression (2). We sort the selling processes according to four dealer characteristics (generous, active, pushy, and average age of male dealers). We replace the female dummy by six group dummies for sex and terciles of the particular dealer characteristic (Low, Mid, and High). The estimates of the respective dummies are reported in the columns "Wom" and "Man". The rows labeled "H-L" report differences between High and Low per sex. The columns labeled "W-M" report differences between females and males holding fixed the particular level of a dealer characteristic. The corresponding regressions are based on 1,060 observations of selling processes. There are 567 male participants and 493 female participants. The regressions involve year dummies and item dummies as well as Ln_appraisal. The regressions also involve intercepts. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

First, we obtain very significant results for generosity. As can be seen in the row H-L, men benefit significantly from being assigned to a generous group of dealers, but the performance of women is not significantly affected (0.149 vs. 0.066). This result for generous groups is consistent with *Hypothesis 2c* saying that women do not exploit their opportunities if there is ambiguity about the outcome.

By contrast, women are significantly affected by the inactiveness of a group of dealers, whereas men perform similarly across different levels of activeness (0.100 vs. 0.037). For inactive groups, women lose 12.5% compared to men. A similar pattern arises for the pushiness of a group of dealers. Women significantly lose if a group is classified as pushy, but men are less affected. To summarize, both measures for a tough selling environment have significant effects on the performance of women, whereas men are less influenced by them. These findings support *Hypothesis 6*, which says that women perform worse in tough selling environments.

For the average age of the male dealers, there is no significant effect across females. But compared to male participants they perform worse if the groups of dealers involve older males (-0.100 and -0.087).

10 Female Teams

Up to now we have focused on the performance of single women in relation to single men. We now want to investigate whether women can improve their performance if they team up with a second person. The majority of our observations are single females and males. We have however 94 additional observations of female teams consisting of two women and 113 observations of couples where we know that the female is the seller.[44] We thus run our benchmark regression (2) augmented by two dummies for female teams and couples where the female is the seller. Table 20 shows the relevant dummies. It is remarkable that female teams perform best and significantly outperform single women. Also couples where the female is the seller do not perform significantly worse than single males. Apparently, it pays off for single females to team up.

Single_man	-0.047
Single_woman	-0.117***
Team_woman	0.000
Couple_wom_sell	-0.078

Table 20: **Performance of Single Women and Female Teams**. The table reports results if we extend our benchmark regression (2) by replacing the female dummy with dummies for the following four categories: single man, single woman, female team, and couple where the woman is the seller. We use female teams as reference group. The estimates of the four dummies are reported in the table. The regression is based on 1,328 observations of selling processes. There are 598 male participants, 529 female participants, 94 female teams, and 113 couples where the woman is the seller. The regression involves dummies for the groups of dealers, year dummies and item dummies as well as Ln_appraisal. The regression involves an intercept. Robust t-statistics (hc3) are reported in brackets. Asterisks correspond to the following p-values: $*p < 0.1, **p < 0.05, ***p < 0.01$.

The good performance of female teams provides evidence for *Hypothesis 7*, since female teams can be considered as an example for an advocacy setting. Therefore, our findings are in line with Amanatullah and Morris (2010) and Bowles, Babcock, and McGinn (2005) who document that women negotiating on behalf of others perform well.

---

[44]For the majority of the remaining observations, we do not know whether the female is the seller. Besides, there are only a few male teams and we find no evidence that teaming up for males pays off.

11 Conclusion

This paper offers a comprehensive analysis of the selling performance of non-professional individuals selling personal belongings. We focus on the relative performance of women compared to men and document a significant gender revenue gap that is as high as 7.3%.

The performance across female sellers is however heterogeneous. We document a new relationship between age, education and negotiation outcomes: Midlife women working in the industry perform as well as men, whereas most other women obtain much less than men. Interestingly, a university degree is not a guarantee for women to perform well. Profession matters more.

The selling environment is also important. Women perform worse if dealers are generous. This result suggests that women do not exploit their opportunities. Together with our findings concerning profession, an important policy implication arises: Offering negotiation training courses to women might empower them to achieve better performances.

Furthermore, we document that female teams show the same performance as males. Hence, teaming up seems to be a good recommendation for women as well.

References

Adams, R., R. Kräussl, M. Navone, and P. Verwijmeren, 2021, Gendered prices, *Review of Financial Studies* 34, 3789–3839.

Amanatullah, E.T., and M.W. Morris, 2010, Negotiating gender roles: gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others, *Journal of Personality and Social Psychology* 98, 256–267.

Amanatullah, E.T., and C.H. Tinsley, 2013, Punishing female negotiators for asserting too much ... or not enough: exploring why advocacy moderates backlash against assertive female negotiators, *Organizational Behavior and Human Decision Processes* 120, 110–122.

Ayres, I., 1991, Fair Driving: gender and race discrimination in retail car negotiations, *Harvard Law Review* 104, 817–872.

Ayres, I., and P. Siegelman, 1995, Race and gender discrimination in bargaining for a new car, *American Economic Review* 85, 304–321.

Bear, J., 2011, "Passing the buck": incongruence between gender role and topic leads to avoidance of negotiation, *Negotiation and Conflict Management Research* 4, 47–72.

Bear, J.B., and L. Babcock, 2012, Negotiation topic as a moderator of gender differences in negotiation, *Psychological Science* 23, 743–744.

Beyer, S., 1990, Gender differences in the accuracy of self-evaluations of performance, *Journal of Personality and Social Psychology* 59, 960–970.

Blau, F. D., and L. M. Kahn, 2017, The gender-wage gap: extent, trends, and explanations, *Journal of Economic Literature* 55, 789–865.

Bowles, H.R., L. Babcock, and L. Lai, 2007, Social incentives for gender differences in the propensity to initiate negotiations: sometimes it does hurt to ask, *Organizational Behavior and Human Decision Processes* 103, 84–103.

Bowles, H.R., L. Babcock, and K.L. McGinn, 2005, Constraints and triggers: situational mechanics of gender in negotiation, *Journal of Personality and Social Psychology* 89, 951–965.

Bowles, H.R., and K.L. McGinn, 2008, Untapped potential in the study of negotiation and gender inequality in organizations, *The Academy of Management Annals* 2, 99–132.

Callahan-Levy, C.M., and L.A. Messé, 1979, Sex differences in the allocation of pay, *Journal of Personality and Social Psychology* 37, 433–446.

Clematide, S., and M. Klenner, 2010, Evaluation and extension of a polarity lexicon for German, *Proceedings of the First Workshop on Computational Approaches to Subjectivity and Sentiment Analysis* pp. 7–13.

Croson, R., and U. Gneezy, 2009, Gender Differences in Preferences, *Journal of Economic Literature* 47, 448–474.

Curhan, J. R., M. A. Neale, L. Ross, and J. Rosencranz-Engelmann, 2008, Relational accomodation in negotiation: Effects of egalitarianism and gender economic efficiency and relational capital, *Organizational Behavior and Human Decision Process* 107, 192–205.

Damnals, J.E., J.J. Zlatev, N. Halevy, and M.A. Neale, 2021, The dynamics of gender and alternatives in negotiation, *Journal of Applied Psychology* 106, 1655–1672.

Daubman, K.A., L. Heatherington, and A. Ahn, 1992, Gender and the self-presentation of academic achievement, *Sex Roles* 27, 187–204.

Day, N. E., 1993, Performance in salespeople - The impact of age, *Journal of Managerial Issues* 5, 254–273.

Diekman, A.B., and A.H. Eagly, 2000, Stereotypes as dynamic constructs: Women and men of the past, present, and future, *Personality and Social Psychology Bulletin* 26, 1171–1188.

Eagly, A.H., 1987, *Sex differences in social behavior: a social-role interpretation.* Lawrence Erlbaum Associates.

Eagly, A.H., and S.J. Karau, 2002, Role congruity theory of prejudice toward female leaders, *Psychological Review* 109, 573–598.

Eagly, A.H., and W. Wood, 2012, Social role theory, in P.M. Van Lange, A.W. Kruglanski, and E.T. Higgins, eds.: *Handbook of theories of social psychology* (Sage, Thousand Oaks, CA).

Eagly, A. H., and W. Wood, 1999, The origins of sex differences in human behavior: Evolved dispositions versus social roles, *American Psychologist* 54, 408–423.

Eckel, C. C., and P.J. Grossman, 2008, Sex and risk: experimental evidence, in C. R. Plott, and V.L. Smith, eds.: *Handbook of Experimental Economics Results* (Elsevier, New York).

Goldsmith-Pinkham, P., and K. Shue, 2023, The gender gap in housing returns, *Journal of Finance* 78, 1097–1145.

Gould, R. J., and C.G. Slone, 1982, The "feminine modest" effect: a self-presentational interpretation of sex differences in causal attribution, *Personality and Social Psychology Bulletin* 8, 477–485.

Hüffmeier, J., P.A. Freund, A. Zerres, K. Backhaus, and G. Hertel, 2014, Being tough or being nice? A meta-analysis on the impact of hard-and softline strategies in distributive negotiations, *Journal of Management* 40, 866–892.

Kray, L.J., and L.L. Thompson, 2005, Gender stereotypes and negotiation performance: an examination of theory and research, in B.M. Staw, and R.M. Kramer, eds.: *Research in organizational behavior: an annual series of analytical essays and critical reviews* (Elsevier, Oxford, UK).

Kray, L.J., L. Thompson, and A. Galinsky, 2001, Battle of the sexes: gender stereotype confirmation and reactance in negotiations, *Journal of Personality and Social Psychology* 80, 942–985.

Kugler, K.G., J.A.M. Reif, T. Kaschner, and F.C. Brodbeck, 2018, Gender differences in the initiation of negotiations: a meta-analysis, *Psychological Bulletin* 144, 198–222.

Kulik, C.T., and M. Olekalns, 2012, Negotiating the gender divide: lessons from the negotiation and organizational behavior literatures, *Journal of Management* 38, 1387–1415.

Levy, M., and A. Sharma, 1994, Adaptive selling: The role of gender, age, sales experience, and education, *Journal of Business Research* 31, 39–47.

Long, J. S., and L. H. Ervin, 2000, Using heteroscedasticity consistent standard errors in the linear regression model, *The American Statistician* 54, 217–224.

Lundeberg, M. A., P.W. Fox, and J. Punccohar, 1994, Highly confident but wrong: gender differences and similarities in confidence judgments, *Journal of Educational Psychology* 86, 114–121.

Maccoby, E.E., and C.N. Jacklin, 1974, *The psychology of sex differences*. Stanford University Press.

MacKinnon, J. G., and H. White, 1985, Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties, *Journal of Econometrics* 29, 53–57.

Major, B., and E. Konar, 1984, An investigation of sex differences in pay expectations and their possible causes, *Academy of Management Journal* 27, 777–792.

Major, B., D.B. McFarlin, and D. Gagnon, 1984, Overworked and underpaid: on the nature of gender differences in personal entitlement, *Journal of Personality and Social Psychology* 47, 1399–1412.

Mazei, J., J. Hüffmeier, P.A. Freund, A.F. Stuhlmacher, L. Bilke, and G. Hertel, 2015, A meta-analysis on gender differences in negotiation outcomes and their moderators, *Psychological Bulletin* 141, 85–104.

Mischel, W., 1977, The interaction of person and situation, in D. Magnusson, and N. S. Endler, eds.: *Personality at the crossroads: current issues in interactional psychology* (Erlbaum, Hillsdale, NJ).

Mobius, M.M., and T.S. Rosenblat, 2006, Why beauty matters, *American Economic Review* 96, 222–235.

Reif, J.A.M., K.G. Kugler, and F.C. Brodbeck, 2019, Why are Women less likeyl to negotiate? The Influence of Expectancy Considerations and Contextual Framing on Gender Differences in the Initiation of Negotiation, *Negotiation and Conflict Management Research* 13, 287–303.

Rudman, L.A., 1998, Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management, *Journal of Personality and Social Psychology* 74, 629–645.

Rudman, L.A., and P. Glick, 1999, Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers, *Journal of Personality and Social Psychology* 77, 1004–1010.

Rudman, L.A., and J.E. Phelan, 2008, Backlash effects for disconfirming gender stereotypes in organizations, *Research in Organizational Behavior* 28, 61–79.

Solnick, S.J., and M.E. Schweitzer, 1999, The influence of physical attractiveness and gender on ultimatum game decisions, *Organizational Behavior and Human Decision Processes* 79, 199–215.

Stevens, C. K., A. G. Bavetta, and M. E. Gist, 1993, Gender differences in the acquisition of salary negotiation skills: The role of goals, self-efficacy, and perceived control, *Journal of Applied Psychology* 78, 723–735.

Stuhlmacher, A.F., and E. Linnabery, 2013, Gender and negotiation: a social role analysis, in M. Olekalns, and W.L. Adair, eds.: *Handbook of research on negotiation* (Edward Elgar, London, UK).

Stuhlmacher, A.F., and A.E. Walters, 1999, Gender differences in negotiation outcome:a meta-analysis, *Personnel Psychology* 52, 653–677.

Tinsley, C.H., S.I. Cheldelin, A. Schneider, and E.T. Amanatullah, 2009, Women at bargaining table: pitfalls and prospects, *Negotiation Journal* 25, 233–248.

Twenge, J.M., 1997, Changes in masculine and feminine traits over time: a meta-analysis, *Sex Roles* 36, 305–325.

Wade, M.E., 2001, Women and salary negotiation: the costs of selfadvocacy, *Psychology of Women Quarterly* 25, 65–76.

Walters, A.E., A.F. Stuhlmacher, and L.L. Meyer, 1998, Gender and negotiator competitiveness: a meta-analysis, *Organizational Behavior and Human Decision Processes* 76, 1–29.

Zerres, A., J. Hüffmeier, P.A. Freund, K. Backhaus, and G. Hertel, 2013, Does it take two to tango? Longitudinal effects of unilateral and bilateral integrative negotiation training, *Journal of Applied Psychology* 98, 478–491.