Women in a Men's World

Risk Taking in an Online Card Game Community

Czibor, Eszter; Claussen, Jörg; Van Praag, Mirjam

Document Version
Accepted author manuscript

Published in:
Journal of Economic Behavior & Organization

DOI:
10.1016/j.jebo.2018.11.011

Publication date:
2019

Women in a Men's World: Risk Taking in an Online Card Game Community

Eszter Czibor, Jörg Claussen, and Mirjam Van Praag

DOI: 10.1016/j.jebo.2018.11.011

Women in a men's world: Risk taking in an online card game community*

Eszter Czibor[1], Jörg Claussen[2,3] and Mirjam van Praag[3,4]

[1]University of Chicago
[2]LMU Munich
[3]Copenhagen Business School
[4]Vrije Universiteit Amsterdam

November 12, 2018

Abstract

Analyzing a large data set from an online card game platform, a traditionally masculine environment with low female representation, we provide novel field evidence for gender differences in risk taking. Our paper complements existing laboratory experiments by studying a setting where selection into and out of the choice environment is endogenous, choices and outcomes are publicly observable and decisions are repeated over hundreds of rounds. We show that despite the possibility of sorting, imitation or learning, female players persistently choose lower risk-return profiles than men. We argue that the observed gender differences in risk taking result from true preference differences rather than a gap in skill, confidence or beliefs.

JEL-codes: D03, J24, M52
Keywords: gender, risk preferences, experience, selection, natural experiment

*Corresponding author: Eszter Czibor; E-mail: eczibor@uchicago.edu, Mailing address: 1126 E. 59$^{\text{th}}$ Street, Chicago, IL 60637, USA. Acknowledgments: We are grateful for *sauspiel.de* for making the dataset analyzed in this study available for the purpose of our research. We thank participants at the Incentives and Behavior Change workshop, the MBEES-2015 conference and the Tinbergen Institute Organizations & Markets Research Group Workshop, and at seminars at Carlos III Madrid University, Copenhagen Business School, Max Planck Institute for Innovation and Competition, Sussex University and the University of Chicago for their insightful comments. We are particularly grateful for the helpful suggestions of Thomas Buser, Tommaso Ciarli, Piotr Denderski, Catherine Eckel, Uri Gneezy, Ingrid Huitfeldt, Fatemeh Momeni, Julie Pernaudet and Randolph Sloof, and two anonymous referees. A previous version of this paper was circulated under the title: "Women do not play their aces - The consequences of shying away".

1 Introduction

Gender differences in risk preferences have been researched extensively, leading to a consensus that women are, on average, more risk averse than men (Beckman et al., 2016; Borghans et al., 2009; Charness and Gneezy, 2012; Donkers et al., 2001; Eckel and Grossman, 2008). A large share of the supporting evidence comes from experiments where decisions are one-shot, choices and outcomes are private, and there is no selection into or out of the decision context (Croson and Gneezy, 2009).[1] In real life, however, people tend to have some flexibility in choosing the environment in which they make decisions, and the risky decisions they face are often repeated. Over the course of these repetitions, people may receive feedback, observe others' choices and results, and may be subject to their peers' scrutiny. The aim of this paper is to study whether a gender gap in risk taking persists even in such settings. Specifically, we assess whether women who self-select into and gain experience in a traditionally masculine environment persistently choose lower risk-return profiles than men.

There are many reasons to suspect that contextual factors affect the size of the gender gap in risk taking (Adams and Funk, 2012). One channel that could mitigate the gap is differential sorting: it is possible that while few women choose to enter male-dominated environments, those who do have similar characteristics to men (Croson and Gneezy, 2009). Alternatively, even if women who select into these domains start out with different decisions, the gap may disappear over time through learning and adaptation (Croson and Gneezy, 2009). This aspect is particularly relevant in light of recent findings suggesting that risk preferences are malleable and respond to one's social environment (Andersen et al., 2013; Booth et al., 2014). Finally, familiarity with and expertise in risky decision making may reduce the gap, particularly if it originally resulted from gender differences in knowledge or skills rather than preferences (Atkinson et al., 2003; Dwyer et al., 2002).

A small literature attempts to address these issues by focusing on the gender gap in risk taking among finance professionals. These studies tend to rely on self-reported, unincentivized data, such as survey responses elicited from fund managers (Beckmann and Menkhoff, 2008; Olsen and Cox, 2001) and CEOs (Adams and Funk, 2012), or case study solutions of management students (Johnson and Powell, 1994). An exception is Atkinson et al. (2003) who compare male- and female-managed funds in terms of performance, risk and other characteristics. However, their sample only contains 72 female managers. Overall, evidence from this approach has been inconclusive: while some studies found no gender gap in risk taking (Atkinson et al., 2003; Johnson and Powell, 1994), others reported women to be more risk averse (Beckmann and Menkhoff, 2008; Olsen and Cox, 2001) or more risk loving (Adams and Funk, 2012).

Our paper takes a different approach by analyzing the decisions of players in a large online community for the German card game *Schafkopf*: a majority-male environment where par-

[1] Observational studies of individuals' and households' private investment decisions also show a robust gender gap in risk taking (e.g. Bernasek and Shwiff, 2001; Jianakoplos and Bernasek, 1998; Sunden and Surette, 1998). However, as Dwyer et al. (2002) point out, a large share of the gender gap in financial risk taking observed in the general population is attributable to differences in knowledge of markets, leaving the question open to what extent a true gender gap in risk *preferences* drives the findings of the above-listed papers.

ticipants make choices affecting their risk exposure over several rounds. To the best of our knowledge, we are the first to study the gender gap in risk taking in a naturally occurring setting that involves self-selection into and repeated interaction in a mixed-sex, but traditionally masculine domain. Such environments are interesting to study as most high-profile jobs were historically reserved for men, and women remain underrepresented in many occupational categories even today (Blau and Kahn, 2017). The setting we study is characterized by a well-defined set of rules that allow us to link choices to risk preferences without having to address confounds present in more complex environments.

Our paper belongs to a strain of literature within behavioral economics that aims to test the generalizability of results obtained in the lab by analyzing naturally occurring data. Examples include television game shows (e.g. Antonovics et al., 2009; Baltussen et al., 2016; Beetsma and Schotman, 2001; Gertner, 1993; Lindquist and Saeve-Soederbergh, 2011; Metrick, 1995; Post et al., 2008), sports (e.g. Bartling et al., 2015; Foellmi et al., 2016; Garratt et al., 2013; Pope and Schweitzer, 2011), and - similar to our study - online card games (Eil and Lien, 2014; Smith et al., 2009). Our setting is characterized by sorting and the possibility for adaptation, features that are typically missing from laboratory experiments on risk taking but are important attributes of real life decision making. While lab experiments in the recruitment phase typically do not disclose information to prospective participants on the type of task, decision or environment awaiting them, our study involves players who knowingly and willingly enter a male dominated environment to play a game in which risk taking is an important feature.[2]

Additionally, most experimental studies on risk preferences involve a one-shot decision or a limited number of repetitions due to time constraints. In our study, the number of games played is endogenously determined and for many players exceeds a thousand rounds, allowing us to study the role of experience in risky decisions. Players in our data receive immediate performance feedback that could facilitate learning: if initial gender differences in risky choices are partly driven by disparities in knowledge or skills, we should expect repetition to mitigate the gap. Further, in our setting participants observe the choices and outcomes of their (mostly male) peers, presenting a scope for adaptive behavior.

There are several other features that make our setting interesting. The dataset we analyze contains more than 4 million games by approximately 15 thousand individual players, yielding enough power to detect potentially small effect sizes, test for heterogeneity and analyze interaction effects. Players represent a wide age range, providing an interesting addition to studies based on student samples. The game we study involves a clear, objective and easy-to-interpret measure of risk taking: the choice at the beginning of each round to raise the stakes. In *Schafkopf* games, unlike in poker, this decision is unaffected by strategic concerns and as such, provides a good reflection of players' risk preferences. Alternative measures for risk taking are

[2]See Al-Ubaydli and List (2015) for a discussion on researchers' control over subjects' participation decision in lab vs. natural field experiments. For a demonstration of how information disclosure may affect the participation decision and the subsequent choices made in an experiment, consider the 'self-selection' condition in Camerer and Lovallo (1999)'s market entry game. Slonim et al. (2013) estimate the magnitude of the so-called 'participation bias' and find that lab participants are not representative of the population they were recruited from on hardly any of the hypothesized characteristics.

also available for robustness checks. Moreover, even though our context is admittedly stylized, its features resemble certain aspects of a professional work context, particularly higher echelons of the hierarchy: it is a traditionally masculine environment with a low share of women, and it involves repeated interaction among players in a setting where status incentives are relevant, choices are public, and there is scope for learning and adaptation. Finally, the data allow us to speculate about the consequences of risky choices for players' performance by analyzing the scores male and female players accumulate. Despite these appealing features of our setting, we also need to address two limitations: gender is self-reported, and stakes are hypothetical. We discuss these aspects in detail in the paper and argue that they do not compromise the validity of our findings.

Our results provide clear evidence that gender differences in risk taking may persist even in a traditionally masculine domain featuring repeated interaction with feedback. We find that female players are significantly less likely to raise the stakes than their male peers. This result is robust to controlling for the gender of players' opponents: women play as cautiously in female-only games as they do against male opponents. We observe gender differences in behavior according to alternative measures of risk taking as well. The gender gap in risky choices is not mitigated by experience and persists even among the most active players, suggesting that women do not adapt their behavior over time.[3]

Besides simply documenting gender differences in playing strategies, our data also allow us to speculate about the source of the discrepancy. We discuss whether the gender gap we observe in risk taking reflects genuine preferences or is rather a product of differences between male and female players in playing skill or beliefs about winning chances. We find that female players are as skilled as their male counterparts, making it unlikely that differences in game knowledge drive the gender gap in risk taking. Other alternative explanations we have tested but found no evidence in support of include gender differences in confidence (Kamas and Preston, 2012; Niederle and Vesterlund, 2007), stereotypes (Daruvala, 2007) or self-fulfilling beliefs about gender differences in the propensity to raise stakes (Babcock et al., 2017). Lastly, we find that female players on average obtain lower but less volatile scores, a result that is consistent with women having more risk-averse playing styles.

Our results confirm that the gender gap in risk taking observed in experimental studies is not an artifact of their special context (anonymous, one-shot decision making) and is robust to selection into the choice environment. This conclusion echoes the findings of Gerdes and Graensmark (2010) who document gender differences among expert chess players in the propensity to use an "aggressive" vs. a "solid" opening strategy, and of Böheim et al. (2016) who find differences in risky strategies between male and female professional basketball teams. We complement these papers by focusing on individual decisions in mixed-sex environments using an objective, straightforward measure of risk taking. Our study also addresses Dwyer

[3] However, we would like to caution against interpreting our results as evidence that risk attitudes are irrelevant for the decision to sort into the community. It is possible that women who join the card game platform are more risk tolerant than women in the general population such that the gap in a non-selected sample would be even larger, or that male players in our setting are also more risk loving than men in general.

et al. (2002)'s critique of observational studies of the gender gap in financial risk taking by showing that the gap is driven by true preference differences between men and women, not by confounds such as knowledge or skills.

The remainder of this paper is structured as follows. Section 2 describes in detail the card game, the online platform and the sample of players we study. Section 3 presents our results related to the risk taking and performance of male and female players. Section 4 discusses the internal and external validity of our findings. Section 5 concludes.

2 Context and data

2.1 The game

We use data from an online community for playing *Schafkopf*: a popular, traditional Bavarian card game. The game is known in other regions and countries as well (though with minor variations), under different names, such as *Doppelkopf, Skat* or *Sheepshead. Schafkopf* is a zero-sum game, played by four participants at a (virtual) table, using the unique German/Bavarian deck of cards.[4]

Each round consists of a selection stage in which players announce their willingness to initiate a game and have the option to raise stakes, and the actual playing stage during which all distributed cards are played out. *Schafkopf* is a trick-taking game: the unit of play is a 'trick' that consists of players laying down one card each in a given order and according to a set of rules. Each trick is evaluated to determine a winner or 'taker' of that trick who collects its point value. A *Schafkopf* game consists of eight tricks, and the winner(s) of the game are determined by the cumulative points collected over the course of the eight tricks. The way points map into earnings ('cents') depends on the chosen stakes. The aim of the game is to gain as many cents as possible since the sum of the cents collected constitutes a player's score.

The selection stage begins with the dealer (a role that rotates one position clockwise each round) distributing four cards to each player. The players then evaluate the strength of their first four cards and decide whether to double the stakes of the game by knocking on the table. Players make this choice simultaneously, and stakes are doubled for each knock. Afterwards, the remaining 16 cards are distributed among the four players, such that each player eventually has a hand of 8 cards. Players can then take the offensive role by actively initiating a game or the passive role by playing as a partner/opponent in a game initiated by someone else at the table. Players initiating a game have a choice between three game types: the standard two-against-two-players *Sauspiel* game and the more risky and competitive one-against-three *Wenz* and *Solo* game types. *Sauspiel* thus involves a competition between teams, while *Solo* and *Wenz* require the individual to compete alone against all other players at the table. Consequently, the stakes are also higher in the latter two game types. The partner of a Sauspiel's initiator cannot be freely chosen but is randomly determined by the initiator calling a specific suit of ace. Team composition in Sauspiel games is thus only revealed when the specific ace is played out. Team

[4]The Bavarian deck has eight different values (in increasing rank: 7, 8, 9, 10, Jack (Unter), Queen (Ober), King (König) and Ace) in four different suits (in decreasing rank: acorn, grass, heart, bell) each, with the Queens, Jacks and heart cards being trumps.

composition for Wenz or Solo games is immediately revealed as the initiator plays against the three other players.[5]

Figure A1 in Appendix A shows a screenshot from the selection stage in the online game: each player is asked if they want to initiate a *Sauspiel*, a *Wenz* or a *Solo* game, or if they prefer not to initiate a game at all (*"Weiter"* i.e. pass). Passing does not result in dropping out of the game: as long as at least one person initiates a game, all four players at the table will join as partners or opponents. The highest announced game type will be played (*Solo* being the highest type, followed by *Wenz* and *Sauspiel*) and if multiple players want to initiate the same game type, the player closest to the dealer will be given priority. If nobody initiates a game, the cards are reshuffled and a new round begins.

The actual playing stage consists of eight tricks such that every player has to contribute one card to each trick. Before playing the first trick, players in the opponent role are allowed to give a "Contra", which doubles the stakes in the game. "Contra" also serves as our proxy for risk taking. The player sitting to the left of the dealer starts the first trick and subsequent tricks are started by the winner of the last trick. The other players then all have to lay down a card in clockwise order according to a set of rules.[6] The player who contributes the card with the highest rank wins the trick.[7]

To determine the winner of the game, each card is allocated a point value, with a total of 120 points for the whole card deck.[8] To win the game, the initiator (together with his/her partner in case of a *Sauspiel* game) needs to win tricks worth at least 61 points in total. Therefore it is possible to win five out of eight tricks but still lose the game if these tricks do not contain enough points. Each player of the losing party has to pay 10 cents to the winner(s) when losing a *Sauspiel* game (the initiator and their partner thus earn the same) and 50 cents when losing a *Wenz* or *Solo*. The amounts are higher when players win with a large margin.[9] Finally, as mentioned earlier, each knock on the table after the distribution of the first four cards and each "Contra" results in a doubling of payouts. The cents that players accumulate over the rounds constitute their score.

After each round, each player decides if they want to stay in for another round or leave the table. Distribution of new cards only starts when four players sit at the table.

2.2 The online platform

Our data was provided by *sauspiel.de*, the largest online *Schafkopf* gaming community. The platform was founded in 2007 by a group of four students with the goal of bringing this tra-

[5]The difference between Solo and Wenz is that only Jacks count as trumps for a Wenz game while the Solo is more similar to the Sauspiel in that Queens, Jacks and one designated color count as trumps.

[6]Players have to contribute a card of the same color as the first played card of the trick, but can play any other card if they don't possess a card of the same color. If the first played card of the trick is a trump card, the other players also have to play a trump card if they still possess one.

[7]If a trick contains multiple Queens or Jacks, their rank order is determined by suit.

[8]11 for the Ace, 10 for the 10, 4 for the King, 3 for the Queen, 2 for the Jack, and 0 for all other cards.

[9]The above amounts are increased by 10 cents if the losing party obtained less than 30 points and by 20 cents if no tricks were won by the losing party. Furthermore, if the winning party had a sequence of at least the three highest trumps, the sum is increased by 10 cents per trump. As we discuss further in Section 2.2, payoffs in the setting we study are hypothetical.

ditional card game to the online world. The founders implemented the online version of the game with exactly the same rules as for regular *Schafkopf*, lowering entry (i.e. learning) costs for experienced players of the game. The platform became popular rapidly and has already hosted more than 500 million games as of 2015. Registration and the use of the platform is free. Virtual 'cents' collected in the online games have no monetary value, but increase players' status on the platform. First-time players have to register a profile on the platform, which includes choosing a pseudonym (user name) and customizing a male or female avatar. Players can also voluntarily report their gender, date of birth and their ZIP code at registration.

As *Schafkopf* is always played by a group of four players, each player joins a virtual table of four. Figure A2 in Appendix A shows the table selection stage: players can either set up a new table by clicking on the plus symbol and wait until three other players join, or join a table with less than four players. The only available information on potential playing mates at this stage is their pseudonym. Tables usually fill up within a few seconds and no indicator of other players' past performance is displayed at this point, thus there is little room for strategically selecting the set of players with whom to share the table.[10]

Once four players have joined a table, the regular game begins: as described in the previous section, players decide if they want to raise the stakes by knocking after the first four cards are distributed, then announce if they want to initiate a game after all eight cards have been dealt, and if at least one player announces a game, all eight tricks are played out consecutively. Players make their choices being informed of their own card quality, the past performance of other players (indicated by their cumulative scores displayed below their user names, as shown in Figure A1), and the gender composition at the table (represented by the avatars). The median game duration including the selection stage is only 81 seconds, so players usually stay at the same table for multiple game rounds. If one player leaves the table, the remaining players have the option to stay at the same table and wait until a new participant joins them. As we have seen in Figure A1, each player's cumulative score is publicly displayed on the screen during the games and thus influences the status in the community. The displayed score provides an imperfect indication of a player's skills and performance because these scores can be reset to zero whenever they have fallen below zero.

2.3 Data

This section introduces our dataset in more detail, providing an overview of the variables recorded in our dataset, and a discussion of the operationalization of the key variables (gender and risk taking) in our study.

[10]Players' decision to stay at a table for several rounds or to leave the table and join a different one could, in theory, be affected by the strength of players at their original table. However, there is no guarantee for players that their new table will feature easier opponents. Moreover, if their opponents are also strategic in their choice of staying or leaving, then each player only has limited influence over the quality of players at their tables. In our data, players spend on average 18 rounds at a table before either taking a break or moving to an entirely new table with three new opponents (and 7.5 rounds on average before one of their opponents is replaced). As such, we expect table selection to be of limited concern for our results.

We observe all games played between September 5th 2007 and January 9th 2008.[11] For each round during these five months, we observe the type of game played (*Sauspiel*, *Wenz* or *Solo* or no game if everyone "passed"), who initiated the game and who was in a partner (for *Sauspiel*) or opponent role. We can also see who raised the stakes of the game by "knocking" and in which order players announced their game choice (i.e. players' position at the table). For each initiator we observe whether they were given "Contra" by their opponent(s), but the identity of the exact opponent(s) who gave the "Contra" is not recorded. Therefore the individual decision to give "Contra" cannot be analyzed in our setting.

If a game is played, our dataset also contains the exact cards played out by each player, allowing us to assess whether good or bad cards were dealt to the players. We calculate a joint measure of card quality by regressing multiple indicators of card value on the probability of winning a *Sauspiel* and then using the coefficients to derive the joint measure of card quality.[12] If all players decide to "pass" and cards are therefore not played out, the distribution of cards is not recorded. For each game, we observe the winner(s) and the number of points achieved (between 0 and 120) as well as the final score in cents for each participant. Our dataset contains players' *actual* scores, which we can use to infer their *displayed* scores.[13]

We use the gender of the avatar players have chosen as a proxy for their true gender. For approximately 15% of the players in our sample we also observe the gender they report upon registration. Among the 3,199 users for whom this information is available and can be matched with game data, in 97.5% of cases the reported gender coincides with the avatar's gender. 58% of the mismatches result from users registered as women having male avatars. In Section 4, we discuss our measurement of gender in more detail, and present evidence in support of the claim that avatar gender provides an accurate representation of player gender.

We treat players' decision to raise the stakes by knocking as our primary measure of risk taking. As discussed in Section 2, at the beginning of each round, upon observing the first four cards dealt to them, players simultaneously decide whether or not to raise the stakes of the game by "knocking" on the virtual table. Other than doubling the stakes, knocking does not affect the ensuing game in any other way. Knocking thus increases the variance of the game's potential outcomes while leaving the probabilities of winning unchanged.[14]

[11] Our dataset also contains information on the first 1.6 million games played on the platform before 5 September 2007, but this data are incompletely recorded and lack important variables such as card quality. For this reason we decided to exclude these observations from our analysis. For players who registered before September 2007, we constructed the variable *number of games played* to take into account these initial games as well, so our analysis of experience does not suffer from left-censoring.

[12] The indicators are the cumulative point value of all owned cards as well as dummy variables for the number of trump cards, the number of suits a player does not possess (as this allows to go in with a trump if another player plays this suit), the number of aces, and the consecutive number of highest trumps in a row. The results presented in the paper are robust to including the different measures of card quality separately instead of the joint card quality measures.

[13] As discussed in Section 2.2, players can reset their displayed scores to zero whenever it is to fall below zero.

[14] An example: As noted in Section 2.1, in a *Sauspiel* game by default each member of the losing party has to pay 10 cents to the winners. This amount increases to 20 cents in case one player at the table knocks, to 40 cents if two knock, etc. Unlike in poker, the decision to knock is unaffected by strategic considerations. First, players do not rely on their opponents' knocking behavior to decipher the quality of their cards the way they would in poker. This is because players have better ways to infer others' card quality during the playing stage: remember that all cards of the deck are in hands, making it easier to count cards. Also, if a player contributes a card of a

Besides raising the stakes by knocking, we also consider two alternative measures for risk taking. First, as discussed in Section 2, players also have an opportunity to raise the stakes at the beginning of the playing stage: after all the cards are dealt, but before the first trick is played out, players in the opponent role can double the current stakes by giving a "Contra". Our data only contain information on "Contras" from the point of view of the player who initiated the game. We thus cannot calculate the gender gap in the propensity of a player in the opponent role to *give* a "Contra". Instead, we can compare initiators' likelihood of *receiving* a "Contra" by the gender composition of their group of opponents.[15] Second, we focus on players' choice to initiate a game. We would like to emphasize that the choice between "passing" and actively initiating a game is a complex decision that is likely affected by competitive and other-regarding preferences and beliefs about others' actions as much as it influenced by risk preferences. The following analysis is thus intended only as supportive evidence for our main result derived from the analysis of increasing stakes.[16]

2.4 Descriptive statistics

The dataset contains over 4 million games recorded from the perspective of each player at the table, resulting in 16,655,344 observations in total. Our data is generated by more than 15 thousand individual players. The share of female players in the community is low, around 8.5%, reflecting Schafkopf being a traditionally "masculine" activity.[17] Female players are, however, very active on the platform: as shown in Table 1, the number of games per player is much higher for women (mean: 2,340) than it is for men (mean: 1,391). Consequently, women generate 13.52% of all observations in our dataset (see also Figure B1 in Appendix B for the distribution of the number of games per player).

As mentioned earlier, our dataset contains limited information about the characteristics or background of the players. A subsample of 2,271 registered users voluntarily reported their age and the ZIP code of their residence. Among this group we observe no substantial difference between men and women in terms of age: as reported in Table 1, the mean age is 30.19 for men and 29.80 for women. Figure B2 in Appendix B, depicting the age distribution of our players, shows that our sample is very diverse, with some players younger than 15 and others older than 70 years. A little more than 10% of the registered users come from Munich, and women are slightly more likely than men to live in the state capital.

Even though we do not observe the employment or marital status of the players in our sample, we can still draw some cautious inferences regarding their lifestyles based on the hour

different color to a trick, his opponents can be certain that he has no card of the called color in hand. Second, knocking is uninformative of important dimensions of a player's cards (e.g. the color of cards the opponent is strong in), and it is based on the strength of the first four cards only. Finally, players' choice of which card to contribute is governed by strict rules, and depends more on the point value of the trick to be won than on the perceived general strength of the opponent.

[15] We are grateful for an anonymous reviewer who suggested this strategy to us.

[16] Based on the finding that gender differences in the willingness to compete are largely reduced when players compete in teams (Healy and Pate, 2011), at least in case of the two-against-two Sauspiel games we expect risk preferences to be the dominant factor in the decision to initiate a game.

[17] To be precise, the share of players *with a female avatar* is low - as pointed out in the previous subsection, we use avatar gender as a proxy for actual gender.

of the day when they play games on the online platform. Figure B3 in the Appendix suggests that there is no systematic difference between men and women in the game community: both are most likely to play in the evening hour. Moreover, we find no indication for women playing more often during the typical working hours, suggesting that they are as likely as men to be employed (assuming, perhaps optimistically, that people do not log onto the platform during their work hours).

Table 1: OVERVIEW OF THE DATASET

	MALE			FEMALE			DIFF.
	Median	Mean	SD	Median	Mean	SD	p-value
Number of rounds per player	238	1,391	2,878	447	2,340	4469	0.000
Age[1]	28	30.19	10.63	27	29.80	10.63	0.514
Location Munich[1]	-	0.128	0.33	-	0.178	0.38	0.033
Number of players		13,941			1,291		
Number of observations		14,402,768			2,252,576		

[1]: Based on a subsample of 2,271 players.
The final column reports p-values from t-tests with unequal variances comparing male and female players.

Table B1 in Appendix B tabulates the gender composition of tables for all rounds. In more than half of the cases (55.5%), there were only male players at the table. Approx. 36% of rounds involved one female player, 8% involved two, less than 1% involved three, and a mere 0.03% of the rounds were played by four women at the table. This latter share, however, still corresponds to 1,110 unique rounds, allowing us to study behavior in female-only environments.

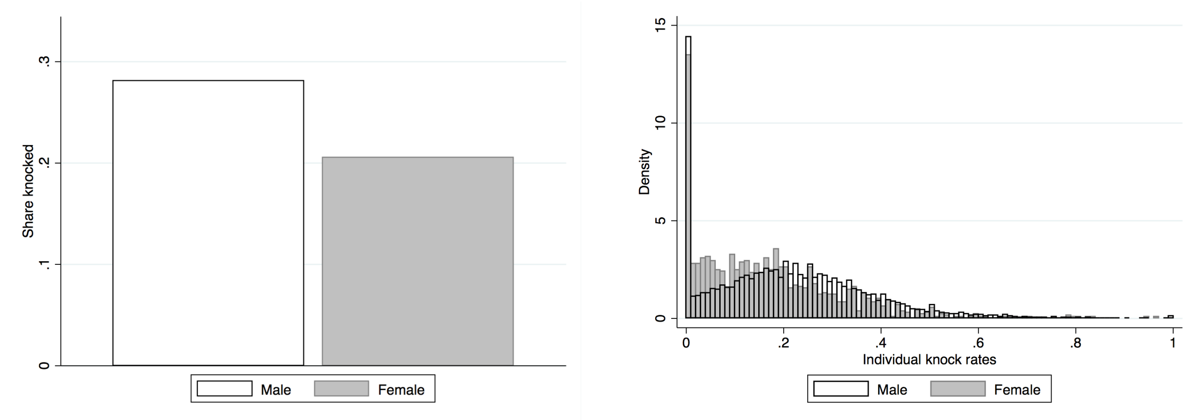
3 Results

In this section we assess whether gender differences in risk attitudes observed in the general population can also be detected in our sample of card game players who voluntarily joined a traditionally masculine environment to participate in a contest that involves risk-taking. We begin our analysis in Section 3.1 by studying choices in the selection stage of the game that we argue reflect risk preferences. As explained in Section 2.3, our main focus is on players' decision at the beginning of each round to raise the stakes of the game. In Section 3.2 we consider gender differences in "Contra" behavior and the choice to initiate a game as robustness checks. We first analyze all games in our sample combined, then in Section 3.3 we explore how experience affects gender differences in risk taking. In Section 3.4 we speculate whether it is possible to attribute the observed gender gap in risk taking to differences between men and women in attitudes towards uncertainty. To do so, we need to exclude possible alternative explanations such as gender disparities in playing skills and confidence. We are also interested in the consequences of players' choices: if women indeed take less risks, how does this affect their success in the game? In Section 3.5 we explore how gender differences in risky choices translate into disparities in scores accumulated, studying both the mean and the variance of scores.

A brief comment on terminology: throughout the paper, "focal player" refers to the individual whose perspective we consider in the analysis (recall that all rounds are recorded from the viewpoint of all four participants), and "opponents" denote the three other players sitting at the table.[18] "Game" denotes a round in which cards were actually played out (rather than re-shuffled because no-one initiated a game). The unit of analysis in the study, unless otherwise specified, is thus defined by the combination of the focal player and the round.

3.1 Documenting a gender gap in risk taking

Panel 1a of Figure 1 compares the propensity to raise the stakes between male and female players over all rounds in our dataset. Such comparison is meaningful given the large number of observations and the fact that cards are dealt at random so it is reasonable to assume that there is no difference in the overall quality distribution of players' first four cards by gender. A clear gender gap emerges from the figure: while men on average knock 28.1% of the time, women do so in only 20.5% of rounds, a highly significant difference of 7.6 percentage points or 37 percent (p-value < 0.000). Panel 1b of Figure 1, in turn, depicts the distribution of individual 'knocking rates', calculated for each player as the share of rounds in which they chose to raise the stakes over all rounds they were involved in. The figure suggests that the mean difference in overall knocking propensity is not driven by a few particularly risk loving men, rather, for females the entire distribution is shifted to the left compared to male players. A Kolmogorov-Smirnov test confirms that the two distributions are significantly different from each other (p-value $<$ 0.001).[19]



(a) Propensity to knock

(b) Distribution of individual knock rates

Figure 1: Gender differences in increasing stakes by knocking

[18]We decided to use the general term "opponent" because even though in the *Sauspiel* game one of the fellow players will actually be the initiator's partner, in the selection stage it is not yet known who this partner may be.

[19] Surprisingly, Figure 1b shows that the share of players who never knock is *higher* among male than among female players. We believe this curious finding is attributable to the fact that players need some time to understand the platform and explore its features, so many players only learn about the option to knock once they have gained a little experience. Players who quit the platform after just a few games thus tend to be the ones whose knocking rates are zero or very low. As Figure B1 shows, the share of men who only play a few games on the platform is much higher than the corresponding share among women, explaining why we observe a higher share of men than women who never knock. To test this explanation, we re-estimate Figure 1b, excluding the first 11 rounds (corresponding to the first decile of the distribution of the total number of rounds played on the platform) and calculating the individual knocking rates among the remaining players and rounds. The results, presented in Figure B4 in Appendix B, indeed support our intuition: once the learning period is excluded, women are overrepresented among players who never knock.

In the following we show that the raw gender differences observed in the propensity to raise the stakes are robust to the inclusion of various control variables. We begin by pointing out that due to data limitations the most important predictor of knocking, the quality of the first four cards a player receives, is unobserved to us: as noted in Section 2.3, players' cards are only recorded in our dataset in case cards are actually played out and not in rounds where all players choose to "pass". We therefore exploit the fact that card quality is negatively correlated between players by design: if many opponents have good enough cards to knock, the focal player is likely to be left with the lower quality cards of the deck. Players decide to raise the stakes simultaneously, so they themselves cannot base their knocking decisions on opponents' choices at this stage. In our analysis, for each focal player we use their opponents' decisions to raise the stakes as a(n admittedly imperfect) proxy for card quality. Later in this subsection we discuss the impact of card quality on game initiation rates without having to rely on such a shortcut.

Table 2 displays average marginal effects from probit models explaining the propensity to raise the stakes, where each column corresponds to a different model specification.[20] Unless otherwise specified, throughout the analysis we use cluster robust standard errors to correct for the fact that error terms may be correlated on the player level.[21]

Column (1) reiterates the result that female players are on average 7.6 percentage points less likely to knock than males, who have a 'baseline' knocking rate of 28.1% (*"Predicted probability male players"* reports the estimated average probability of knocking in a given round for a male player). The second column shows that the gender gap is insensitive to the focal player's position at the table or the presence of a female opponent.[22]

The model presented in column (3) analyzes the role of player characteristics by controlling for experience on the platform (i.e. the (log) number of prior rounds the player participated in) and the focal player's strength compared to the opponents (i.e. their relative rank in terms of displayed score).[23] While player rank is found to be significantly correlated with the choice to raise the stakes, its inclusion in the model decreases the estimated gender difference only slightly, to 6.8 percentage points.

[20] The corresponding estimated coefficients from the probit models are displayed in Table B2 in Appendix B. In the main text we choose to present averages of individual marginal effects instead of marginal effects at means mainly to avoid referring to nonexistent or inherently nonsensical observations such as "88% male" (see e.g. Bartus (2005)). Given the large number of observations in our sample AME and MEM lead to very similar estimated effects (Greene, 2008). Estimates from linear probability models also lead to similar conclusions (results from OLS regressions available from the authors upon request).

[21] We note that there is another non-nested dimension of potential clustering as there is strategic interaction in each round between the four players at the same virtual table. We chose clustering standard errors on the player instead of the table level as this approach results in more conservative estimates. An alternative would be a multi-way clustering approach (Cameron and Miller, 2015).

[22] In Table B4 in Appendix B we explore whether male and female players respond differently to their position at the table or to the presence of a female opponent, and find no differential response by gender (note that the interaction terms between player gender and position at the table are significant but negligible in size). We elaborate further on this finding in Section 3.4.

[23] While it is an interesting avenue to study how opponents' displayed scores may affect risk taking by serving as "social reference points" (Linde and Sonnemans, 2012), the fact that each focal player observes three others at the table dos not facilitate a straightforward analysis of this question in our data. Note that we only included rank as a covariate in the model to ensure that gender differences in rank do not drive the gender gap in risk taking. Players' rank depends on their cumulative score, which reflects their past knocking and game initiation behavior. Therefore, we cannot assign a causal interpretation to the estimated coefficients associated with the different ranks.

Table 2: LIKELIHOOD OF RAISING THE STAKES

	(1)	(2)	(3)	(4)	(5)
Female player	-0.076***	-0.076***	-0.068***	-0.070***	-0.086***
	(0.009)	(0.009)	(0.008)	(0.008)	(0.018)
Position at table = 2		-0.010***	-0.010***	-0.009***	-0.010***
		(0.001)	(0.000)	(0.000)	(0.001)
Position at table = 3		-0.010***	-0.010***	-0.009***	-0.011***
		(0.001)	(0.001)	(0.001)	(0.001)
Position at table = 4		-0.008***	-0.008***	-0.007***	-0.009***
		(0.000)	(0.000)	(0.000)	(0.001)
Female opponent		-0.002	-0.006***	-0.014***	-0.019***
		(0.001)	(0.001)	(0.001)	(0.002)
Num. rounds played (log)			0.003	0.004**	-0.002
			(0.001)	(0.001)	(0.003)
Rank = 2			-0.029***	-0.024***	-0.024***
			(0.003)	(0.003)	(0.006)
Rank = 3			-0.052***	-0.044***	-0.046***
			(0.004)	(0.004)	(0.008)
Rank = 4			-0.081***	-0.069***	-0.073***
			(0.005)	(0.005)	(0.011)
Num. opponents knocked = 1				-0.119***	-0.120***
				(0.001)	(0.001)
Num. opponents knocked = 2				-0.202***	-0.205***
				(0.001)	(0.002)
Num. opponents knocked = 3				-0.242***	-0.251***
				(0.002)	(0.004)
Age (Q2)					0.007
					(0.010)
Age (Q3)					0.017
					(0.011)
Age (Q4)					0.023
					(0.015)
Location Munich					-0.018
					(0.010)
Predicted prob. male players	0.281***	0.281***	0.280***	0.280***	0.287***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)
Number of observations	16,655,344	16,655,344	16,655,344	16,655,344	5,499,377
Number of players	15,232	15,232	15,232	15,232	2,271
Pseudo R^2	0.00307	0.00315	0.00748	0.0306	0.0331
Log likelihood	-9.699e+06	-9.698e+06	-9.656e+06	-9.431e+06	-3.129e+06

Note: The table displays average marginal effects from probit models where the dependent variable is the indicator for raising stakes. Indicator variables for age quartiles are included in column (5), the youngest quartile being the omitted category.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

In column (4) we also control for the number of opponents who knocked in the given round, and find that the size of the gender gap in knocking remains virtually unchanged. Since players make the decision to raise the stakes all at the same time, the large negative effect of others' choices on the focal player's tendency to knock is not an indicator of strategic behavior, rather, as explained above, results from the negative correlation between players' card quality.[24]

Finally we test whether observable characteristics such as age or place of residence can explain the difference between men and women in risk taking. In column (5) we estimate our model on the subsample of 2,271 players who voluntarily disclosed this information at registration. While age and location do not seem to influence the likelihood of knocking, the impact of gender is higher in this subsample: the estimated gender gap increases to 8.6 percentage points.[25]

3.2 Alternative measures of risk taking

We explore whether our finding that female players follow less risky strategies than males is specific to the measure of risk taking we have chosen (i.e. the propensity to raise the stakes by "knocking"). In the following, we show that the gender gap in risk taking is robust to considering alternative proxies, such as "Contras" and game initiation, measures introduced in Section 2.3.

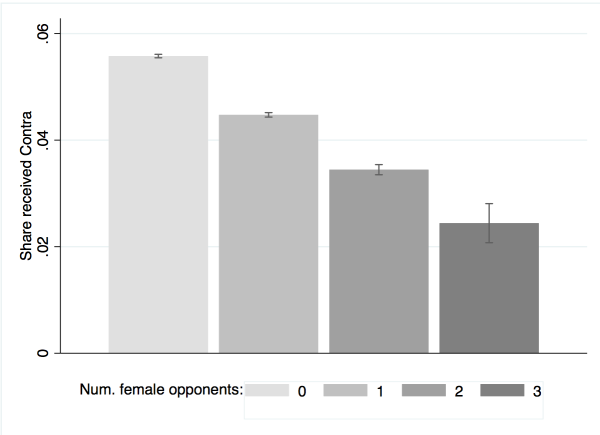


Figure 2: Likelihood of receiving Contra by number of female opponents (means and 95% CIs)

Figure 2 shows that initiators' likelihood of receiving a "Contra" decreases in the number of female opponents they face, suggesting that female opponents are less likely than males to

[24]As shown in column (4) of Table B4, there is no gender difference in the way card quality - as captured by the opponents' knocking behavior - affects the propensity to raise the stakes, either. We delay the discussion of the interaction between gender and experience until Section 3.3. Another way to show the robustness of our results is to estimate models with game fixed effects: such specifications also lead to very similar estimates of the gender gap in knocking. Calculations are available from the authors upon request.

[25]Table B3 in Appendix B compares the gender gap in the full sample and in the subsample where demographic information is available without adding extra controls for age and location, and confirms that the change in the estimated impact of gender is attributable to the sample restriction, not to the inclusion of age and location as covariates.

double the stakes at this stage.[26] Thus, the gender gap in the propensity to increase the stakes is not limited to the selection stage of the game, when players only know four out of the eight cards in their hand, but also shows up in the playing stage, when uncertainties around the quality of one's cards are resolved.

Next, we consider whether female players are also less likely than male players to initiate a game. Figure 3 compares male and female players' odds of initiating a game for different levels of card quality. By card quality, we refer to the quality of the 8 cards in the focal player's hand, calculated according to the procedure outlined in footnote 12. As mentioned in Section 2.3, card quality is only recorded in our dataset for rounds when a game was initiated, but not when all players decided to pass. We therefore cannot directly calculate initiation rates by card quality. Instead, we can exploit the large size of our dataset together with the fact that cards are dealt at random to simulate the card quality distribution we would have observed if a game had been initiated in every single round. We then contrast this *simulated* distribution with the *observed* card quality distribution in rounds when a game was actually initiated to obtain our estimated game initiation rates by card quality.[27]

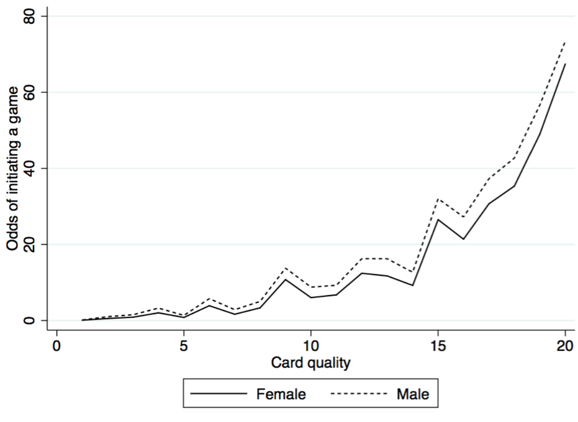


Figure 3: Odds of initiating a game by card quality quantiles, by gender

Figure 3 shows a small but persistent gender gap in game initiation rates that appears throughout the card quality distribution, including the best cards as well.[28] The gender gap

[26]The results are very similar when controlling for the gender of the focal player, their position at the table, their relative rank, their experience on the platform, and for the number of opponents who knocked. Regressions results are available from the authors upon request.

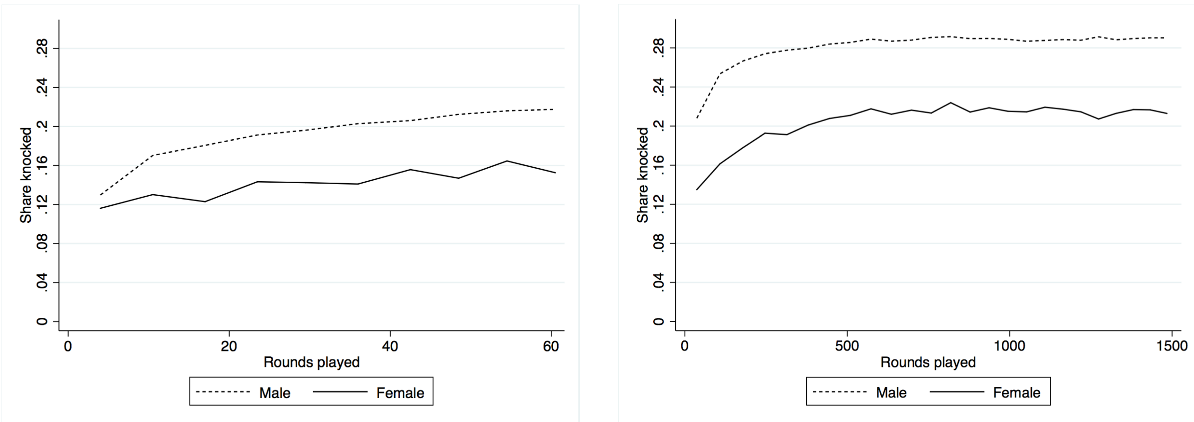
[27]As each player gets 8 out of 32 cards, there are 10,518,300 different possible combinations of 8 cards, each occurring with the same probability. We calculate a measure of card quality for each of these 10,518,300 card combinations to arrive at the simulated card quality distribution we would have observed if a game had been initiated in every single round. We estimate the quantiles of this distribution and form 20 bins such that the odds that a player's card quality in a given round falls into any one of the bins is 5%. We then compare the number of rounds when a game was actually initiated in each quality bin with the total number of rounds expected to occur in that bin to obtain game initiation rates by quality bin. To arrive at Figure 3, we repeat this exercise separately by gender.

[28]The x-axis of the graph in Figure 3 shows card quality vigintiles. The apparent non-monotonicity of initiation rates by card quality likely results from the particular way we created the single proxy for card quality. As explained in Section 2.3, our card quality measure is calculated from several different card attributes, but is likely still missing some information that is related to the choice of initiating a game. So while the odds of initiating

in game initiation rates is highly significant and robust to the inclusion of various controls (see Table B5 in Appendix B) and it increases with the riskiness of game types (see Table B6 in Appendix B).

3.3 The role of experience

In the previous subsection we considered gender differences in risk taking across all games in our dataset. Such a combined approach, however, may mask important differences in behavior over time. Given the large number of rounds players participate in, our setting allows us to assess whether repetition with feedback mitigates the gap in risky choices between men and women. In particular, we can test whether female players, observing the higher propensity of their predominantly male peers of increasing stakes and initiating games, adapt their behavior and become more similar in choices to men over time.[29]



(a) Players with ≥63 total rounds played

(b) Players with ≥1510 total rounds played

Figure 4: The impact of experience on the gender gap in knocking

We begin by a simple overview of the evolution of risky choices by experience. Figure 4 shows the odds of increasing the stakes over the number of rounds played. In order to ensure that we capture the pure effect of experience clean from the impact of sorting out of the community, we restrict our attention to players who participated in at least as many rounds as captured in the figure, such that the odds are calculated over the same number of players both in the first and the last round studied. This criterion presents a trade-off: the longer the period we observe, the smaller the subsample of players that we can observe throughout. We present figures for the first and the third quartiles of the distribution of the total number of rounds per player (corresponding to players who participated in at least 63 and at least 1510 rounds, respectively), such that our estimates are based on the most 'active' 75% and 25% of players on the platform. Panel 4a focuses on learning throughout the first 63 rounds, while Panel 4b shows how playing

a game generally increase with card quality, uncaptured differences in card quality that are correlated with the captured measures can lead to the observed apparent non-monotonic relationship.

[29]Relatedly, relative performance feedback and advice have been shown to affect the decision to enter competitions, shrinking the gender gap (Brandts et al., 2015; Wozniak et al., 2014).

strategies evolve over a longer period covering 1510 games.[30]

Table 3: THE IMPACT OF EXPERIENCE ON THE GENDER GAP IN RISK TAKING

Outcome:	Raise the stakes (1)	Initiate game (2)
Position at table = 2	-0.009*** (0.000)	-0.021*** (0.000)
Position at table = 3	-0.009*** (0.001)	-0.030*** (0.000)
Position at table = 4	-0.007*** (0.000)	-0.035*** (0.001)
Female opponent	-0.009*** (0.000)	-0.003*** (0.000)
Num. opponents knocked = 1	-0.121*** (0.001)	-0.109*** (0.000)
Num. opponents knocked = 2	-0.206*** (0.001)	-0.170*** (0.001)
Num. opponents knocked = 3	-0.251*** (0.002)	-0.203*** (0.001)
Rank = 2	0.003*** (0.001)	0.002*** (0.000)
Rank = 3	0.008*** (0.001)	0.005*** (0.000)
Rank = 4	0.010*** (0.001)	0.009*** (0.000)
Num. rounds played (log)	0.012*** (0.001)	-0.006*** (0.001)
Female player * Num. rounds played (log)	-0.003 (0.002)	0.001 (0.001)
Number of observations	16,655,344	16,655,344
Number of players	15,232	15,232
R^2 (within)	0.0304	0.0281

Notes: The table displays estimated coefficients from fixed effects (within) regressions. The dependent variable in column (1) is the indicator for raising stakes and in (2) the indicator for initiating a game. Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05

Neither panels of Figure 4 support the hypothesis that the gender gap in risk taking is mitigated by learning: if anything, in Panel 4a we observe a widening of the gap over rounds, while in Panel 4b we see a persistent and rather constant gender gap among the most active and experienced players on the platform. The smaller initial gap and the increase in knocking rates observed in Panel 4a might be attributable to players requiring some time to understand the features of the platform (see footnote 19). Alternatively, it might signal gender differences

[30]In Figure B5 in Appendix B we also document the evolution of knocking over the first 11 games among players who played at least 11 rounds, covering 90% of our sample. Note, however, that most players need some time to familiarize with the platform, resulting in a lot of noise in choices over the first few rounds.

in the evolution of risk taking, similar to gender differences in response to wins and losses observed in tournament performance (Gill and Prowse, 2014), challenge seeking (Buser, 2016b) and tournament entry dynamics (Buser, 2016a).

Under the assumption that attrition from the community is only related to time-invariant unobserved characteristics of players we can exploit the panel nature of our data and estimate a model with individual fixed effects to study the impact of experience on male and female players' risky choices. The conclusion from Figure 4 is supported by results from within regressions presented in Table 3. Note that in a fixed effects regression we are unable to estimate the impact of time-invariant characteristics such as gender, so in this model we are limited to analyzing its interaction with rounds played. The gender gaps in the propensity to knock (column (1)) or initiate games (column (2)) do not change over time, as indicated by insignificant interaction terms in each model between the female dummy and the (log of the) number of rounds played.

3.4 Alternative explanations

In our analyses so far we have demonstrated that male and female players differ in the amount of risk they choose to take. We have also seen no evidence of adaptation: the gender gap in risky choices does not seem to disappear over time and it is also detectable among the most active and experienced players on the platform. It remains a question, however, to what extent the documented gap in risk taking is driven by risk preferences or by other underlying differences between the genders. We start by discussing a potential explanation inspired by Dwyer et al. (2002) who show that the gender gap observed in financial risk taking is to a large extent attributable to investor knowledge of financial markets and investments. Translating their findings to our setting, we need to ensure that gender differences in the propensity to raise the stakes are not mediated by gender disparities in game knowledge and playing skills. We also consider gender differences in confidence as alternative explanation.

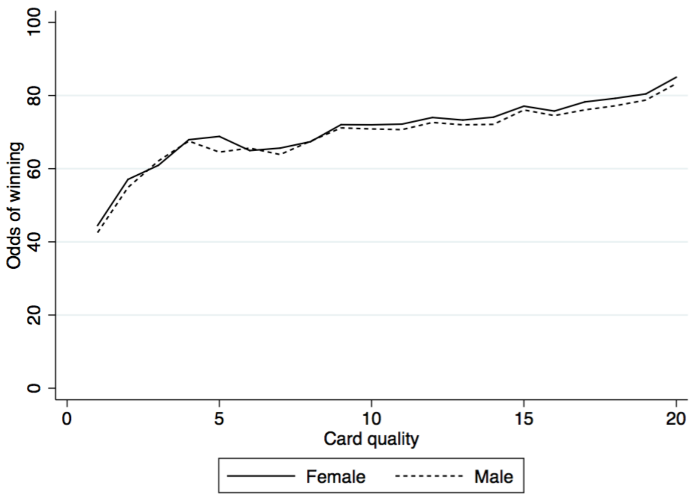


Figure 5: Odds of winning a self-initiated game by card quality quantiles, by gender

In assessing the claim that men and women differ in their knowledge and expertise of the game we are unable to rely on actions in the selection stage: since players' cards are only recorded in our dataset in rounds when a game is initiated (and not when all players chose to 'pass'), it is not possible for us to measure ability through the optimality of knocking decisions. Instead, we test for gender differences in skills by assuming that player ability is correlated across the two stages of the *Schafkopf* game. In particular, we assume that more able players both make better choices in the selection stage (i.e. in deciding whether to raise the stakes and initiate a game) and are also better at playing out cards, taking tricks and collecting points in the ensuing games. This premise allows us to use conditional winning probabilities in games as indicators for "*Schafkopf* skills" more generally.[31] Figure 5 focuses on games initiated by the focal player and compares for different levels of card quality male and female players' probabilities of winning. According to this measure, there is no clear gender difference in ability: irrespective of the quality of players' cards, men and women seem to have very similar odds of winning games they started themselves.

Table 4: SUCCESS IN THE GAME

Role:	Initiator (1)	Partner (2)	Opponent (3)	Solo 4^{th} (4)
PANEL A: Winning				
Female player	0.001	-0.002*	-0.000	0.001
	(0.002)	(0.001)	(0.001)	(0.002)
Predicted prob. male players	0.766***	0.799***	0.243***	0.303***
	(0.001)	(0.000)	(0.000)	(0.001)
Adjusted R^2	0.136	0.153	0.105	0.0865
Log likelihood	-1.403e+06	-864288	-3.438e+06	-239488
PANEL B: Points				
Female player	0.171	-0.245***	0.071	0.240
	(0.158)	(0.052)	(0.066)	(0.151)
Constant	70.776***	84.277***	-0.930***	84.206***
	(0.318)	(0.228)	(0.182)	(0.303)
Adjusted R^2	0.258	0.270	0.195	0.175
Controls	✓	✓	✓	✓
Number of observations	2,984,123	2,032,717	6,919,652	427,093
Number of players	14,302	14,049	14,991	12,086

Note: The table displays average marginal effects from probit models (Panel A) and estimated coefficients from an OLS regression (Panel B). The dependent variable is the indicator for winning a game (Panel A) and points collected in a game (Panel B). Samples restricted to players in the following positions: (1): initiator (2): partner (3): opponent (4): opponent in a Solo game at the fourth position at the table. The following covariates are included in all models: female player, position at table, female opponent, number of opponents who knocked, number of rounds played (log), card quality measures: own, partner's (when applicable), opponents', Contra received.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

[31] A second method to assess the role of playing skills in explaining the gender gap in risky choices relies on the assumption that repetition and feedback results in improved decision making. In particular, if gender differences in the propensity to raise the stakes or initiate games largely result from either male or female players making disproportionately more mistakes, then we should expect learning to mitigate the gap as players converge to the profit-maximizing strategy over time. However, as discussed in Section 3.3 we find no decrease in the gender gap with experience.

Column (1) of Table 4, presenting average marginal effects from a probit model explaining winning probabilities, confirms this interpretation: we find that even after controlling for various player and game characteristics, women are as likely to win self-initiated games as men.[32] Columns (2) and (3) present the corresponding estimates from specifications focusing on games played as a partner or an opponent, respectively. We find no gender differences in winning probabilities in these roles, either (while the point estimate for female in column (2) is negative, it is negligible in size and only significant at the 5% level).

The above results, however, suffer from potential endogeneity problems. Players self-select into the initiator role (and implicitly also into the other roles, too) and this sorting is unlikely to be random: the unobservable characteristics that influence initiating behavior could also affect winning probabilities and could be correlated with gender. To address this issue, we consider performance in a situation where the scope for self-selection is virtually absent: among players in the fourth position at the table playing as opponent in a *Solo* game. To see why the opponent role is exogenously assigned in this case, consider the following. Players in the fourth position are the last to announce their intent to initiate games. The only way these players could prevent being an opponent is to call a 'higher' game type than the one already initiated. Since *Solo* is the highest type, there is no way for the last player at the table to avoid the opponent role in this game type.[33] Column (4) of Table 4 analyzes the winning odds of opponents in the fourth position of *Solo* games and confirms that there is no significant gender difference in performance in this arguably exogenous role, either.

Finally we test for gender differences in ability using a slightly different outcome measure: instead of the binary indicator for winning a game, we focus on the number of points a player collected in a game. As a reminder, players can receive between 0 and 120 points in a game, depending on the point value of the tricks they won. As we see in Panel B of Table 4, there is no indication of gender differences in ability according to this continuous measure of performance, either. The only exception is the role of Partner, where women earn on average 0.245 points less than men, a difference that is highly statistically significant but small in size (note that on average players collect more than 80 points in the partner role).

Having found no skill differences between men and women that could explain their different strategies, we consider an alternative explanation for the gender gap in risk taking related to confidence. In our setting by confidence we refer to players' propensity for under- or overplacement, i.e. their self-evaluation of performance relative to others (Moore and Healy, 2008).[34] In the following we discuss to what extent women's lower propensity to raise the stakes or initiate

[32]Estimated coefficients from probit models are presented for all covariates in Table B7 in Appendix B.

[33]Another way to address the endogeneity of winning probabilities in self-initiated games would be to estimate a Heckman sample selection model. However, we did not find a variable in our dataset that would influence game initiation but not performance in the game. In the absence of such an exclusion restriction, identification of the estimated parameters depends on untestable functional form assumptions. Moreover, the most important predictor of winning, card quality, is only observed in cases where a game is actually initiated. Adding it as a control variable in the main but not in the selection equation would potentially make it an endogenous covariate. Based on these considerations we decided not to include a probit model with sample selection in our analysis.

[34]For a discussion on different definitions of overconfidence from an economics perspective, please refer to e.g. Hvide (2002).

games is driven by their less optimistic evaluation of winning chances. Women are consistently shown to be less overconfident than men (Croson and Gneezy, 2009; Kamas and Preston, 2012; Niederle and Vesterlund, 2007). Our data offer no direct way to measure and control for players' beliefs, so instead we review suggestive evidence that confidence is not the main driver of the gender gap in risky choices. First, we argue that players' publicly displayed scores serve as signals on relative ability and affect players' expected chances of winning. In column (3) Table 2 we indeed find the likelihood of knocking decreases in one's relative rank - but the inclusion of rank as a control only reduces the estimated gender gap very slightly. Second, since player ability arguably has the highest impact on winning chances in case cards are of intermediate quality (even the most able player is likely to lose a game with bad cards and players with low skills may still win when they are dealt exceptionally good cards), we expect confidence in one's own skills to influence players' choices the most when their cards are neither too good nor too bad. Relatedly, if confidence is the major source of the gender gap in risk taking, we predict the gap to be the largest in case of intermediate card quality. We find no evidence for such patterns either in increasing the stakes (the interaction between gender and the number of opponents who knocked, our proxy for card quality, is insignificant, see column (4) of Table B4) or in initiating games (see Figure 3).

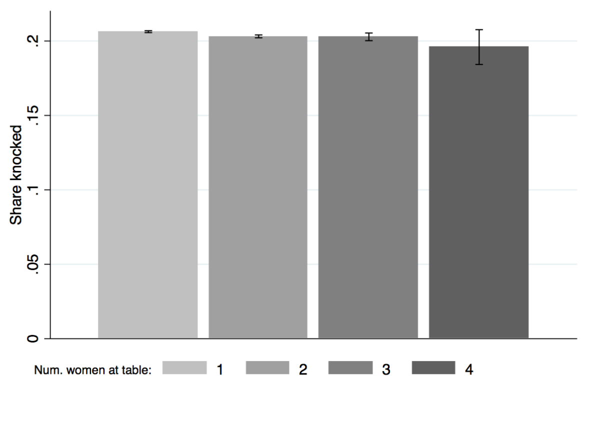


Figure 6: Female players' propensity to raise the stakes by number of women at the table (means and 95% CIs)

Next, we discuss whether differences in beliefs about male and female players' propensity to knock might be driving our results (see e.g. (Daruvala, 2007) for a discussion on gender stereotypes about risk taking). In particular, the observed gender gap in raw knocking rates could be consistent with an equilibrium where all players have the same preferences over risk exposure but everyone expects women to raise the stakes less often (Babcock et al. (2017) demonstrate such an asymmetric equilibrium). Given that a man by construction faces one more female opponent than a woman sitting at the same virtual table, these beliefs could cause men at mixed-gender tables to expect lower knocking rates from their opponents than women at the same table do. As a result, even if both genders had the same optimal risk exposure, expectation of others' behavior could lead men to knock more often on average. Our detailed analysis of the data, however, does not support this explanation: as we have seen in Table B4, both female and male players respond to the presence of a woman at the table by *decreasing* their knocking rates. Further, as Figure 6 shows, women are no more likely to knock in female-only games than they are when facing three male opponents (if anything, they are slightly less likely to raise the stakes when they play exclusively against women).

Finally, we need to rule out the possibility that the overall gender gap in knocking and initiation rates documented in Section 3.1 merely reflects the choices of the most active players in the sample. To exclude the possibility that our results are driven by a small share of players who play thousands of games on the platform, we re-estimate the regressions explaining the propensity to increase the stakes and to initiate games using weights that are inversely proportional to the focal players' total number of rounds played. In other words, we assign larger weights to players who participate in fewer rounds to ensure that our estimates do not merely reflect the preferences of the most active players. As shown in columns (1) and (2) of Table 5, this approach results in slightly smaller estimates for the gender gap, however, the differences between male and female players are still highly significant, and, especially in case of knocking, important in size. Column (3) of Table 5 finds no gender gap in winning probabilities using weights to correct for differences between players in the total number of rounds they participated in, either.

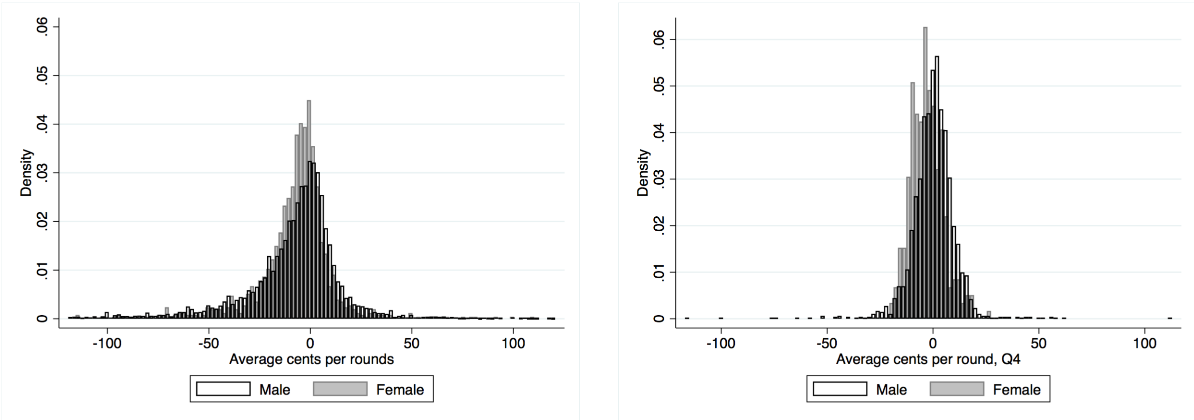
Table 5: REGRESSIONS WITH WEIGHTS

Outcome:	Raise the stakes (1)	Initiate game (2)	Win self-initiated game (3)
Female player	-0.056***	-0.026***	-0.013
	(0.004)	(0.003)	(0.007)
Controls	✓	✓	✓
Predicted prob. male players	0.214***	0.189***	0.681***
	(0.001)	(0.001)	(0.002)
Number of observations	16,655,344	16,655,344	2,984,123
Number of players	15232	15232	14302
Pseudo R^2	0.0416	0.0256	0.190
Log likelihood	-6379	-6091	-1232

Note: The table displays average marginal effects from probit models where each observation is weighted by the inverse of the total number of rounds the focal player participated in. The dependent variable is (1): raise the stakes (2): initiate a game (3): win a game initiated by the focal player. Covariates included in all three models: female player, position at table, female opponent, number of rounds played, number of opponents who knocked. Additional covariates: (1) and (2): rank in displayed score (3): card quality controls, opponents gave Contra.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

3.5 The consequences of different strategies

In our analyses so far we have established the existence of a gender gap in risk taking and argued that it reflects differences in preferences rather than in skills or in confidence. In this subsection we proceed to analyze the consequences of male and female players' strategies for their 'earnings' (as measured by their scores). As mentioned before, a minor disadvantage of our setting is that players' cards in the selection stage are only partially recorded, making it hard to evaluate the optimality of players' knocking decisions. Therefore we cannot conclude whether women raise the stakes and initiate games "too little" and/or men do so "too much" (cf. Niederle and Vesterlund (2007)'s seminal study). Instead, we can infer the *relative profitability* of their playing strategies by comparing scores by gender.



(a) Full sample

(b) Most active players (4th quartile)

Figure 7: Average cents accumulated per round

As discussed in Section 2, a player's score at any point in time is the sum of cents he or she has accumulated over all previous rounds, while cents collected in each game depend on the outcome (won or lost) and the number of players who decided to raise the stakes: each 'knock' doubles the payoffs of the given round. The margin with which a player won, the type of the game played (stakes are higher in *Wenz* and *Solo*) and whether anyone gave *Contra* also influence the cents collected in a round.[35] Score is the most important indicator of success in the community since it is the only publicly displayed measure of performance: as shown in Figure A1 in Appendix A, besides a player's pseudonym and avatar, the only information readily available to opponents is a player's score.

In Figure 7, we present the distribution of players' average cents obtained per round. In Panel 7a we show results for the full sample, while in Panel 7b we restrict our attention to the most active quartile of players (i.e. those who played at least 1510 games in total) to control for the impact of experience on success in the game. In both samples, men tend to outperform women in general. This result may seem surprising at first given that we found no gender differences in conditional winning probabilities or points collected per round in the different roles. In the following, we explore two channels through which gender differences in the propensity to raise the stakes and to initiate games explain women's lower scores.

[35]Note that displayed scores can be reset to zero if they have turned negative, but in our data true scores are observable. We present analysis based on players' true scores, but our conclusions are unchanged if we look at displayed scores instead.

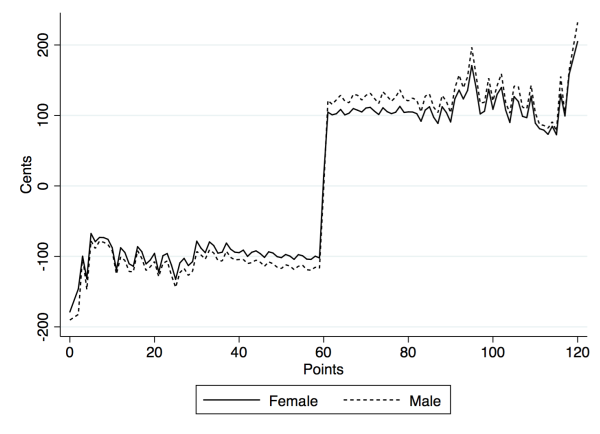


Figure 8: Mean cents earned by points collected, by gender

First, even though female players are as likely to win games as males in each role, due to their lower propensity to initiate games they end up more often either in rounds where no game takes place or in the difficult opponent role (see columns (1) - (3) in Table 4 for baseline winning probabilities in each role). Consequently, even though the winning probabilities are the same for male and female players *in each role*, the *overall* winning rates for women are lower. Second, women profit less from winning games than men as the stakes in these games are lower, in part because women 'knock' less often and also because they disproportionately shy away from initiating the more risky but potentially more lucrative game types *Wenz* and *Solo* (see Table B6 in Appendix B). This idea is illustrated in Figure 8, plotting mean cents earned by points obtained in the game separately for male and female players.[36] The figure conveys a clear message: women gain less from winning, lose less from losing a game than men. The difference is larger in absolute terms in case of winning: as winners, men earn on average 19.5 cents more than women, while in case of a loss women lose on average 14.3 cents less than men.[37]

Note that besides demonstrating a possible explanation for the gender gap in scores, Figure 8 has another important take-away: cents (and consequently scores) earned by female players are less volatile than those of males. This outcome is also captured by comparing the within-person variation in cents and scores between the genders: the within standard deviation in cents is 173.90 for women and 207.13 for men, while the corresponding within standard deviations for scores are 15317.33 for women and 16248.66 for men.

Table 6 summarizes the two channels discussed above by analyzing the determinants of the gender gap in cents per game (over all game types and positions). Column (1) shows that controlling for game- and player-specific characteristics such as position at the table, presence of a female opponent, number of rounds played before and number of opponents who knocked, there is still a significant gender difference in payoffs: women on average earn 4.3 cents less per game than men. Including additional controls for card quality in column (2) hardly affects the estimated gender gap. However, as soon as we account for the focal player's choice to raise the

[36]The discontinuity in cents results from the rules of the game: collecting at least 61 points - in case of the *Sauspiel* game type, together with one's partner - ensures victory. Please refer to Section 2 for more details.

[37]Calculations are available from the authors upon request.

stakes we see a sizable reduction in the estimated difference between male and female players: in column (3) we find that the gender gap drops to 2.8 cents after adding an indicator variable for the focal player's own knocking decision. Finally, if we control for game type (*Sauspiel, Wenz or Solo*) and player role (initiator, partner or opponent) as well as their interactions, the gender gap in cents earned disappears.

Table 6: EXPLAINING EARNINGS PER ROUND

	(1)	(2)	(3)	(4)
Female player	-4.299***	-4.112***	-2.799***	-0.624
	(0.444)	(0.599)	(0.587)	(0.607)
Round-specific controls	✓	✓	✓	✓
Card quality controls		✓	✓	✓
Control: Focal player knocked			✓	✓
Control: Game type # role				✓
Constant	-1.576	14.092***	13.649***	58.422***
	(0.869)	(1.213)	(1.201)	(1.523)
Number of observations	16,655,344	11,936,492	11,936,492	11,936,492
Number of players	15232	15185	15185	15185
Adjusted R^2	0.00509	0.0850	0.0858	0.107

Note: The table displays estimated coefficients from OLS regressions explaining cents earned per round. The following covariates are included in all models: female player, position at table, female opponent, number of opponents who knocked, number of rounds played (log). Additional covariates: (2) card quality measures: own, partner's (when applicable), opponents', Contra received; (3): indicator for focal player's own choice to raise the stakes by knocking, (4): game type and focal player's role in game. Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

4 Discussion

In this section we discuss two aspects of our setting that may affect the internal and external validity of our results: first, players' gender is not directly observable to us; second, there are no monetary incentives for good performance on the platform. In the following we argue that these features do not threaten the generalizability of our findings.

As noted in Section 2.3, in our analysis we infer players' true gender from the gender of the avatar they have chosen. While anecdotal evidence from the *Schafkopf* community suggests that avatar gender accurately indicates players' actual gender, in the absence of direct measures we cannot rule out that some male players select a female avatar or vice versa. If "mismatches" occurred at random, a classic linear regression mismeasurement model would predict our estimates of the gender differences in behavior to be downward biased in magnitude toward zero (Hausman, 2001). However, our estimates would be biased upwards if the most risk loving female players systematically played with a male avatar and the least risk tolerant male players played with a female avatar. In the following, we explain why we find this scenario unlikely.

As mentioned in Sections 2.2 and 2.3, a subsample of players not only selected their avatar's gender but also reported their gender at registration. This information was elicited prior to them choosing their avatar and is not publicly displayed, so there are no incentives to misreport at this stage. For 3,199 players, we are able to link their registration information to their avatar choice and game behavior. Reassuringly, we find that in 97.5 percent of the cases our two measures of gender coincide.[38]

In Table 7 we compare the estimates we obtain for the gender gap in raising the stakes when we use the two different proxies for player gender: avatar gender and gender reported at registration.

Table 7: COMPARING DIFFERENT MEASURES OF GENDER

Gender measure:	Avatar		Reported
	Full sample	Restricted sample	Restricted sample
	(1)	(2)	(3)
Female player (avatar)	-0.070***	-0.078***	
	(0.008)	(0.013)	
Female player (registration)			-0.092***
			(0.013)
Controls	✓	✓	✓
Predicted prob. male players	0.280***	0.291***	0.293***
	(0.002)	(0.004)	(0.004)
Number of observations	16,655,344	7,746,645	7,746,645
Number of players	15232	3199	3199
Pseudo R^2	0.0306	0.0324	0.0336
Log likelihood	-9.431e+06	-4.436e+06	-4.430e+06

Note: The table displays average marginal effects from probit models. The dependent variable is the indicator for initiating a game. In columns 2 and 3, sample restricted to players who provided gender information at registration. The following covariates are included in all models: female player, position at table, female opponent, number of opponents who knocked, number of rounds played (log), card quality measures: own, partner's (when applicable), opponents', Contra received.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

To facilitate comparison with our previous findings, column (1) reiterates the results from our main specification, while column (2) shows the estimated gap, using avatar gender, estimated among the subsample of players who provided gender information upon registration. In the restricted sample, the estimated gender gap in knocking rate is 7.8 percentage points (vs. 7 p.p. in the full sample), suggesting that gender differences in behavior are not restricted to, but are slightly larger, among those who disclose their gender at registration. In column (3), we use reported gender as our explanatory variable, and find a 9.2 percentage points gap in the

[38]Assuming that people report their true gender at registration and the share of mismatches is the same in the full player population as it is among those who disclosed the additional gender information, we can bound the share of mismatches by gender. We find that at most 14.8% of female avatars belong to male players and at most 1.1% of male avatars belong to female players. Comparing the distribution of individual knock rates displayed in Figure 1b we conclude that mismatches are unlikely to fully drive our results. An obvious caveat to this calculation is the possibility that users who plan to choose a gender-incongruent avatar refrain from reporting gender information at registration.

propensity to raise the stakes between male and female players. We find it reassuring that the estimated gap is higher when we use self-reported gender, as it helps alleviate the concern that the strategic choice of avatar gender drives our main findings.[39]

Table B9 in Appendix B reports estimated gender gaps for the most important outcome measures tested in this paper (raising stakes by knocking, initiating a game, winning self-initiated games) using gender reported at registration as the explanatory variable, and shows that our results are all robust to using this alternative gender measure.

The other special feature of our setting we need to address is the lack of monetary incentives: as explained in Section 2.2, cents collected on the online platform have no value outside of the community. Despite the fact that cents do not represent pecuniary incentives, there are different reasons why we believe that our findings are also valid outside our particular context. First, monetary incentives do not necessarily alter subjects' choices over risky gambles: in a major survey on incentive effects by Camerer and Hogarth (1999, p.8) "the overwhelming finding is that increased incentives do not change average behavior substantively although the variance of responses often decreases." In line with this conclusion are the recent results of Falk et al. (2016) who show for a variety of economic behaviors (including risk taking) that the single best predictor for preferences elicited in an incentivized manner is a hypothetical version of the experiment itself, and of Gneezy et al. (2015) who find no difference between estimates for risk and ambiguity parameters elicited through incentivized or hypothetical questions.[40]

Relatedly, we argue that the feature that cumulative scores are publicly displayed provides strong incentives, through affecting players' status in the community. During games, players' scores are prominently shown on the screen (see Figure A1), sending a strong signal of ability to other players, and allowing for immediate comparison with the opponents at the table. Relative performance information (without monetary incentives attached) has been shown in various settings to affect behavior and increase performance (Azmat and Iriberri, 2010; Blanes i Vidal and Nossol, 2011; Gill et al., 2016; Kosfeld and Neckermann, 2011). Strikingly, Delfgaauw et al. (2013) find no additional impact of financial rewards on top of the incentive effect of a tournament that provides feedback on relative performance. Moreover, Gallus (2017) demonstrates that status incentives are also effective in online communities: purely symbolic awards that enhance recognition are "powerful motivators [even if] only the award recipients themselves know that they are the individuals behind the pseudonyms that were awarded" (p.2). Of course, it would be problematic for the interpretation of our findings if men and women differed in their responsiveness to relative performance feedback. Encouragingly, none of the

[39]Relatedly, we argue that in our setting, strategic incentives to select an avatar gender different from one's true gender are low because there are no systematic differences in the way male and female players are treated in the community. Even though players are slightly less likely to knock, initiate a game or initiate the most risky game type (*Solo*) in the presence of a female opponent, the differences are actually negligible in size (see columns (1)-(3) of Table B8 in Appendix B). Moreover, female players are as likely as males to experience "aggressive" treatment by their opponents (i.e. to receive a *Contra*) when they initiate a game, even in case of *Solo*, the most competitive game type (columns (4)-(5) of Table B8).

[40]It is interesting to note that Becchetti et al. (2014) who also study the behavior of card game (in their case, bridge and poker) players in trust games report similar results across incentivized and non-incentivized experiments.

studies listed above find a gender difference in response to information on rank. While gender differences in response to tournament incentives have been extensively documented (Gneezy et al., 2003; Gneezy and Rustichini, 2004; Günther et al., 2010; Price, 2008), we believe our setting is closer to the studies on relative performance feedback than to the literature on gender and tournaments: in the environment we study, even though publicly displayed scores facilitate social comparison, there are no prizes awarded to players with the highest score, and there are no clear winners/losers. Moreover, Dohmen et al. (2011) using fMRI to study activation in the brain region associated with rewards find that the effect of social comparison and the "joy of winning" are very similar for men and women. Combined, these studies suggest that our results are indeed driven by gender differences in risk preferences and not by disparities in the importance attached to status.

Finally, the fact that stakes are hypothetical presents a clear *advantage* in our setting: it eliminates gender differences in wealth as a potential confound for different playing styles. In particular, if men had more disposable income to spend on card playing (either due to the gender pay gap in salaries that they earn outside the platform, or because of different rates of labor market participation), we could not attribute our results to genuine preference differences. Given that there are no real stakes, and thus players do not face an actual budget constraint that can limit their ability to raise the stakes in accordance with their preferences, such concerns play no role.[41]

5 Conclusion

This paper exploits a large dataset from an online card game community to study gender differences in risk preferences in a setting where players voluntarily sort into and interact over several rounds of a game that involves risk-taking in a male-dominated environment. We find that despite the possibility for sorting, learning and adaptation, women tend to take less risk than men, a result that we argue is not explained by playing skills, beliefs or confidence but rather attributable to true gender differences in risk attitudes.

While our context is admittedly special in that we analyze players in an online card game community who interact through their pseudonyms and avatars in the stylized setting of a zero-sum game, it still encompasses many interesting features of certain real-life work environments. Players in our sample are intrinsically motivated, engage in strategic interaction with each other over several rounds, participate in competition as individuals or as team members, and strive to gain status in the community by signaling high performance. As such, our study helps alleviate the concern that the gender gap in risk taking observed in the existing literature results from certain features of the laboratory environment (anonymity, one-shot decisions) or, in case of observational studies on financial risk taking, from gender differences in market knowledge, and thereby contributes to our understanding of gender differences in preferences.

[41] We thank an anonymous reviewer for their helpful comment highlighting this advantage of our setting.

References

Adams, R. B. and Funk, P. (2012). Beyond the Glass Ceiling: Does Gender Matter? *Management Science*, 58(2):219–235.

Al-Ubaydli, O. and List, J. A. (2015). Do Natural Field Experiments Afford Researchers More or Less Control Than Laboratory Experiments? *American Economic Review*, 105(5):462–466.

Andersen, S., Ertac, S., Gneezy, U., List, J. A., and Maximiano, S. (2013). Gender, Competitiveness, and Socialization at a Young Age: Evidence From a Matrilineal and a Patriarchal Society. *Review of Economics and Statistics*, 95(4):1438–1443.

Antonovics, K., Arcidiacono, P., and Walsh, R. (2009). The Effects of Gender Interactions in the Lab and in the Field. *The Review of Economics and Statistics*, 91(1):152–162.

Atkinson, S. M., Baird, S. B., and Frye, M. B. (2003). Do Female Mutual Fund Managers Manage Differently? *Journal of Financial Research*, 26(1):1–18.

Azmat, G. and Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7-8):435–452.

Babcock, L., Recalde, M. P., Vesterlund, L., and Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107(3):714–747.

Baltussen, G., van den Assem, M. J., and van Dolder, D. (2016). Risky Choice in the Limelight. *Review of Economics and Statistics*, 98(2):318–332.

Bartling, B., Brandes, L., and Schunk, D. (2015). Expectations as Reference Points: Field Evidence from Professional Soccer. *Management Science*, 61(11):2646–2661.

Bartus, T. (2005). Estimation of marginal effects using margeff. *Stata Journal*, 5(3):309–329(21).

Becchetti, L., Fiaschetti, M., and Marini, G. (2014). Card games and economic behavior. *Games and Economic Behavior*, 88:112–129.

Beckman, S. R., DeAngelo, G., Smith, W. J., and Wang, N. (2016). Is social choice gender-neutral? Reference dependence and sexual selection in decisions toward risk and inequality. *Journal of Risk and Uncertainty*, 52(3):191–211.

Beckmann, D. and Menkhoff, L. (2008). Will Women Be Women? Analyzing the Gender Difference among Financial Experts. *Kyklos*, 61(3):364–384.

Beetsma, R. M. W. J. and Schotman, P. C. (2001). Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo. *The Economic Journal*, 111(474):821–848.

Bernasek, A. and Shwiff, S. (2001). Gender, Risk, and Retirement. *Journal of Economic Issues*, 35(2):345–356.

Blanes i Vidal, J. and Nossol, M. (2011). Tournaments Without Prizes: Evidence from Personnel Records. *Management Science*, 57(10):1721–1736.

Blau, F. D. and Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3):789–865.

Böheim, R., Freudenthaler, C., and Lackner, M. (2016). Gender Differences in Risk-Taking: Evidence from Professional Basketball. IZA Discussion Papers 10011, Institute for the Study of Labor (IZA).

Booth, A., Cardona-Sosa, L., and Nolen, P. (2014). Gender differences in risk aversion: Do single-sex environments affect their development? *Journal of Economic Behavior & Organization*, 99:126–154.

Borghans, L., Heckman, J. J., Golsteyn, B. H. H., and Meijers, H. (2009). Gender Differences in Risk Aversion and Ambiguity Aversion. *Journal of the European Economic Association*, 7(2-3):649–658.

Brandts, J., Groenert, V., and Rott, C. (2015). The Impact of Advice on Women's and Men's Selection into Competition. *Management Science*, 61(5):1018–1035.

Buser, T. (2016a). How does the Gender Difference in Willingness to Compete evolve with Experience? Tinbergen Institute Discussion Papers 16-017/I, Tinbergen Institute.

Buser, T. (2016b). The Impact of Losing in a Competition on the Willingness to Seek Further Challenges. *Management Science*, 62(12):3439–3449.

Camerer, C. and Lovallo, D. (1999). Overconfidence and Excess Entry: An Experimental Approach. *The American Economic Review*, 89(1):306–318.

Camerer, C. F. and Hogarth, R. M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty*, 19(1):7–42.

Cameron, C. A. and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–372.

Charness, G. and Gneezy, U. (2012). Strong Evidence for Gender Differences in Risk Taking. *Journal of Economic Behavior & Organization*, 83(1):50–58.

Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2):448–74.

Daruvala, D. (2007). Gender, risk and stereotypes. *Journal of Risk and Uncertainty*, 35(3):265–283.

Delfgaauw, J., Dur, R., Sol, J., and Verbeke, W. (2013). Tournament Incentives in the Field: Gender Differences in the Workplace. *Journal of Labor Economics*, 31(2):305 – 326.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schüpp, J., and Wagner, G. G. (2011). Individual risk attitudes: measurement, determinants and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Donkers, B., Melenberg, B., and Van Soest, A. (2001). Estimating Risk Attitudes using Lotteries: A Large Sample Approach. *Journal of Risk and Uncertainty*, 22(2):165–195.

Dwyer, P. D., Gilkeson, J. H., and List, J. A. (2002). Gender differences in revealed risk taking: evidence from mutual fund investors. *Economics Letters*, 76(2):151–158.

Eckel, C. C. and Grossman, P. J. (2008). Differences in the Economic Decisions of Men and Women: Experimental Evidence. In Plott, C. R. and Smith, V. L., editors, *Handbook of Experimental Economics Results*, volume 1, chapter 57, pages 509–519. Elsevier.

Eil, D. and Lien, J. W. (2014). Staying ahead and getting even: Risk attitudes of experienced poker players. *Games and Economic Behavior*, 87:50–69.

Falk, A., Becker, A., Dohmen, T., Huffman, D. B., and Sunde, U. (2016). The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. IZA Discussion Papers 9674, Institute for the Study of Labor (IZA).

Foellmi, R., Legge, S., and Schmid, L. (2016). Do Professionals Get It Right? Limited Attention and Risk-taking Behaviour. *The Economic Journal*, 126(592):724–755.

Gallus, J. (2017). Fostering Public Good Contributions with Symbolic Awards: A Large-Scale Natural Field Experiment at Wikipedia. *Management Science*, 63(12):3999–4015.

Garratt, R. J., Weinberger, C., and Johnson, N. (2013). The State Street Mile: Age And Gender Differences In Competition Aversion In The Field. *Economic Inquiry*, 51(1):806–815.

Gerdes, C. and Graensmark, P. (2010). Strategic behavior across gender: A comparison of female and male expert chess players. *Labour Economics*, 17(5):766–775.

Gertner, R. (1993). Game Shows and Economic Behavior: Risk-Taking on "Card Sharks". *The Quarterly Journal of Economics*, 108(2):507–521.

Gill, D. and Prowse, V. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, 5(2):351–376.

Gill, D., Prowse, V., Kissova, Z., and Lee, J. (2016). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. Economics Series Working Papers 783, University of Oxford, Department of Economics.

Gneezy, U., Imas, A., and List, J. (2015). Estimating Individual Ambiguity Aversion: A Simple Approach. NBER Working Papers 20982, National Bureau of Economic Research, Inc.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance In Competitive Environments: Gender Differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

Gneezy, U. and Rustichini, A. (2004). Gender and Competition at a Young Age. *American Economic Review*, 94(2):377–381.

Greene, W. H. (2008). *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, 6th edition.

Günther, C., Ekinci, N. A., Schwieren, C., and Strobel, M. (2010). Women can't jump? An experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization*, 75(3):395–401.

Hausman, J. (2001). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *The Journal of Economic Perspectives*, 15(4):57–67.

Healy, A. and Pate, J. (2011). Can Teams Help to Close the Gender Competition Gap? *The Economic Journal*, 121(555):1192–1204.

Hvide, H. K. (2002). Pragmatic beliefs and overconfidence. *Journal of Economic Behavior & Organization*, 48(1):15 – 28.

Jianakoplos, N. A. and Bernasek, A. (1998). Are women more risk averse? *Economic Inquiry*, 36(4):620–630.

Johnson, J. E. V. and Powell, P. L. (1994). Decision Making, Risk and Gender: Are Managers Different? *British Journal of Management*, 5(2):123–138.

Kamas, L. and Preston, A. (2012). The importance of being confident: Gender, career choice, and willingness to compete. *Journal of Economic Behavior & Organization*, 83(1):82 – 97. Gender Differences in Risk Aversion and Competition.

Kosfeld, M. and Neckermann, S. (2011). Getting More Work for Nothing? Symbolic Awards and Worker Performance. *American Economic Journal: Microeconomics*, 3(3):86–99.

Linde, J. and Sonnemans, J. (2012). Social comparison and risky choices. *Journal of Risk and Uncertainty*, 44(1):45–72.

Lindquist, G. S. and Saeve-Soederbergh, J. (2011). Girls will be Girls, especially among Boys: Risk-taking in the Daily Double on Jeopardy. *Economics Letters*, 112(2):158–160.

Metrick, A. (1995). A Natural Experiment in "Jeopardy!". *The American Economic Review*, 85(1):240–253.

Moore, D. A. and Healy, P. J. (2008). The Trouble With Overconfidence. *Psychological Review*, 115(2):502.

Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Olsen, R. A. and Cox, C. M. (2001). The Influence of Gender on the Perception and Response to Investment Risk: The Case of Professional Investors. *Journal of Psychology and Financial Markets*, 2(1):29–36.

Pope, D. G. and Schweitzer, M. E. (2011). Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes. *American Economic Review*, 101(1):129–157.

Post, T., van den Assem, M. J., Baltussen, G., and Thaler, R. H. (2008). Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show. *American Economic Review*, 98(1):38–71.

Price, J. (2008). Gender Differences in the Response to Competition. *Industrial and Labor Relations Review*, 61(3):320–333.

Slonim, R., Wang, C., Garbarino, E., and Merrett, D. (2013). Opting-in: Participation bias in economic experiments. *Journal of Economic Behavior & Organization*, 90:43–70.

Smith, G., Levere, M., and Kurtzman, R. (2009). Poker Player Behavior After Big Wins and Big Losses. *Management Science*, 55(9):1547–1555.

Sunden, A. E. and Surette, B. J. (1998). Gender Differences in the Allocation of Assets in Retirement Savings Plans. *The American Economic Review*, 88(2):207–211.

Wozniak, D., Harbaugh, W. T., and Mayr, U. (2014). The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices. *Journal of Labor Economics*, 32(1):161–198.

A Screenshots from the online game

Figure A1: Screenshot from the online game

Figure A2: Table selection in the online card game community

B Additional tables and figures

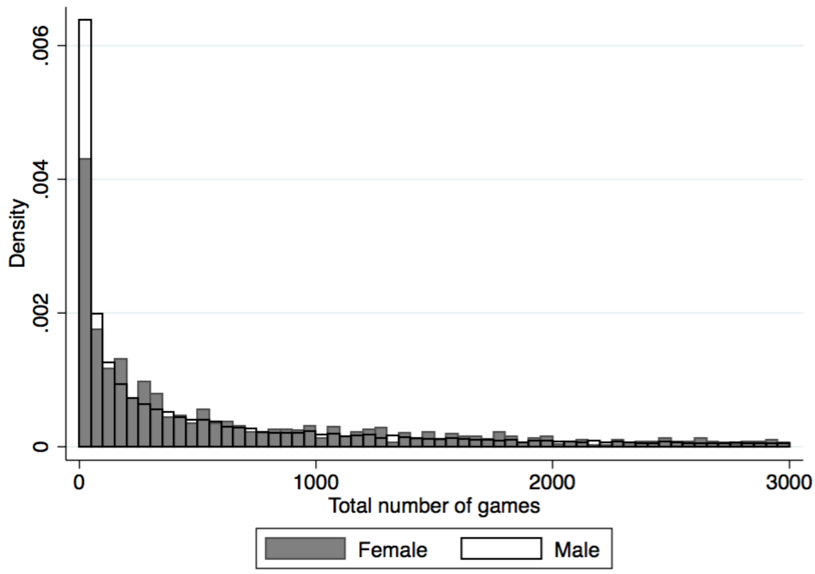


Figure B1: Distribution of total number of games played, by gender (truncated at 3,000)

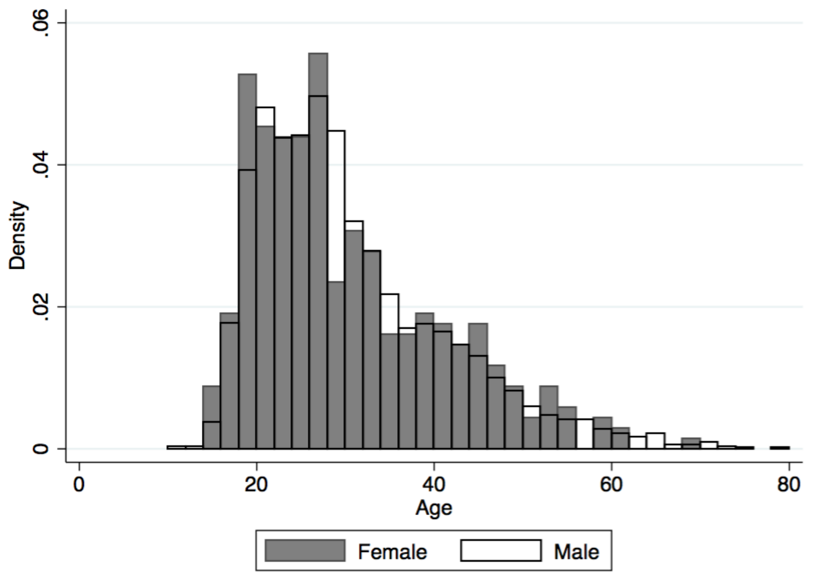


Figure B2: Distribution of age, by gender

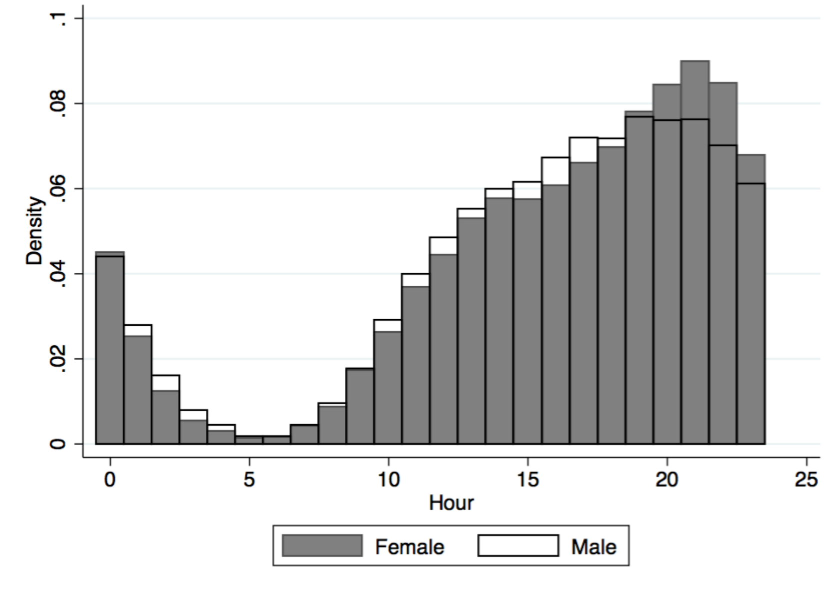


Figure B3: Distribution of the hour of the day when games are played, by gender

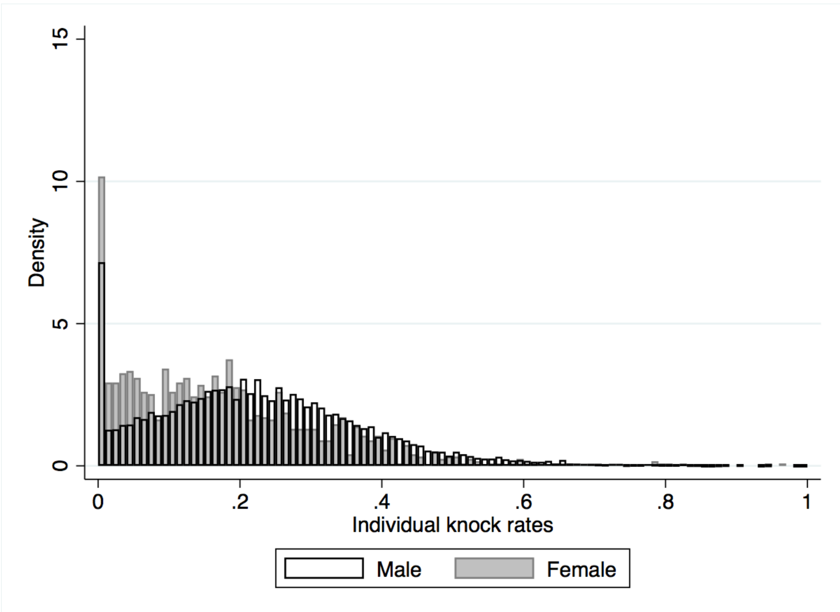


Figure B4: Distribution of individual knock rates, excluding the first 11 games

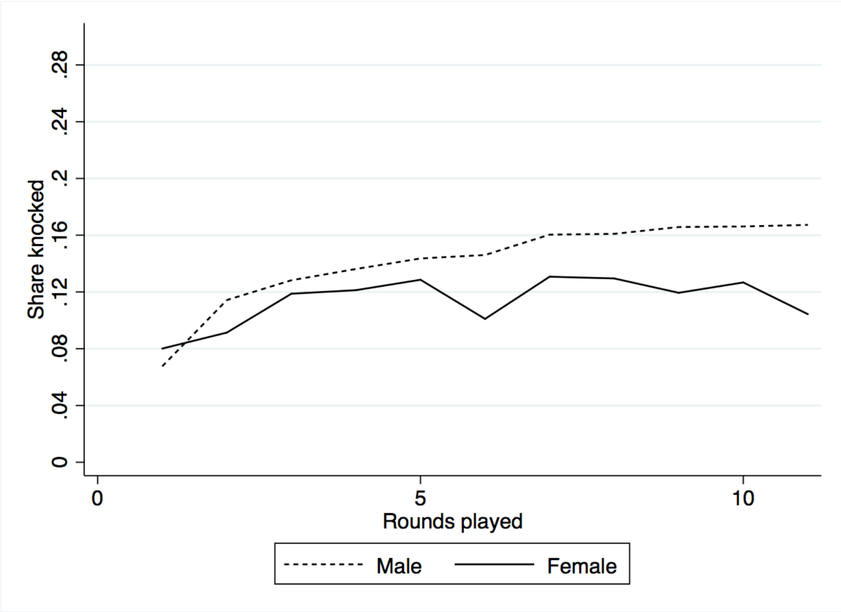


Figure B5: The gender gap in knocking over the first 11 games

Table B1: GENDER COMPOSITION OF ROUNDS

Number of women at table	Number of rounds	Share of rounds, %
0	2,308,689	55.45
1	1,491,212	35.81
2	333,553	7.96
3	31,273	0.75
4	1,110	0.03
Total	4,163,837	100

Table B2: LIKELIHOOD OF RAISING THE STAKES - ESTIMATED COEFFICIENTS

	(1)	(2)	(3)	(4)	(5)
Female player	-0.244***	-0.244***	-0.218***	-0.230***	-0.283***
	(0.030)	(0.030)	(0.028)	(0.029)	(0.065)
Position at table = 2		-0.029***	-0.029***	-0.027***	-0.032***
		(0.001)	(0.001)	(0.002)	(0.003)
Position at table = 3		-0.031***	-0.031***	-0.028***	-0.033***
		(0.002)	(0.002)	(0.002)	(0.003)
Position at table = 4		-0.023***	-0.023***	-0.021***	-0.028***
		(0.001)	(0.001)	(0.001)	(0.003)
Female opponent		-0.005	-0.018***	-0.045***	-0.060***
		(0.003)	(0.003)	(0.003)	(0.006)
Num. rounds played (log)			0.008	0.012**	-0.008
			(0.004)	(0.004)	(0.009)
Rank = 2			-0.084***	-0.073***	-0.071***
			(0.009)	(0.009)	(0.016)
Rank = 3			-0.155***	-0.134***	-0.139***
			(0.012)	(0.013)	(0.025)
Rank = 4			-0.246***	-0.215***	-0.228***
			(0.015)	(0.016)	(0.034)
Num. opponents knocked = 1				-0.347***	-0.348***
				(0.002)	(0.004)
Num. opponents knocked = 2				-0.646***	-0.654***
				(0.003)	(0.007)
Num. opponents knocked = 3				-0.827***	-0.867***
				(0.009)	(0.015)
Age (Q2)					0.022
					(0.031)
Age (Q3)					0.054
					(0.035)
Age (Q4)					0.071
					(0.045)
Location Munich					-0.056
					(0.032)
Constant	-0.579***	-0.557***	-0.495***	-0.277***	-0.126
	(0.007)	(0.007)	(0.033)	(0.034)	(0.073)
Number of observations	16,655,344	16,655,344	16,655,344	16,655,344	5,499,377
Number of players	15232	15232	15232	15232	2271
Pseudo R^2	0.00307	0.00315	0.00748	0.0306	0.0331
Log likelihood	-9.699e+06	-9.698e+06	-9.656e+06	-9.431e+06	-3.129e+06

Note: The table displays estimated coefficients from probit models where the dependent variable is the indicator for raising stakes.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table B3: LIKELIHOOD OF RAISING THE STAKES - SAMPLE RESTRICTION

Sample:	Unrestricted (1)	Restricted (2)
Female player	-0.070***	-0.087***
	(0.008)	(0.018)
Position at table = 2	-0.009***	-0.010***
	(0.000)	(0.001)
Position at table = 3	-0.009***	-0.011***
	(0.001)	(0.001)
Position at table = 4	-0.007***	-0.009***
	(0.000)	(0.001)
Female opponent	-0.014***	-0.019***
	(0.001)	(0.002)
Num. rounds played (log)	0.004**	-0.001
	(0.001)	(0.003)
Rank = 2	-0.024***	-0.025***
	(0.003)	(0.006)
Rank = 3	-0.044***	-0.047***
	(0.004)	(0.009)
Rank = 4	-0.069***	-0.075***
	(0.005)	(0.011)
Num. opponents knocked = 1	-0.119***	-0.120***
	(0.001)	(0.001)
Num. opponents knocked = 2	-0.202***	-0.205***
	(0.001)	(0.002)
Num. opponents knocked = 3	-0.242***	-0.251***
	(0.002)	(0.004)
Predicted prob. male players	0.280***	0.287***
	(0.002)	(0.005)
Number of observations	16,655,344	5,499,377
Number of players	15232	2271
Pseudo R^2	0.0306	0.0327
Log likelihood	-9.431e+06	-3.131e+06

Note: The table displays average marginal effects from probit models where the dependent variable is an indicator for raising stakes. Sample restrictions: (1): none, (2): only players who provided their age and ZIP code at registration are included in the estimation.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table B4: LIKELIHOOD OF RAISING THE STAKES - GENDER INTERACTIONS

	(1)	(2)	(3)	(4)
Female player	-0.230***	-0.241***	-0.230***	-0.236***
	(0.029)	(0.029)	(0.028)	(0.028)
Position at table = 2	-0.027***	-0.028***	-0.027***	-0.027***
	(0.002)	(0.002)	(0.002)	(0.002)
Position at table = 3	-0.028***	-0.030***	-0.028***	-0.028***
	(0.002)	(0.002)	(0.002)	(0.002)
Position at table = 4	-0.021***	-0.023***	-0.021***	-0.021***
	(0.001)	(0.002)	(0.001)	(0.001)
Female opponent	-0.045***	-0.045***	-0.045***	-0.045***
	(0.003)	(0.003)	(0.003)	(0.003)
Num. rounds played (log)	0.012**	0.012**	0.012**	0.012**
	(0.004)	(0.004)	(0.004)	(0.004)
Rank = 2	-0.073***	-0.073***	-0.073***	-0.073***
	(0.009)	(0.009)	(0.009)	(0.009)
Rank = 3	-0.134***	-0.134***	-0.134***	-0.134***
	(0.013)	(0.013)	(0.013)	(0.013)
Rank = 4	-0.215***	-0.215***	-0.215***	-0.215***
	(0.016)	(0.016)	(0.016)	(0.016)
Num. opponents knocked = 1	-0.347***	-0.347***	-0.347***	-0.349***
	(0.002)	(0.002)	(0.002)	(0.002)
Num. opponents knocked = 2	-0.646***	-0.646***	-0.646***	-0.647***
	(0.003)	(0.003)	(0.003)	(0.003)
Num. opponents knocked = 3	-0.827***	-0.827***	-0.827***	-0.831***
	(0.009)	(0.009)	(0.009)	(0.009)
Female player * Position = 2		0.013***		
		(0.004)		
Female player * Position = 3		0.015***		
		(0.004)		
Female player * Position = 4		0.014***		
		(0.003)		
Female player * Female opponent			-0.002	
			(0.011)	
Female player * Opponents knocked = 1				0.008
				(0.007)
Female player * Opponents knocked = 2				0.016
				(0.013)
Female player * Opponents knocked = 3				0.036
				(0.029)
Constant	-0.277***	-0.275***	-0.277***	-0.276***
	(0.034)	(0.034)	(0.034)	(0.034)
Number of observations	16,655,344	16,655,344	16,655,344	16,655,344
Number of players	15232	15232	15232	15232
Pseudo R^2	0.0306	0.0306	0.0306	0.0306
Log likelihood	-9.431e+06	-9.431e+06	-9.431e+06	-9.431e+06

Note: The table displays estimated coefficients from probit models where the dependent variable is the indicator for raising stakes.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table B5: LIKELIHOOD OF INITIATING A GAME

	(1)	(2)	(3)	(4)
Female player	-0.033***	-0.033***	-0.027***	-0.029***
	(0.003)	(0.003)	(0.003)	(0.003)
Table controls		✓	✓	✓
Individual controls			✓	✓
Opponents controls				✓
Predicted prob. male players	0.184***	0.184***	0.183***	0.183***
	(0.001)	(0.001)	(0.001)	(0.001)
Number of observations	16,655,344	16,655,344	16,655,344	16,655,344
Number of players	15232	15232	15232	15232
Pseudo R^2	0.000980	0.00233	0.00331	0.0317
Log likelihood	-7.823e+06	-7.812e+06	-7.804e+06	-7.582e+06

Note: The table displays average marginal effects from probit models where the dependent variable is the indicator for raising stakes. Covariates included in the models: (1): female player, (2): as before, plus position at table and female opponent, (3): as before, plus number of rounds played and rank in displayed score, (4): as before, plus number of opponents who knocked.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table B6: GENDER GAP IN INITIATION BY GAME TYPE

Game type:	Sauspiel (1)	Wenz (2)	Solo (3)
Female player	-0.014***	-0.006***	-0.009***
	(0.002)	(0.001)	(0.001)
Controls	✓	✓	✓
Predicted prob. male players	0.124***	0.023***	0.036***
	(0.001)	(0.000)	(0.000)
Number of observations		16,655,344	
Number of players		15232	
Pseudo R^2		0.0252	
Log likelihood		-1.010e+07	

Note: The table displays average marginal effects from a multinomial probit model. The dependent variable is *game type initiated* with four categories: Sauspiel; Wenz; Solo; none (baseline). Covariates included: female player, position at table, female opponent, number of rounds played, rank in displayed score, number of opponents who knocked.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table B7: LIKELIHOOD OF WINNING GAMES - ESTIMATED COEFFICIENTS

	(1) Initiator	(2) Partner	(3) Opponent	(4) Solo 4^{th}
Female player	0.004	-0.009*	-0.001	0.003
	(0.009)	(0.004)	(0.003)	(0.007)
Position at table = 2	-0.072***	-0.051***	-0.035***	
	(0.002)	(0.003)	(0.002)	
Position at table = 3	-0.010***	-0.036***	-0.006***	
	(0.002)	(0.003)	(0.002)	
Position at table = 4	0.053***	-0.008*	0.021***	
	(0.003)	(0.003)	(0.002)	
Female opponent	-0.014***	0.000	-0.019***	-0.013**
	(0.002)	(0.002)	(0.001)	(0.004)
Num. rounds played (log)	0.054***	0.017***	0.005***	0.002
	(0.002)	(0.001)	(0.001)	(0.002)
Num. opponents knocked = 1	-0.031***	-0.010***	-0.046***	-0.053***
	(0.002)	(0.003)	(0.001)	(0.005)
Num. opponents knocked = 2	-0.057***	-0.090***	-0.038***	-0.019**
	(0.003)	(0.003)	(0.002)	(0.007)
Num. opponents knocked = 3	-0.098***	-0.205***	-0.014**	0.041*
	(0.017)	(0.010)	(0.005)	(0.020)
Card quality	0.529***	0.622***	2.071***	0.219***
	(0.010)	(0.010)	(0.005)	(0.012)
Opponents' card quality	-3.823***	-4.348***	0.339***	-1.846***
	(0.018)	(0.017)	(0.004)	(0.014)
Partner's card quality	0.896***	0.262***	2.693***	
	(0.011)	(0.009)	(0.005)	
Contra	-0.847***	-0.809***	0.851***	1.021***
	(0.004)	(0.004)	(0.003)	(0.009)
Constant	1.319***	2.149***	-2.960***	0.991***
	(0.019)	(0.017)	(0.008)	(0.019)
Observations	2,984,123	2,032,717	6,919,652	427,093
Number of players	14302	14049	14991	12086
Pseudo R^2	0.136	0.153	0.105	0.0865
Log likelihood	-1.403e+06	-864288	-3.438e+06	-239488

Note: The table displays estimated coefficients from probit models.The dependent variable is the indicator for winning a game. Samples restricted to players in the following positions: (1): initiator (2): partner (3): opponent (4): opponent in a Solo game at the fourth position at the table.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table B8: THE RESPONSE TO FEMALE PLAYERS

Outcome:	Knock (1)	Initiate game (2)	Initiate Solo (3)	Receive Contra (4)	Contra to Solo (5)
Female player				-0.002**	-0.003*
				(0.001)	(0.002)
Female opponent	-0.014***	-0.004***	-0.002***		
	(0.001)	(0.000)	(0.000)		
Controls	✓	✓	✓	✓	✓
Predicted prob. male players	0.280***	0.183***	0.036***	0.051***	0.057***
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)
Number of observations	16,655,344	16,655,344	16,655,344	2,984,123	574,852
Number of players	15232	15232	15232	14302	12321
Pseudo R^2	0.0306	0.0317	0.0270	0.0228	0.0196
Log likelihood	-9.431e+06	-7.582e+06	-2.432e+06	-589567	-122940

Note: The table displays average marginal effects from probit models. (Estimates only reported for relevant covariates.) Dependent variables: (1): raise the stakes by knocking; (2): initiate any game type; (3): initiate a *Solo* game, (4): receive Contra from one or more opponents, (5): receive Contra from one or more opponents when initiating a *Solo* game. Covariates included in all models: female player, position at table, female opponent, number of rounds played, rank in displayed score, number of opponents who knocked.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table B9: ALTERNATIVE GENDER MEASURE

Outcome:	Knock (1)	Initiate (2)	Win (3)
Female player (registration)	-0.092***	-0.037***	0.007
	(0.013)	(0.004)	(0.004)
Controls	✓	✓	✓
Number of observations	7,746,645	7,746,645	1,392,814
Number of players	3199	3199	3146
Pseudo R^2	0.0336	0.0331	0.130
Log likelihood	-4.430e+06	-3.529e+06	-646922

Note: The table displays average marginal effects from probit models, using the gender reported at registration as explanatory variable. Dependent variables: (1): raise the stakes by knocking; (2): initiate any game type; (3): win a self-initiated game. Covariates included in all models: female player, position at table, female opponent, number of rounds played, rank in displayed score, number of opponents who knocked.
Standard errors (clustered on player ID) in parentheses, *** p<0.001, ** p<0.01, * p<0.05.