STOCKHOLM SCHOOL OF ECONOMICS

Department of Economics

5350 Master's thesis in economics

Academic year 2014–2015

Discrimination and behavioral impact of gender group composition in a competitive game show setting

Pontus Andersson (21952) Jesper Lundström (40598)

Abstract

Despite the topicality of gender and discrimination issues in the current public debate, research on these issues in a non-experimental setting is conspicuously absent in behavioral economics. Hence, this study attempts to fill this gap by investigating a number of hypotheses regarding 1) discrimination in terms of age, gender, education, and place of residence as well as 2) whether a person's behavior shifts with the gender composition of the group. This two-headed approach makes it possible to draw connections to gender gaps in e.g. the labor market. The analysis is based on data from 320 episodes of the Swedish quiz game show *Vem vet mest*. Results show that academics discriminate against workers. Additionally, the results demonstrate that men behave more competitively in an all-male group and women behave pointedly less competitively in an all-female group. In mixed gender groups, the behavior between men and women do not differ. This is an indication that previous studies might have overlooked the role of gender group composition in explaining competitive behavior.

Keywords: behavioral, game shows, discrimination, gender, group behavior

JEL classification: C93, D03, D81, J16

Supervisor: Magnus Johannessson

Date submitted: May 14, 2015

Date examined: May 27, 2015

Examiner: Karl Wärneryd

Discussants: Maral Shadloo and Cecilia Tollin

Acknowledgements

First and foremost, we want to offer our gratitude to supervisor Professor Magnus Johannesson for providing us with guidance, inspiration and great conversation throughout the entire process. Special thanks are also directed to Kungliga Biblioteket, who kindly and conveniently provided us with the episodes of *Vem vet mest*, which was obviously vital to make this study possible. Lastly, we want to acknowledge the game show host Rickard Olsson and the participants of the show for entertaining us throughout the hours of data collection process. All remaining shortcomings are our own.

Contents

List of Tables

1 Introduction

In recent years, gender and discrimination issues have been increasingly prominent topics in the public as well as the academic debate. The relevance of these topics for researchers in economics is clear. For example, the well-known gender pay gap reduces women's lifetime earnings, their pensions and, directly or indirectly, affects the general life opportunities of all members of society. Of course, this has motivated studies in the past, but these previous studies rarely dig deeper into potential explanations to *why* the different gender gaps exist, merely stating their actuality. However, this particular study aims at shedding light on potential factors behind the observed gender differences, an issue that has puzzled policymakers and academics for a long time. In this paper, we evaluate two potential explanations for gender gaps, namely discrimination and pure behavioral differences between the genders.

The relative absence of studies of this kind in previous literature is understandable as empirical data on these issues without significant noise is hard to find. Hence, the question of underlying factors behind the gender gaps has been more or less left on the table by disciplines having a distinct focus on empirical data, such as labor economics. Further, going through the existing literature on the topic in behavioral economics, the research area in which we would place this paper, it is evident that, also here, the studies are rather monotonous in their choice of method and data source. Even if the literature has brought some attention to these issues, there are few examples using non-laboratory data. Hence, results from this field are arguably often deprived of external validity.

Limited answers from both traditional economics and behavioral economics bring us to use data from a TV game show in our attempt to elucidate the above-mentioned matters. Our choice to conduct a study using non-laboratory data and a more traditional experimental method is well-aligned with recent developments in the field of behavioral economics. In the past decade or so, there has been a clear hike in the number of studies using already available empirical data to test various behavioral hypotheses (e.g. Antonovics, Arcidiacono, and Walsh 2005; Post et al. 2008). In this very spirit, several researchers have been drawn to TV game shows (e.g. List 2006; Anwar 2012). Participants in game shows are often engaging in well-defined high stakes situations, which is very convenient for the purpose of behavioral economics, as the controlled environment is somewhat similar to an experiment. However, an added benefit is that the participants in the study are removed from the obvious experimental setting, which, hopefully, allows the researcher to study behavior more closely resembling that of a real life situation. This, together with a usually broader subject pool, increases the external validity of game show studies.

The Swedish trivia quiz show called *Vem vet mest* (Swedish for "Who knows the most") is almost an ideal show for our purposes. Above all, the show presents us with a very diverse subject pool in terms of several demographic variables with the stated mission to accurately reflect the Swedish population. This is highly beneficial for both external validity and fits the purpose of our study nicely, presenting contestants with the opportunity of discrimination and a higher likelihood of different group compositions.

Further, *Vem vet mest* is fairly intense in how many decisions are made per show and the show has been running for several years. Together, these means that we are presented with an immense set of shows to pull data from. Also, as it is a long-running and widely watched show, the risk of the participants misunderstanding the rules, a concern often voiced in laboratory studies, is minimized. Even though the game show format is international and multifaceted, the field has so far mostly used American shows with various risk-taking behavior as the main area of study. Hence, *Vem vet mest* is an opportunity to also broaden the academic field of game show studies could be broadened both in terms of geography and type of game show apart from being a nice candidate for our specific study.

The part of the show used for our study is the last stage where only three contestants remain. In this stage, the contestants take turn in answering questions. Even though the game flow and scoring system is quite complex, one contestant is generally in charge of the game flow and has the privilege to either keep the question and answer it herself or send it to one of her opponent. Further, a contestant generally receives points when answering questions correctly and when other contestants answer incorrectly. At the end, the contestant with the highest score wins a substantial amount of money. The structure of the game is covered much more in depth in section 3.1.

To sum up, *Vem vet mest* is a lucrative data source in our attempt to study our two potential root causes of gender gaps, namely gender discrimination and gender-related behavioral differences. For the purposes of this study, we will focus on behavioral differences in a competitive setting due to the nature of the game show. General differences in competitiveness between the genders have been studied previously in behavioral economics before, but not in a non-experimental setting, and has been proposed as a potential explanation to gender gaps (e.g. Gneezy and Rustichini 2004; Niederle and Vesterlund 2007). Moreover, we will extend previous literature and also study how the differences in competitiveness vary with the gender composition of the group. Further, we also magnify the scope of our thesis to encompass other types of discrimination as well. Apart from gender discrimination, the paper addresses both extensively studied types of discrimination (e.g. age) and more innovative discrimination concepts (e.g. education). This has led us to the following research questions:

- Do people discriminate with regards to gender, age, place of residence or education in this setting?
- Do people behave differently depending on the gender composition in the group?

The remainder of this paper will be organized as follows: after this introduction, section 2 presents a theoretical background of the topics touched upon in this thesis. Section 3 describes the game show, provides a detailed walk-through of the method employed to carry out the thesis as well as outlines the hypotheses used to test our research questions. Section 4 reports the results, which subsequently are analyzed and discussed in section 5. Section 6 concludes this paper.

2 Theoretical framework

First, we make the case why game show data is a highly viable alternative to experimental data. This section is followed by an overview of results from studies in behavioral economics using this particular data. Also, we account for previous work in relevant areas of study, i.e. discrimination and how group behavior is affected by the gender composition of the group.

2.1 TV game shows as data source

In this subsection, we argue that game show data is advantageous as it combines the high level of external validity in a natural experiment and the high level of internal validity in a laboratory experiment, resulting in a favorable middle course.

In behavioral economics, studies using empirical data are increasingly considered to be of more interest than studies using experimental data. This study, using game show data, would in the context of behavioral economics be considered to be a mix between the two. A TV setting is, of course, not an entirely natural situation and the mere fact that the game is broadcasted on national television might change individuals' behavior. Of course, contestants might be more nervous knowing that thousands of people watch their performance.

However, an important difference from laboratory experiments is that the participants are not aware that their actions are *studied* by an experimenter. Hence, the show behavior is not impacted by common experimental confounding factors. These factors affect behavior directly due to an awareness of being studied or a

desire to comply with what the participant thinks is the goal of the study or the wish of the experimenter. Often, these factors are lumped together with other interferences in the internal validity of an observational or behavioral study and referred to as the Hawthorne effect (Wickström and Bendix 2000; Benson 2001). With using game show data, we avoid the more drastic cases of the Hawthorne effect and the behavior observed is arguably more similar to real world behavior.

Another benefit of using game show data over traditional experimental data is the possibility to broaden the subject sample. Traditionally, experimental economics has, to a large extent, used students as sample. This is considered to be a limitation since the student population is fairly homogeneous and not necessarily a good representation of the total population (Belot, Bhaskar, and van de Ven 2010). In game shows, the sample consists of individuals with more varied characteristics with regards to demographics as well as socioeconomic background.

Yet, there is still some self-selection bias in game show data as people apply to and enter the game show on a voluntary basis. While this could be argued to be a drawback compared to a natural experiment, there is little difference compared to a laboratory experiment where the participants also choose to take part. Nevertheless, we would argue that most natural experiments also focus on a fairly small part of the population. Imagine a natural experiment in a labor market or house purchasing situation. Then the study would only include those in that particular labor market or those purchasing a house in that particular area at the time. We argue that the willingness to participate in a game show is a trait widely spread across the population, giving game show studies a great deal of external validity. Therefore, a game show population should be more representative of the total population than a student population, and in some cases even more representative compared to a population in a natural experiment.

Furthermore, in contrast to natural experiments, the decisions and actions allowed by the participants in a game show are in general much more limited than in a real life setting. Most game shows are structured in such a way that the players have a well-defined set of choices. As a result, the effect desired to study is more isolated in such a setting.

A distinguishing feature of most game shows is that they have relatively high stakes, which gives the contestants large incentives to perform well. Though the prize sum varies, the mere fact that there is a substantial monetary reward up for grabs should make the incentives similar across different prize sums. In an early game show study, Metrick (1995) estimates that the cost of running an equivalent experiment in a laboratory would be over five million dollars, making game shows an attractive proposition for researchers.

Another major feature of a game shows is the degree of anonymity, which is

significantly lower than in most other possible settings. This can in itself have a large impact on individuals' behavior. Belot, Bhaskar, and van de Ven (2010) argue that reputational concerns might be of high importance when making decisions in such a setting. The participants are not only interacting face to face (opposite to many laboratory experiments), but also broadcasted on television, which could been seen as one of the lowest form of anonymity possible since thousands of individuals observe every action. Researchers (e.g. Charness and Gneezy 2008) show that individuals behave according to the economic man model to a lesser extent when the degree of anonymity is lower.

Furthermore, Charness, Rigotti, and Rustichini (2007) find a strong audience effect. In the presence of an audience, individuals increase their aggressiveness in two games, prisoner's dilemma and battle of sexes. Interestingly, while having an audience reduces the efficiency in one game, it increases the efficiency in the other. On the same topic, Sutter, Lindner, and Platsch (2009) find that the degree of co-operation in a prisoner's dilemma increases when the game is observed by an anonymous audience, which is similar to an anonymous television audience. While many studies in behavioral economics prefer anonymity in order to isolate a certain behavior, decisions in real life are rarely completely anonymous. Therefore, experiments with lower degree of anonymity should better mimic the real world. The fact that behavior varies with anonymity could partly be explained by preferences for social reputation. It is well-established in the literature that individuals care about their social image when making economic decisions (e.g. Andreoni and Petrie 2004; Ariely, Bracha, and Meier 2009).

An important note made by Gertner (1993) is that game shows often encourage risk taking and gambling, which should be taken into consideration when analyzing results based on game shows. However, Baltussen, van der Assem, and van Dolder (2014) show that individuals are *more* risk averse in a TV setting (a live audience, a host, cameras, etc.) than in a laboratory environment. This could potentially be explained by the reduced degree of anonymity, which could increase the fear of "doing wrong" and therefore make individuals take fewer risks. Even though we do not examine pure risk aversion in this thesis, we must be aware of the possibility that individuals' behavior might change slightly while on television.

2.2 Results from TV game shows

Due to its high stake nature, large sample size and diverse pool of contestants, game shows are used in order to study a number of economic decisions observed in both laboratory experiments and real life. The most common economic effect studied is risk preferences. Post et al. (2008) examine individual behavior in the

game show *Deal or no deal*. The most prominent finding is that risk aversion is highly affected by the luck in earlier stages of the game. They suggest that this behavior could be explained by reference-dependent choice theories, such as prospect theory.

Other game show studies (e.g. Beetsma and Schotman 2001) also find support for substantial risk aversion. However, Deck, Lee, and Reyes (2008) find a more modest risk aversion effect than normally suggested by the literature. Gertner (1993) studies the TV show *Card Sharks* and finds higher risk aversion than most other studies. However, he also notes that players ignore their overall wealth in the game when making decisions. Players do not seem to base their decisions on expected final wealth, but rather make decisions based on the expected wealth for every different round of the game. This implies that people are significantly short-sighted.

Metrick (1995) studies the quiz game show *Jeopardy*, where the outcome is not only determined by luck, but also by individual performance, which is similar to *Vem vet mest*. While he cannot prove the existence of risk aversion, he suggests that its absence might be explained by overconfidence. If the players have wrong beliefs about their own ability they might make riskier choices than if they were fully aware of their "real" ability. In a similar study based on the quiz show *Who wants to be a millionaire*, where the outcome is also based on individual performance, Hartley, Lanot and Walker (2006) find that female contestants have a slightly higher degree of risk aversion. However, using the reasoning by Metrick (1995), this could be a sign that women overestimate their ability to a lesser extent than men. By having more accurate beliefs about their ability, women may make more accurate decisions.

Beyond the scope of risk aversion, several papers show that players in game shows make suboptimal decisions when there exists a Nash equilibrium. Berk, Hughson, and Vandezande (1996) find that players of the game show *The price is right* systematically use suboptimal strategies. However, they also find a learning effect, as the number of errors are lower towards the end of the show. Healy and Noussair (2004) find similar results when they perform an identical experiment in a laboratory setting. Moreover, Tenorio and Cason (2002) find that suboptimal decisions in *The price is right* are independent of the current monetary stake in the game, but occur more frequently when it is more difficult to observe which decisions that are optimal.

While many games shows do not include strategic interaction between contestants, Levitt (2004) studies data from the TV quiz show *Weakest Link*. In this show, the contestants do not only answer questions, but also vote off their competitors. Due to this strategic interaction component of the game, it is fairly similar to *Vem vet mest* and both shows allows researcher to study behavioral patterns *between* the contestants. In his study, Levitt finds systematic discrimi-

nation against Hispanics and the elderly, even when controlling for performance in the show and background factors such as education level. He also finds that the discrimination towards Hispanics are information-based, i.e. the discrimination arises from beliefs that the group has a lower ability. The discrimination towards the elderly is, on the other hand, shown to be taste-based, i.e. not based on beliefs about the group's ability, but the mere fact that the group is disliked. However, there is no data suggesting any type of discrimination against blacks or women.

Neither do Antonovics, Arcidiacono, and Walsh (2005) find any discrimination against women nor blacks in a very similar study. However, they find that women consistently discriminate against men and evidence in favor of the discrimination being taste-based. Furthermore, their data shows no evidence for any differences in ability between any groups. While this finding might be a bit surprising, other studies, outside a game show setting, find a similar pattern (Dillingham, Ferber, and Hammermesh 1994). However, in a study of the game show *Friend or Foe*, List (2006) finds evidence that men discriminate against women in the sense that men cooperate slightly less when paired together with a woman. Yet, he does not find any support for racial or age discrimination. This is also in line with the findings of Anwar (2012), who studies another game show, *Street Smarts*. She finds that despite the fact that black contestants perform worse than non-blacks, there is no discrimination, even though performance-based discrimination would have been optimal according to the data.

2.3 Discrimination

Unequal outcomes pervade society in many different aspects. There are without doubt several observed differences in outcomes between races, not only in economic outcomes, but also in social outcomes. To mention a few, there are observed differences in wealth, income, prices paid, residential location or credit extended (Arrow 1998). There are also observed differences between the genders, perhaps most notably in the labor market, where it is clear that women earn significantly less on average than men (e.g. Blau and Ferber 1987; Wright and Ermisch 1991; Hensvik 2014). A common suggested explanation for these observed differences is the existence of discrimination. While gender and racial discrimination are arguably the most debated types of discrimination, there are many observed differences in the labor market outcomes between groups based on other characteristics such as age (Duncan and Loretto 2004), disability (Baldwin and Johnson 1994) or class (Locke 1973).

Since Becker (1957) published his seminal work *Economics of Discrimination*, economists have tried to identify and analyze discrimination and the reasons behind discriminatory behavior. Typically, the literature distinguishes between

two types of discrimination, taste-based discrimination (e.g. Becker 1957) and statistical discrimination (e.g. Arrow 1973). While both types might yield the same disadvantageous outcome for the discriminated group, the reasons differ. Taste-based implies that the individual discriminating against a group has "a taste" or preferences against or in favor of a particular group. For example, an individual who has a taste against a group gets disutility if she has to work together with a person from that specific group. Therefore, an individual with this type of preference is willing to sacrifice some income in order to satisfy these preferences (Stiglitz 1973).

On the other hand, statistical discrimination does not involve any preferences, instead the beliefs of the "quality" of a person who belongs to a demographic group is based on the average "quality" of a person belonging to that particular group. Models of statistical discrimination suggest that individuals use all available information in order to maximize utility. Hence, if there is no information regarding a specific individual, a person would assign him or her the average characteristics of the group he or she belongs to. However, even though this type, in contrast to taste-based discrimination, can be efficient for society as a whole, it hurts individuals belonging to a group with statistically lower ability (Schwab 1986). However, this is not a purely moral trade-off between efficiency and egalitarianism, as several studies show that statistical discrimination actually reduces efficiency (e.g. Lundberg and Startz 1983; Schwab 1986). These alternative findings suggest that discouraging discrimination could also be in the interest of rational economic agents.

Since the labor market is a highly important field for an individual and a field where the literature has observed discrimination, it is natural that the lion's share of the literature on the topic of discrimination focuses on the labor market (e.g. Oaxaca and Ransom 1994). However, in the labor market, there is no doubt that there is an unequal relationship between the individual (or firm) discriminating and the individual subject to discrimination. It is the employer that has the conclusive power, both *de facto* and *de jure*, whether an individual should, for example, be employed, receive a pay raise or receive a promotion. While there is a possibility that an individual discriminates against a firm (e.g. based on the characteristics of the firm's employees), it does, in general, not have the same impact on the discriminated firm as it would have on a discriminated individual. Therefore, we argue that by putting too large of a focus on labor market-related discrimination, the literature misses one important notion, namely discrimination among equals. However, this is not simple to study in a pure labor market setting. It is hard to find a field setting where individuals are equal apart from the characteristics we would like to examine. Instead, attempts have been made in the field of behavioral economics, where experiments can be designed to designate experimental subjects with the exact same degree of power, making the subjects equals.

For example, Fershtman and Gneezy (2001) find, in an experiment using undergraduate students at the same university, systematic discrimination against students of "Eastern origin." These result are, according to the authors, due to mistrust towards those of "Eastern origin." In another laboratory experiment by Anderson and Haupert (1999), subjects playing the role of an employer was allowed to choose between "green workers" and "yellow workers". Cards with a certain color and a made-up productivity level were randomly assigned to subjects acting as the employees in the game. While the subjects acting as "employers" could pay a small fee to find out the true productivity of each player, the "employers" still hired fewer of the workers associated with the color known to have the lowest average productivity.

Another behavioral experiment by Tajfel (1970) suggests that the mere division into different groups within the experiment triggers discriminatory behavior, despite the fact that there was no difference between the subjects before the experiment. A similar discriminatory behavior between made-up groups was found by Vaughan, Tajfel and Williams (1981). Here, no group was designed to be superior to the other. Further, List (2004) finds evidence for statistical discrimination against men, non-whites and Californians (compared to non-Californians), using a prisoner's dilemma game where the subjects do not know each other and have identical possibilities to affect the game. This result is especially interesting since it captures residential discrimination, which has not been analyzed much in economic literature otherwise.

In our thesis we have a unique setting where all players are more or less equals in the sense that the same rules apply for everyone and that all players have approximately the same chance of winning at the start of the final stage. Any discrimination among equals is closely related to ingroup-outgroup effects, also referred to as social identity effects. When individuals belong to a group, even though there is no clear hierarchy between groups, they are likely to identify themselves with other individuals of that particular group. Even though standard economic analysis are based on individual actions based on individual incentives, group identity has been shown to play a major role in individuals' decision making (Chen and Li 2009). The economic literature has found strong support proposing that group membership can affect economic outcomes (e.g. Charness, Rigotti, and Rustichini 2007). More specifically, there is strong evidence for ingroup favoritism and outgroup derogation where individuals have preferences for individuals within the own group or preferences against individuals outside the own group. This also has been modeled formally by Akerlof and Kranton (2000).

Hargreaves Heap and Zizzo (2009) find evidence for outgroup discrimination and that the level of trust is lower between individuals from different groups. Similar results are found by Fershtman and Gneezy (2001) and Bernhard, Fehr, and Fischbacher (2006). Furthermore, Chen and Li (2009) show that individuals,

not only display more generosity towards ingroup members, but also make welfare maximizing actions to a larger extent when matched together with ingroup members. Dasgupta (2004) finds ingroup favoritism to a larger extent among people who belong to socially advantageous groups compared to people who belong to disadvantageous groups. This behavior could be a partial explanation for observed discrimination. An individual could be subject to discrimination because they carry certain characteristics, e.g. gender or race, that are classified as "negative." However, an individual subject to discrimination could also simply belong to an outgroup relative to those discriminating.

One of the most distinguishing economic ingroup-outgroup conflicts is the conflict between workers and employers, or between blue collar and white collar workers. While such conflicts could partly be explained by the ingroup-outgroup effects, it is reasonable to assume that there are some differences between blue collar and white collar workers apart from the simple fact that they belong to different groups. White collar work is typically associated with a higher education level, something that could signal or actually imply a higher ability.

In fact, there are some previous evidence demonstrating the existence of discrimination *in favor* of well-educated individuals. Levitt (2004) finds that individuals with a doctoral degree are subject to statistical discrimination in a competitive setting. He shows that competitors (regardless of education level) would rather compete against individuals without such a degree in a knowledge quiz, suggesting that individuals have beliefs that well-educated people have a higher ability. This might be considered rational as a high education is in some sense a proof of someone's ability in a specific field. Even though there is no rock solid proof of the person's ability in other fields, it is still an indication of the person's general ability to acquire knowledge and information. A more surprising finding by Ball and Eckel (1998) shows that individuals seem to discriminate against groups with lower status, even though the status has nothing to do with past performance in itself. By assigning experiment participants a star to wear on their shirt during the experiment, they find that individuals with a star on their shirt (indicating a higher status) are treated better than individuals without a star.

Another highly pertinent type of observed discrimination, even if not debated as often as race or gender discrimination, is treating people differently because of their age. Several studies provide evidence of age discrimination in the labor market, not only against younger but also against elderly people (e.g. Snape and Redman 2003; Riach 2015). Further, Riach and Rich (2010) find strong evidence for discrimination against older applicants in employment processes for "younger-type jobs" such as *graduate* or *waiter*. A similar result is also found both by McGoldrick and Arrowsmith (1993) as well as Singer and Sewell (1989). The latter also who finds that younger people are preferred for low status jobs.

On the other hand, for somewhat higher status jobs, there are some signs suggesting discrimination in favor of the elders. Gibson, Zerbe and Franken (1993) find, in an interview study, that older people are perceived as more stable and experienced. They also find that older people are perceived better overall by other elders, which points to ingroup favoritism. Also for "older-type jobs" there is discrimination in favor of elders (Perry, Kulik, and Bourhis 1996). Furthermore, Arvey et al. (1987) find that older interview candidates are perceived as better than younger interview candidates. While the literature has established this problem in the labor market and within organizations, age discrimination has not yet been examined in a pure behavioral economics setting.

2.4 Behavioral impact of gender group composition

As discussed in the previous section, gender differences exist in several domains of society such as consumption, investments, and labor market outcomes. In order to effectively shape policies to narrow the gender gap, a consensus of which factors are the main drivers of the gender differences must be reached. However, without deeper knowledge about the underlying reasons for the observed differences, policy changes might not yield the desired outcome. Therefore, it is in the interest of everyone to uncover the driving forces behind the observed outcomes. The literature has traditionally tended to focus on discrimination when trying to explain the gender gap in the labor market (e.g. Aigner and Cain 1977; Altonji and Blank 1999). In the last decade, however, there has been an increased literature trying to find other explanations. One of the most commonly suggested factors is the potential of various psychological and behavioral differences between men and women. Differences in risk-aversion, self-confidence or response to competition have all been suggested as important factors to explaining the gender gaps (Bertrand 2011).

The differences in attitude to competition between the genders have been widely studied. Many laboratory experiments (e.g. Gneezy, Niederle, and Rustichini 2003) show that men are much more competitive than women and that men respond better to competitions in terms of improving their performance. In a paper by Niederle and Vesterlund (2007), men and women are allowed to perform a task. After completion, they find out their absolute score and choose if they want to enter a competition stage or not. Despite no difference in performance, twice as many men choose to enter the competitive tournament stage. Similar results are found in field studies as well. Manning and Saidi (2010) find that men in the labor market are slightly more likely to have contracts that include performance-based pay. On the same topic, Buser, Niederle, and Oosterbeek (2014) show that teenage boys are more likely to choose more prestigious study profiles, even when controlling for earlier academic ability. These results suggest that men are more overconfident about their performance. Prince (1993) show

that the difference in overconfidence is largest in stereotypically "masculine" tasks, such as financial decisions.

However, several studies suggest that the difference cannot solely be explained by overconfidence. Price (2008) shows that men actually respond better to competition in the sense that men actually improve their results in a competitive setting. In his experiment, men increased their performance with 10 percent when put in a competitive situation while women did not change their performance. Similarly, Gneezy, Niederle, and Rustichini (2003) find that women perform worse when competing in mixed-gender tournaments. However, in single-gender tournaments there is no difference between the genders, suggesting that women only respond worse to competition when competing against men. These results are in line with those of Antonovich, Arcidiacono, and Walsh (2003). They show that men perform better than women in a high stake game show and men perform relatively better when they compete against women instead of other men. They also find the performance of women does not seem affected by the gender of their opponents. This seems true also for young boys and girls, as Gneezy and Rustichini (2004) find that boys in fourth grade outperform girls in fourth grade when competing together, even though there is no difference between the boys and girls when competing alone.

Although there are some support for the view that women respond relatively worse to competition, there is no clear consensus on this topic. Dreber, von Essen, and Ranehill (2011) perform a very similar study, but with Swedish 7-10 year old children where they do not find any gender difference in performance. Niederle and Vesterlund (2007) also fail to confirm the notion that women respond worse to competition.

The findings bearing witness of gender differences in performance in competitive setting are often explained by differences in risk preferences. It is well-established, both in psychology (e.g. Byrnes, Miller, and Schafer 1999) and economic literature (e.g. Powell and Ansic 1997; Croson and Gneezy 2009) that men are less risk averse than women. Therefore, men might see risks as challenges rather than threats and, hence, become more motivated by a situation where risk is involved, such as a competition. Similarly, if women see risks as unpleasant threats, they might instead respond to risky situations by avoidance, something that could explain differences in competitive outcomes (Arch 1993). Similar to the negative response to competition, the risk aversion among women seems to be stronger when they compete against men, and women are also less likely to make risky decisions when competing against men (Lindquist and Säve-Söderbergh 2011). Genetic differences between men and women could be a partial explanation to these findings, however Gneezy, Leonard and List (2009) show that women respond relatively better to competition in Indian matrilineal societies, suggesting that cultural factors play an important part in why we observe differences in competitiveness. Findings in this area have a signif-

icant impact in many aspects of society, but it is perhaps most applicable in the field of education policy. A common argument for single-gender schools is that individuals, especially girls, have greater chance to perform well in a single-gender environment. Since research show that boys are more competitive than girls, a girls-only school should therefore be a less competitive environment, and as a result, more suitable for girls. This is based on the assumption that girls dislike competition *per se*, regardless of the context. Therefore, a single-gender environment, where everyone has the same attitude towards competitive behavior, would be beneficial for those who dislike competition and thereby a more efficient learning environment can be created. However, there are claims (e.g. by Ceraulo 1999) that girls do not dislike competition *per se*. Girls simply dislike competing against boys. Hence, single-gender schools creates an environment where girls potentially respond well to competition, which also could create a more efficient learning environment (Gneezy, Niederle, and Rustichini 2003).

Even though the debate regarding single-gender schools is not very relevant in Sweden, the purpose with this research (i.e. if men and women behave differently in single-gender environments compared to mixed-gender environments) is still highly relevant. If policy makers want to change the unequal outcomes in the labor market, it is vital to have an understanding of the causes of these differences to increase the possibility of shaping effective policies aiming to lessen the gender gap.

While it is well-established that men and women have different preferences for risk and competition in mixed sex-groups, there is still no clear consensus on how the gender group composition affects this behavior, which is one of the primary motivations for this study. Several papers on the effect of single-gender education find that single-gender schools deliver better education results, especially for girls (e.g. Lee and Bryk 1986). Furthermore, Lee and Marks (1990) demonstrate that women who went to single-gender schools tend to overcome social barriers in their professional career to a larger extent. However, other studies have found no significant differences in the performance and outcomes between individuals who went to single-gender schools or mixed-gender schools (e.g. Harker 2000). There are also researchers claiming that single-gender schools simply increase negative gender stereotyping and do not increase neither boys' nor girls' performance (e.g. Halpern et al. 2011).

As shown above, there is no consensus in this matter in studies using schools as the arena of study. This is also the case for studies using other arenas. Datta Gupta, Poulsen, and Villeval (2013) find, in a laboratory experiment, that men chose to compete against other men to a larger extent than they choose to compete against women. This suggests that, not only do men respond better to competition than women, but also do they respond to it even better in a single-gender setting. Price (2008) studies a program that intended to increase educational performance. He concludes that the largest gains were made by men

in departments with the highest proportion of females. The same result, but to a less extent, is true for women. The women that increase their performance the most were women in department with a high proportion of females, indicating that female perform better in a single-gender setting.

Moving beyond differences in competitiveness and performance, Ortmann and Tichy (1999) show women to be more cooperative in a mixed-gender setting. A similar result is found by Nowell and Tinkler (1994), who use a laboratory public goods game to show that all-female groups are more cooperative and more willing to share than all-male and mixed-gender groups.

Yet, most of these studies suggesting differences in behavior are based on laboratory experiments, with all their drawbacks and pitfalls, see section 2.1 for a further discussion. This thesis addresses these theories, widely supported by experimental research, but in more real life-like setting. While *Vem Vet Mest* is still an artificial setting, it provides a starting point from which we can further examine whether these laboratory results hold up in a less artificial setting. The pressure to perform well on a TV-show is arguably more similar to the pressure to perform well in a real workplace or school environment compared to an anonymous laboratory experiment.

3 Method and hypotheses

This section starts off by a description of the game in more detail, including the different phases of the game and an example on how scoring and game flow works in practice. Thereafter, we present our regression equations together with our hypotheses. In the third and final section, we present how the data gathering process was conducted.

3.1 Description of Vem vet mest

As stated, the target for our analysis is the Swedish game show *Vem vet mest*. The show is a quiz show televised every weekday on Swedish national television, based on the British long-running format *Fifteen to One*. As of the spring of 2015, the show has been running for over 7 years and is on its 15th season. Each Monday-Thursday, a number of new contestants take up the challenge and for each show, a select few advance to the weekly finals on Friday where real prize money is up for grabs. This process is then repeated for several number of weeks. During the first four years (8 seasons) of the show, each show contained 12 contestants with 3 from each day advancing to the weekly final. However, from season 9 and onwards, the format was changed slightly to accommodate a

new set-up of only 8 contestants per day and with only 2 progressing from each weekday show.

Nevertheless, the show has always consisted of the same three stages and the common theme through all the stages is that contestants are faced with general knowledge questions with answers given orally and with little time for consideration. Please see List A23 in the appendix for a few example questions. These are the three stages as they are currently constructed with the 8 person set-up:

1. Elimination stage:

   Each of the contestants begin the quiz with three "light bulbs." Each contestant is asked a question in order and given approximately three seconds to answer. If the contestant is wrong, he/she loses one of the three light bulbs. A question is posed first to contestant number 1, followed by a question to contestant number 2, etc. This process is repeated three times, making each contestant answer three questions. If you fail three times, you do not advance to the next stage.

2. Nomination stage:

   At this point, all surviving contestants receive 1 extra light bulb, making it so everyone has 2-4 light bulbs. Now one player is randomized to start this stage and receives a question. If this contestant is wrong, he/she loses one light bulb and a new player is randomized and the new contestant receives a question. If he/she answers correctly, this contestant qualifies as the "nominator", which means calling out the next player to face a question. If the chosen nominee gives the wrong answer, the nominee loses a light bulb and the privilege of nominating stays with the previous nominator. If the chosen nominee is correct, he/she becomes the new nominator. If you lose all your light bulbs, you are immediately removed from the game. This nomination process is repeated until there are only three contestants still standing.

3. Final stage:

   Here, the remaining light bulbs of the three finalists are converted to points with 1 light bulb equals 10 points, which serves to give players who performed better in the previous stages of the game a small advantage. This stage has two types of questions – opened or closed – and before each question, the contestants are given the topic of the question – e.g. "Christmas gifts." However, most of these topics are rather vague and cover a very broad range of possible questions. Again, please see List A23 in appendix for a number of example questions.

   Open questions are open to all players to answer on the buzzer. If the contestant first to press the buzzer is wrong, the opponents both receive

10 points each and we have a new open question. If the player is correct, we move on to a closed question dedicated to that particular player. If a question is closed, the owner of the question is given a vague topic and can either 1) choose to keep the question and answer it him- or herself or 2) send the question away to one of the two other contestants.

If the contestant owning the closed question chooses to retain the question and gives the correct answer, he/she receives 10 points and the next question will be a closed question to that particular contestant again. If the contestant retains the question and is wrong, the opponents receive 10 points each and the following question will be open.

If the contestant owning the closed question chooses to send the question to an opponent and the opponent is wrong, the sender will receive 10 points and the next question will once again be a closed question to the sender. If the contestant owning the closed question chooses to send the question to an opponent and the opponent is right, the receiver receives 10 points as well as the right to get the next closed question.

The process of the final stage is continued until the end of the show (27-28 minutes). The length of the third stage therefore depends on the length of the second stage. Hence, the number of questions posed in the final stage varies. In the current 8 player format, the contestants with the two highest scores at the end of the show advance to the Friday final. In the previous 12 player format, all three players in the final stage advance to the Friday final. Nevertheless, in both formats the contestant with the highest score also gets the privilege of starting the "Friday final" with one extra light bulb. For the purposes of our study, only the final stage is relevant.

Please find Table 1 below, which is a constructed example illustrating both how points are earned and when the question are open or closed during this final stage. The table also contains examples on when the choice of receiver is classified as correct or incorrect, which is highly critical for the hypotheses on discrimination, which are to be presented later in this paper. In both of these dimensions the example should be complete despite the abnormal brevity of this final stage.

Table 1: Example of the game flow in the final stage of Vem vet mest

Nature of question	Action	Right or wrong answer	Incorrectly or correctly sent	Score		
				P1	P2	P3
	Start of the final stage!			10	10	20
Open	P1 pushes the button	Wrong		10	20	30
Open	No-one pushes the button			10	20	30
Open	P2 pushes the button	Right		10	30	30
Closed	P2 sends the question to P1	Wrong	Corr. (10p vs 30p)	10	40	30
Closed	P2 sends the question to P3	Right	Incorr., type 1 (30p vs 10p)	10	40	40
Closed	P3 keeps the question	Right		10	40	50
Closed	P3 keeps the question	Wrong		20	50	50
Open	P1 pushes the button	Right		30	50	50
Closed	P1 sends the question to P2	Right	Incorr., type 2 (50p vs 50p)	30	60	50
	End of game!					

Our hypothesis testing is based only on actions during the final stage for several reasons. First, in the final round, players have two different choices: to send the question to another player or to keep the question to him/herself. In the earlier rounds there are no such decisions, every player has to keep the question in the first stage and send the question away in the second stage. Second, there are only two opponents available as potential receivers in the final stage, opposite to the second stage where there are several opponents to choose from. So if a player chooses to send, he/she faces a binary option which is easier to analyze compared to a situation where the player has up to eleven different choices.

As suboptimal strategies exist (explained below) in the final stage of the game, we can observe suboptimal actions to examine whether there are specific characteristics of the players involved that systematically result in suboptimal decisions. This is based on the assumption that players only care about winning, i.e. they are indifferent between second and third place as they payoff for second and third place is exactly the same. While there is no single optimal strategy at every decision during the final stage of the show, there exists a suboptimal strategy at every decision. A rational contestant, aiming to win the contest, would never send a question to the competitor with highest points.

To illustrate this, we describe three different scenarios with player 1, player 2, and player 3. If player 1 chooses to keep the question, there are two possible outcomes. In the first outcome, player 1 answers the question correctly and receives ten points while player 2 and player 3 receive zero points. Player 1 also keeps the initiative for next round and gets to choose whether to keep or send

the subsequent question. In the second outcome, player 1 answers the question incorrectly and receives zero points while both player 2 and player 3 receive ten points each. Player 1 then loses the initiative and the subsequent question will be an "open" question that the player that first pushes his/her is let to answer.

However, if player 1 chooses to send the question, there are two possible players to send the question to and for each player that receives the questions there are two possible mirror outcomes:

First, assume that player 1 sends the question to player 2. The first possible outcome is that player 2 answers the question correctly and receives ten points while player 1 and player 3 receive zero points each. Player 2 then gets the initiative and chooses whether to send or keep the subsequent question. The second possible outcome is that player 2 answers the question incorrectly and receives zero points, while player 1 receives ten points and player 3 receives zero points. Here, player 1 keeps the initiative and can, again, choose whether to keep or send the subsequent question. Regardless of the answer of player 2, player 3 is static (i.e. did neither send nor receive a question) and cannot receive any points. Second, the mirrored case is exactly the same as above, expect that player 1 sends the question to player 3 rather than player 2.

Since a player can receive points only when 1) he or she answers a question or 2) when the active player keeps the question and give an incorrect answer, a player not receiving a sent question cannot receive any points. Therefore, players have a clear possibility to "isolate" one opponent from having the opportunity to receive any points by sending the question to the other opponent. As a result, sending a question to the opponent with the lowest score can neither increase the distance to first place (if the active player is in second or third place), nor reduce the distance to the closest opponent (if the active player is in first place). Yet, sending a question to the opponent with the higher score is associated with the risk that the player answers correctly, and therefore increases the distance (if the active player is in second or third place) or decreases the distance (if the active player is in first place).

However, there is a theoretical situation where sending a question to the competitor with the higher score could be optimal. If the player with the highest score has a low enough observed probability of answering correct, it could be rational to send the question to that player, and still have a higher expected value than if the question were sent to the opponent with the lower score. However, this is merely a theoretical situation. It is fairly intuitive to see why this should not be commonly observed. If a player holds the first place, it is very unlikely that he or she has a low frequency of correct answers.

Therefore, a suboptimal action (i.e. sending a question to the highest ranked opponent) must be based on the *belief* that a competitor must have a lower

ability than observed. Hereafter, we will often refer to this action as "incorrectly" sending the question.

Furthermore, no suboptimal strategy exists in cases where the two opponents have identical scores, as there is no opponent that is strictly more attractive to send a question to. Therefore, the decision of sending a question to one of the opponents must either be random or based on a belief that the chosen opponent have a lower ability. As a result, we have defined two types of incorrectly sent questions. The first type is situations where the sender sends a question to an opponent that has a higher score than the other opponent. This is, as explained above, clearly suboptimal. The second type is situations where the sender sends a question to an opponent that have identical points as the other opponent. While this is not suboptimal, if individuals with certain characteristics systematically receive questions when they have equally as many points as the other opponent with other characteristics, we argue that this as well is a sign of an underestimation of the ability of those individuals. In the main analysis we combine these two types of incorrectly sent questions, but we also run the same regressions with both types separately. [1]

It is important to distinguish between questions sent and questions incorrectly sent. While sending a question to the opponent with the highest score is always suboptimal (hence defined as incorrectly sent), sending a question is in itself not necessarily suboptimal. The decision between keeping the question and sending it to the opponent with the lowest score is simply a function of subjective variables such as risk aversion, self-confidence or own ability. This is taken advantage of when formulating the hypotheses on how this decision to send or keep will vary due to gender group composition in this thesis.

As described above, when a player keeps a question and answer it incorrectly, both opponents receive 10 points. However, if the player sends the question, only the receiver of the question can receive 10 points. As a result, keeping a question is associated with more risk since both opponents can receive points instead of just one of the opponents. Therefore, a very risk averse player might want to send the question even though he or she believes that he or she has a good chance of knowing the answer. Similarly, a player knowing that his or her true ability is very low and there is a low probability of knowing the answer to the question, it might be perfectly rational to send the question to an opponent. If so, the player does not have to answer the question and potentially give away 10 points to both opponents. Therefore, we cannot state any optimal strategy regarding the decision to send or keep the question, which in any case is not immediately

[1] In the main analysis we use a composite variable called *incorr_sent* which consists of observations with both types of incorrectly sent questions. In appendix tables A1-A8 we present results for the same regressions, but with the two types separated. The incorrect choices when opponents' scores differ are named *inc_sent_type1* and those when scores are equal are named *inc_sent_type2*.

relevant to our research questions regarding gender group behavior.

Further, the only real information about the competitors' ability is the frequency of correct answers given by the competitors. Therefore, the only way to update the beliefs regarding the competitor's ability is by observing this frequency. Yet, players might take other factors such as gender, age or education level into account when updating the beliefs about the competitor's ability. But since these factors in themselves do not give any information about the individual's ability, we claim that the process of "updating beliefs" based on these demographic factors is simply statistical discrimination. This is taken advantage of when formulating the hypotheses on discrimination.

Another important and separate note is that the players do not face situations that directly affect their monetary payoff, instead they face several situations that affect the probability of receiving a fixed monetary payoff (i.e. the 10.000 SEK prize). To illustrate, if player 1 receives 10 points, player 1 increases his or her probability of receiving the monetary payoff. Similarly, if the opponents receive 10 points, player 1's probability of receiving the monetary payoff decreases. The magnitude of the effect 10 points has on the change in probability is of course different for different junctures in the game. The relative score between the opponents and the time left (i.e. number of questions left) will affect how much a gain of 10 points will increase the probability of receiving the monetary payoff.

Since the players' decisions do not directly affect their monetary payoff, it is possible to claim that the effect affecting individuals' behavior here is ambiguity aversion rather than risk aversion (i.e. preferences for known risk over unknown risk, see e.g. Fox and Tversky (1995) for a detailed description). However, we argue that an optimal strategy if a player wants to win the game and receive the prize payoff is to maximize expected points in every single decision. Therefore, every decision could be seen as a separate event where the contestants only need to care about how to maximize the expected payoff (in terms of points) in the current event. As a result, the aversion to the risk of losing points in the game is to be considered as regular risk aversion. This theoretical reasoning is also supported by empirical data from game shows. As Gertner (1993) show, contestants generally do not seem to make their decisions based on how the immediate decisions affect the final outcome. Instead, he argues, decisions are made to maximize the expected outcome of each separate event of the game.

3.2 Hypotheses: discrimination

We argue that, in the third stage of *Vem vet mest*, there is only statistical discrimination and no taste-based discrimination. Sending a question to an op-

ponent does not create disutility for the opponent, it is merely a belief that the chosen opponent do not have the ability to know the answer to the question. After all, by sending the question to a specific player, that player is at least given a chance of receiving points. Hence, if someone carries a preference against a certain group, sending questions to a person of that group would not be a wise idea. One objection to this line of reasoning is that a person with a hard-lined preference against another group would want to publicly humiliate a person of this group by giving them questions that they would answer incorrectly. However, we argue that this gamble would not be appealing to a person using taste-based discrimination, as the potential recipient is more likely to answer the question correctly than incorrectly. Hence, sending questions to people you do not like in order to humiliate them is quite a bad prospect.

Furthermore, as a fairly large sum of money is on the line, it is reasonable to believe that the contestants try to maximize their own probability of winning the game. If this assumption does not hold, it is possible to imagine a situation where taste-based discrimination *in favor* of another group cannot be easily separated from statistical discrimination. If contestants have very strong preferences for contestshaants with certain characteristics and prefer such a contestant to win rather than to win themselves, it would, hypothetically, be rational to send questions to that opponent to accommodate a situation where this preferred opponent gets a chance to receive points. However, we assume that the financial (and social) incentives to win are large enough to trigger a behavior motivated by the desire to gain as many points as possible for yourself. Therefore, we argue that we can isolate for statistical discrimination.

In this section we address the matter of statistical discrimination in the show. We test for four different types of discrimination, namely *gender discrimination*, *age discrimination*, *education discrimination*, and *residential discrimination*.[2] While there are substantial previous research regarding the two former types, there are still a large gap in the literature on the latter types (see section 2.3 for a detailed discussion).

For every hypothesis we will use the same dependent variable, i.e. whether the question was correctly or incorrectly sent. On the other hand, we will vary the independent variable depending on which hypothesis we aim to test. Please find the general structure for the equations used to test our hypotheses in this section below:

[2]Our initial plans did also include testing for race discrimination, but as the amount of contestants with another racial background than Swedish advancing to the final was very small, the sample size was way too small for us to perform any meaningful analysis. However, this lack of non-Swedes in the final stage could very well be due to discrimination in the second stage (the Nomination stage) rather than non-Swedes just being worse contestants.

$$y = \gamma X + \beta x \tag{1}$$

where X is a set of control variables, including age, gender, place of residence, and education for the sender, the receiver, and the static player respectively. Please note that when testing for e.g. gender discrimination, all control variables for gender are naturally dropped to avoid collinearity. Therefore, the set of control variables will differ slightly between the different categories of discrimination. The control variables are defined below.

We define *gender* based on whether the constant is a man or a woman.

We define *education level* based on whether the individual has a post-upper secondary school education (eftergymnasial utbildning in Swedish).

We define *place of residence* as follows:

Large = Stockholm, Gothenburg or Malmö (including suburbs)
Medium = Cities/Towns with a population between 20,000 and 200, 000
Small = Towns/Villages/Countryside with a population less than 20,000

We define *age* as follows:

Old = Retired
Middle = 30-65 years old
Young = under 30 years old

Please see section 3.4 for a discussion regarding the collection of this data. The control variables are in the form of dummy variables. When gender and education ought to be controlled for, it will suffice to only include one dummy variable for each of the three players. For example, if the sender of a question is a male and the recipient is female, the gender dummy for the sender assumes the value of 1 and the gender dummy for the receiver takes the value of 0. Further, if the sender is an academic and the static player is a worker, then the education dummy variable for the sender assumes the value of 1, while the same variable for the static players assumes the value of 0.

When age and place of residence are used as controls, the process is slightly different as these categories have three outcomes rather than two. We have set a baseline so that all players are old and live in a large city. For each of the sender, receiver and the static player we then have included two dummy variables for both *age* (young and middle-aged) and place of residence (middle- and small-sized towns/villages). Together with the baselines, these four variables cover all possibilities for these two types of variables. If e.g. the sender is middle-aged and lives in a small town, then the sender dummy variables for middle-age and small-sized town would assume the value of 1. On the other hand, if the receiver

is old and lives in a large city, all the receiver dummies for age and place of residence would assume the value of 0 in the regression, as both old and large city are the baseline values.

The reason for controlling for these characteristics in the various hypotheses is to avoid all sorts of biases that arise. If we were only to control for the characteristics of the sender and the receiver, we could still see bias enter from a characteristic that the static player has. Imagine a scenario where a male sends the question incorrectly to a female and a male is static, which would indicate male discrimination against female. However, this wrongful choice to send the question to the female receiver could also stem from that both the sender and the receiver are workers and the static is an academic. Hence, any detected discrimination could also be a discrimination against the worker. If we were to exclude the control variables for any of the three persons involved or for any of the other characteristics, we would fail to separate between potentially interfering types of discrimination.

The dependent variable y in all the different categories throughout this section regarding discrimination is given by:

$$y = incorr\_sent = \begin{cases} 0 & = & \text{correct sent} \\ 1 & = & \text{incorrect sent} \end{cases}$$

In the rest of this section, we will go through each of the four areas of discrimination.

Gender discrimination

Hypothesis 1.1

We hypothesize that females discriminate against men. This would manifest itself in that female senders tend to make erroneous sending decisions, i.e. sending to the "wrong" contestant more often to men than to females. The rate of incorrectly sent questions should be higher among the questions sent to men than to women, suggesting that women believe men to be less able to give the correct answer, taking previous performance in the game into account.

A positive result would be in line with the findings by Dillingham, Ferber, and Hamermesh (1994) and Antonovics, Arcidiacono, and Walsh (2005). Furthermore, a positive result would also be in line with the well-supported ingroup-outgroup effect, suggesting that the individuals in the group *women* would have lower beliefs about the ability of the individuals in the outgroup *men*, (see e.g. Chen and Li 2009). The base equation (1) would for this hypothesis be complemented by the following definition of the independent variable x:

$$x = fem\_mal\_disc = \begin{cases} 0 & = & \text{female sender, male receiver, female static} \\ 1 & = & \text{female sender, female receiver, male static} \end{cases}$$

Here, we pick out those observations where 1) a female is the sender, 2) her opponents are one male and one female and 3) she chooses to send the question to one of her opponents. Those occurrences where she chooses to send it to the male is given the value of 0. Consequently, the remaining observations are when she chooses to send it to the other female in the game and are given the value of 1.

As we hypothesize that the amount of incorrect sending decisions correlate positively with the decision to send the question to a male, we would expect that:

$$\beta < 0$$

in the equation (1) above.

Hypothesis 1.2

Our second hypothesis is that men discriminate against women. Similarly, if our hypothesis is correct, men would send incorrect questions to women more often than they send incorrect questions to other men. This would imply that men have beliefs that men have a higher ability than women. This type of discrimination is well documented in the literature (see section 2.3) and if men incorrectly send questions to women to a larger extent, it would indicate that women are subject to statistical discrimination by men. Here, the base equation (1) would be complemented by the following definition of the independent variable x:

$$x = mal\_fem\_disc = \begin{cases} 0 & = & \text{male sender, male receiver, female static} \\ 1 & = & \text{male sender, female receiver, male static} \end{cases}$$

For the testing of this hypothesis we single out the observations where 1) a man is the sender, 2) his opponents are one male and one female and 3) he chooses to send the question to one of his opponents. Those occurrences where he chooses to send it to the female is given the value of 1. Consequently, the remaining observations are when he chooses to send it to the other male in the game and those are given the value of 0.

As we hypothesize that the amount of incorrect sent decisions correlate positively with the decision to send the question to a female, we would expect that:

$$\beta > 0$$

in equation (1) above.

Hypothesis 1.3

Finally, we believe that the discrimination effect is stronger among men, i.e. men discriminate against women to a higher degree than women discriminate against men. While there are empirical support for discrimination against both genders, the observed gender gap in the labor market is without doubt in the favor of males (Hensvik, 2014). Naturally, if discrimination is a factor explaining this gap, the discrimination against women should be greater. Therefore, our hypothesis is that the effect is stronger in hypothesis 1.2 than in hypothesis 1.1.

$$|\beta_{1.1}| < |\beta_{1.2}|$$

To test hypothesis 1.3, we will use a method to test the equality of two coefficients from two different regressions and samples proposed in Clogg, Petkova, and Haritou (1995) and Paternoster et al. (1998).

In the general case, this method calculates the Z-statistic as follows:

$$Z = \frac{\beta_1 - \beta_2}{\sqrt{(se\beta_1)^2 + (se\beta_2)^2}}$$

where seβ_n is the standard error of β_n. This Z-statistic is then measured against a standard normal probabilities table.

Age discrimination

When testing for age discrimination, we use, as with gender discrimination, equation (1). However, we define the independent x variable differently for the different hypotheses.

Hypothesis 2.1

Our first age-related hypothesis is that young and middle-aged individuals overestimate the ability of older and "wiser" individuals. This overestimation would imply that relatively younger people think that older people have a higher ability than younger people. In terms of behavior in the game, we would see, if the

hypothesis holds, that younger and middle-aged contestants send incorrectly to older contestants less often than to contestants classified as young or middle-aged. This notion is supported in previous literature (Arvey et al. 1987; Gibson, Zerke, and Franken 1993). To test this, we again use our base equation (1), but we define our independent variable x as follows:

$$x = youmid\_old\_disc = \begin{cases} 0 & = & \text{young/middle sender, old receiver, young/middle static} \\ 1 & = & \text{young/middle sender, young/middle receiver, old static} \end{cases}$$

The observations used in this specific testing are those when the 1) the sender is classified as either young or middle-aged, 2) exactly one of the other contestants are old, 3) the other contestant is either young or middle-aged and 4) the person with the initiative chooses to send the question. Those observations where the sender chooses to send the questions to the old opponent are given the value of 0, while the remaining questions sent to the contestant classified as either middle-aged or young are given the value of 1.

As we hypothesize that the amount of questions that are incorrectly sent would correlate negatively with the age of the chosen receiver, we would expect that:

$$\beta > 0$$

in the regression, i.e. equation(1), used for this hypothesis.

Hypothesis 2.2

We also hypothesize that relatively older contestants discriminate against younger opponents. In a similar fashion, a positive result would mean that older and middle-aged senders would send more incorrect questions to younger recipients more often than they send these questions to other contestants that are middle-aged or old. Such a behavior would imply that older contestants underestimate the ability of younger contestants, which would be in line with the theory of in-group favoritism as well as evidence put forward by Gibson, Zerbe, and Franken (1993). To test this, we again use our base equation (1), but we define our independent variable x as follows:

$$x = midold\_you\_disc = \begin{cases} 0 & = & \text{old/middle sender, old/middle receiver, young static} \\ 1 & = & \text{old/middle sender, young receiver, old/middle static} \end{cases}$$

The observations that we use in this specific testing are those when the 1) the sender is classified as either old or middle-aged, 2) exactly one of the other con-

testants is young, 3) the other contestant is either old or middle-aged and 4) the person with the initiative chooses to send the question away. Those observations where the sender chooses to send the questions to the young opponent are given the value of 1, while the remaining questions sent to the contestant classified as either middle-aged or old are given the value of 0.

As we hypothesize that the amount of questions that are incorrectly sent would correlate negatively with the age of the chosen recipient, we would expect that:

$$\beta > 0$$

in the regression, i.e. equation (1), we use for this hypothesis.

Education discrimination

Our third part of the discrimination section is regarding discrimination based on education attainment. We hypothesize that everyone, both academics and workers, underestimate the ability of workers. Previous studies have find some support for such discrimination, but the evidence is very limited and circumstantial in nature (see section 2.3).

Hypothesis 3.1

As stated in the introduction, we hypothesize that contestants with higher education will discriminate against people with lower education. This is reasonable as education can very well be seen as a symbol a person's ability, which also has been empirically supported by Levitt (2004). If this hypothesis holds, we would see that academics have a higher rate of incorrect questions when they choose to send away questions to workers than when they choose to send to other academics. This would also be in line with ingroup favoritism. To test this, the x variable in our base equation (1) is defined as follows:

$$x = aca\_work\_disc = \begin{cases} 0 & = & \text{academic sender, academic receiver, worker static} \\ 1 & = & \text{academic sender, worker receiver, academic static} \end{cases}$$

In order to isolate for this exact behavior, we have again isolated the relevant observations in the total set of questions. In this case, we will use the observations where 1) the sender is classified as an academic, 2) exactly one of the other contestants is an academic too, 3) the other contestant is a worker and 4) the academic nominator chooses to send the question away. Those questions that the sender chooses to send the questions to the academic are given the value of 0, while the remaining questions sent to the worker are given the value of 1.

As we hypothesize that the amount of questions that are incorrectly sent would correlate positively with the decision to send the question to the worker, then we would expect that:

$$\beta > 0$$

in the regression, i.e. equation (1), we use for this hypothesis.

Hypothesis 3.2

As stated in the introduction, the very limited empirical studies in this area suggest that also workers discriminate against other workers (or in favor of academics) (Levitt, 2004). Hence, we would expect that workers will underestimate the ability of other workers. This would lead to that workers will send more incorrect questions to workers relative to how much they send these questions to academics. Our base equation (1) would be modified as follows in order to accommodate this hypothesis:

$$x = work\_work\_disc = \begin{cases} 0 & = & \text{worker sender, academic receiver, worker static} \\ 1 & = & \text{worker sender, worker receiver, academic static} \end{cases}$$

In this case, we will use the observations where 1) the sender is classified as a worker, 2) exactly one of the other contestants is a worker too, 3) the other contestant is an academic and 4) the person carrying the initiative chooses to send the question away. Those questions that the sender chooses to send to the academic are given the value of 0, while the remaining questions, those sent to the worker, are given the value of 1.

As we hypothesize that the amount of questions that are incorrectly sent would correlate positively with the decision to send the question to the worker, then we would expect that:

$$\beta > 0$$

in the regression equation (1).

Hypothesis 3.3

Finally, we believe that the discrimination effect is stronger among academics, i.e. academics discriminate against workers to a higher degree than workers discriminate against other workers. First, the discrimination against workers and discrimination against individuals with lower status is more established in

the literature (e.g. Ball and Ecker 1998). Second, if education is seen as proxy for knowledge or ability, it would be rational to discriminate against non-educated than against educated since this would be associated with a higher probability of receiving points, as shown in Anwar (2012). Therefore, our hypothesis is that the effect is stronger in hypothesis 3.1 than in hypothesis 3.2. Formally, this is given by:

$$|\beta_{3.1}| > |\beta_{3.2}|$$

Just as with the similar hypothesis 1.3 regarding relative strength of gender discrimination, the method suggested by Clogg, Petkova, and Haritou (1995) and Paternoster et al. (1998) is used to evaluate hypothesis 3.3. For further information regarding this method, please see hypothesis 1.3.

Residential discrimination

As the final part of the discrimination section, we test for the innovative concept of discrimination on a residential basis, i.e. whether individuals discriminate against people from certain areas, such as major cities, medium cities/towns, and small towns/countryside.

Hypothesis 4.1

Our first hypothesis related to residential discrimination is that people living in relatively urban environment discriminate against people living in rural areas. Hammarström (2004) states the population in urban areas are more highly educated, hence it is possible that people generalize this notion to all individuals living in large cities and vice versa. Also, a discriminatory effect in this direction would be in line with ingroup favoritism. In this case, we have defined this to be that people living in large or mid-size cities will underestimate people living in rural areas or in very small cities. This would manifest itself in the game in that urban people will send more incorrect questions to contestants from the countryside compared to other contestants from urban areas. Our base equation (1) would be modified as follows in this case:

$$x = midlarg\_sma\_disc = \begin{cases} 0 & = & \text{large/middle sender, large/middle receiver, small static} \\ 1 & = & \text{large/middle sender, small receiver, large/middle static} \end{cases}$$

To test this hypothesis, we only use the observations where 1) the sender lives in a large or medium-sized city, 2) exactly one of the other contestants also lives in a large or medium-sized city, 3) the other contestant lives on the countryside and 4) the person with the initiative chooses to send the question away. Those questions that the sender chooses to send to the other contestant living in a

relatively more urban environment are given the value of 0, while the remaining questions sent to person from the countryside are given the value of 1.

As we hypothesize that the amount of questions that are incorrectly sent would correlate positively with the decision to send the question to the rural contestant, we would expect that:

$$\beta > 0$$

in the regression for this hypothesis, i.e. equation(1).

Hypothesis 4.2

Furthermore, we also believe that individuals from smaller or mid-size cities will underestimate individuals from the three largest cities in Sweden, i.e. those categorized as living in large cities. This line of reasoning stems from an underdog, ingroup behavior, favorizing other people from rural areas. People from more rural environment may carry a chip on their shoulder and think that people from the largest cities are not that skilled and have no knowledge of the "real world", resulting in an underestimation of these individuals. Another, less hostile, interpretation of this effect could be that contestants simply value their own type of knowledge and experience higher and deem it more useful. As the "others" obviously lack this specific type of knowledge that the contestant overvalue, then the contestant will underestimate the "other" by overvaluing subjects in which they excel themselves. If this is true, it could also be applied in hypothesis 4.1. Nevertheless, these beliefs will result in that relatively more rural people will send more incorrect questions to people from the largest cities compared to the rate of incorrectly sent questions to other people from the countryside or medium-sized cities. Here, we modify our base equation (1) as follows:

$$x = smamid\_larg\_disc = \begin{cases} 0 & = & \text{small/middle sender, small/middle receiver, large static} \\ 1 & = & \text{small/middle sender, large receiver, small/middle static} \end{cases}$$

In this case, we will focus only on the observations where 1) the sender lives in a small or medium-sized city, 2) exactly one of the other contestants also lives in a mall medium-sized city, 3) the other contestant lives in one of the three major cities in Sweden and 4) the person that currently has the initiative chooses to send the question away. Those questions that the sender chooses to send the questions to the other contestant living in a relatively more rural environment are given the value of 0, while the remaining questions sent to person from the major cities are given the value of 1.

As we hypothesize that the amount of questions that are incorrectly sent would correlate positively with the decision to send the question to the more urban contestant, then we would expect that:

$$\beta > 0$$

in our regression, i.e. equation (1).

Hypothesis 4.3

While our hypotheses are that both groups underestimate the ability of "the other" group, we believe that the effect is stronger among individuals living in large cities. This is because the underdog mentality of those in smaller areas towards people from the more urban areas may be diffused by some kind of subordination or minor self-contempt. However, this is only our own reasoning and have no particular basis in previous literature. Formally, this hypothesis is given by:

$$|\beta_{4.1}| > |\beta_{4.2}|$$

As previously, the method used to test hypothesis 4.3 is the one suggested by Clogg, Petkova, and Haritou (1995) and Paternoster et al. (1998). For further information regarding this method, please see hypothesis 1.3.

3.3 Hypotheses: behavioral impact of gender group composition

In this section, we test whether individuals' behavior is dependent on gender group composition. Literature suggest that men are more competitive than women and that men tend to be more overconfident than women (e.g. Gneezy, Niederle, and Rustichini 2003; Antonovics, Arcidiacono, and Walsh 2003). However, we employ a different angle with our hypotheses. Instead of testing the pure level of risk preferences or overconfidence, we test whether men or women choose to keep the questions to a larger extent when the group composition have certain characteristics. A decision to do so would indicate that a certain group composition would lead to an increased tolerance/preference for risk and affect the contestants' overconfidence. This study therefore branches from and broaden the current literature, even though it is inspired by some previous studies. For example, Nowell and Tinkler (1994) suggest that females are more cooperative when surrounded by other females, which could be extrapolated to imply a less competitive behavior. Also, Lindquist and Säve-Söderbergh (2011) argue that women are less likely to make risky decisions when placed in a group with men.

This could also be used analogously in our study as keeping the question is to be considered the more risky choice, as argued in section 3.1.

To test our hypotheses for this section we state the following equations, one for male behavior and one for female behavior:

$$y_1 = \gamma X + \beta_1 x_1 + \beta_2 x_2 \tag{2}$$

$$y_2 = \gamma X + \beta_3 x_3 + \beta_4 x_4 \tag{3}$$

where X is a set of control variables explained below.

The dependent variable in both of these equations is the observed outcome for each observation, i.e. whether the question was sent to another opponent or kept by the player with the initiative. This means that questions that we define as "open" those where the opportunity to answer is given to the contestant pushing the button the fastest, are excluded. We will only count observations where a decision to keep or send the question is made. As the literature suggest that men and women respond differently to competitive situations, we will separate male and female behavior into two different regressions to see whether the behaviors shift depending on group composition (e.g. Gneezy, Niederle, and Rustichini 2003). Therefore, we run one regression for observations where the player taking an action is a man, equation (2) and one where that player is a woman, equation (3). As referred to above, the different dependent variables are given by:

$$y_1 = kept\_male = \begin{cases} 0 = \text{send } | male \\ 1 = \text{kept } | male \end{cases}$$

$$y_2 = kept\_female = \begin{cases} 0 = \text{send } | female \\ 1 = \text{kept } | female \end{cases}$$

As can be seen above, they are built the same way, but they separate decisions made by males and females respectively into two different groups of observations. The independent variables represent different possible gender compositions of the groups. These are presented below:

$$x_1 = male, male, male$$

$$x_2 = male, female, female$$

$$x_3 = female, female, female$$

$$x_4 = female, male, male$$

Please note that the first two variables are used in equation (2) where we study male behavior, while the two latter are used in equation (3) where we study female behavior. Hence, the first position in the enumeration of genders is used to denote the gender of the decision maker (male for x_1 and x_2 and female for x_3 and x_4). The other two are not defined by any position, but just indicates a situation where both of the two contestants are of the specified genders. These two groups of two are then compared to a baseline in each regression. For both equations, the baseline is one where the two opponents are of different sex, i.e. for the first regression (2) (males) the baseline is *male, male, female* and for the second regression (3) (females) the baseline is *female, female, male*.

In order to control for characteristics other than gender, we include a set of control variables representing all possible combinations for all the other characteristics, i.e. (*education level*, *age* and *place of residence*). By doing this, it is possible to exclude the possibility that the differences in behavior between the different groupings in terms of gender is driven by the group composition with respect to the other characteristics, e.g. education level. It is possible, for example, that academics might behave differently in an all-academics group compared to a mixed group.

The control variables consist of dummy variables that represent various compositions of different characteristics and take a value of 1 when such a composition occurs in the game. For example, the first dummy variable on age is *young-young-middle*. This represent one of the 18 possible age compositions of the three finalists and should be interpreted as a situation where the player making a decision is young (the first value of the three age types) and compete against one young and one middle-aged player. Remember, the order of the two opponents is not relevant. The second dummy variable on age is *young-young-old* and represent another possible age combination where the player making the decision is young and compete against one young and one old player. In settings where we observe this composition, this dummy variable takes a value of 1, in all other cases, this dummy variable equals 0. We have included dummy variables for all possible age compositions except for one arbitrarily chosen baseline which is *young,-young-young*, representing situations where all three contestants are young. In such cases, all age-related dummy variables take the value of 0.

We have created dummy variables in the same way for education level and place of residence. As a result, we will always have three relevant types of dummy variables - *age, education*, and *place of residence*. Note that education level only have two possible values since it is defined as *academics/worker* compared to age or place of residence which are defined as *young/middle/old* and *large/middle/small* respectively. Therefore, there are of course fewer possible combinations for education. For both *age* and *residence*, we have 18 different combinations (17 dummy variables), but for *education* there is only 6 combinations, resulting in only 5 dummy variables.

We have decided to not control for anything else than various group compositions. One possible control variable left out intentionally is the relative score in the game at the time. It is possible that certain game situations might trigger the contestants to either send or keep the question to a larger extent. For example, a player with a very comfortable lead might be more content to either gamble and take all questions or the player might be more willing to play it safe and send everything away to minimize the scoring potential for his or her opponents. However, we consider it highly unlikely that any potential standings in the game, e.g. a close contest or a landslide victory, should correlate with the gender composition of the group or any other composition of any other demographic variable. Hence, an exclusion of a measure describing the game situation in terms of points would only imply a loss of explanatory power, but would imply no bias. Therefore, it has, without a significant loss, been omitted in favor of a cleaner analysis as well as easing the econometric analysis on part of the authors.

When we test for shifts in behavior due to various gender group compositions, we cannot use the same control variables as when we test for discrimination. In this case, we are interested in the *group composition*, not whether there are any differences between the receiver and the static player. For example, a situation where the player making the decision is old, the first opponent is young and the second opponent is old is in our regressions identical to a situation where the player taking the action is old, the first opponent is old and the second opponent is young. In other words, *old-young-old* is identical to *old-old-young*. Since we are interested in the group composition's effect on peoples' behavior, we do not have to distinguish between the characteristics of the receiver and the static player nor are we interested in whether a question is sent incorrectly or not.

Now when we have presented the structure of and motivated these regressions, we will move on to our two hypotheses in this area:

Hypothesis 5.1

Our first hypothesis is that men keep the question to a larger extent when the group consists only of men compared to the baseline case where the opponents are of mixed gender. This is in line with much of the research (e.g. Datta Gupta, Poulsen, and Villeval 2013), stating that men become more competitive and thereby more confident in their own ability in a single-gender environment.

If our hypothesis is correct, we would expect the following:

$$\beta_1 > 0$$

in the regression above, i.e. equation (2).

Hypothesis 5.2

Similarly, we hypothesize that men keep the question to a larger extent when both the other opponents are women compared to the baseline when the opponents are of two different genders. The line of thinking is that men stereotypically want to prove their manliness and worth. This effect is hypothesized to be even stronger when a man is the only male of the group. Hence, a man would take more questions when facing two females than when facing one man and one female.

As a result, we would expect:

$$\beta_2 > 0$$

in the equation (2) above.

Hypothesis 5.3

Now, moving on the two hypotheses looking at female behavior. We start with the hypothesis that women keep the question to a lesser extent when the group consists only of women. This is compared to the baseline group where the female sender faces one male and one female opponent. We base this on the limited previous research on this topic, expecting females to become less competitive and less confident in an all-female setting (Nowell and Tinkler 1994; Lindquist and Säve-Söderbergh 2011). On the other hand, Harker (2000) argues that we should not see any difference in female behavior depending on the gender group composition.

If this is correct, we would expect the following:

$$\beta_3 < 0$$

in equation (3) stated above.

Hypothesis 5.4

Our last hypothesis regarding female behavior and compares the situation where female contestants who carry the initiative face two male opponents compared to the baseline of facing one male and one female opponent. Here, we believe that the female contestant keeps the question to a lesser extent when both the other opponents are men compared to the baseline. Of course, making these hypotheses assume that the taste for competition is highest when facing only one male and two females. This loose reasoning comes from that in a primal setting, stereotypical females are expected to fight for the attention of the sole male.

If this hypothesis proves to be correct, we would see the following:

$$\beta_4 < 0$$

in equation (3) above.

3.4 Data collection method

The data for this thesis is collected by the authors by simply watching the show. In total, we collected data from 320 shows, which gives us roughly 7,600 questions (observations) in the final stage and 20,000 questions (observations) in the first two stages. Primarily, we have used data from the Friday final shows since the incentives in those differ slightly from the incentives in the qualification shows (Monday – Thursday) as they are constructed in the current format. In the final round of the qualification shows, two out of three contestants reach the Friday final, and hence has a chance of receiving the 10,000 SEK prize. Despite the fact that the winner in the qualification show receives an extra light bulbs for the Friday final, which gives a slight advantage, there is a considerably larger difference between third and second place than between second and first place. Therefore, a contestant has large incentives to finish top two while they do not have such as large incentives to be the number one. However, in the Friday final, there is only one winner of the 10,000 SEK prize and a third place is just as good as a second (or any other) place. Therefore, the incentives are slightly different since the contestants in the Friday final only cares about winning in the current format.

However, in seasons 1-8, aired between 2008 and 2011, all the three finalists in the qualification rounds are qualified for the Friday final. Still, the winner gets an extra light bulb in the Friday final, and hence there is no difference between second and third place. Hence, contestant only cares about winning and are indifferent to second and third place. As a result, the relative incentives in the qualification shows in seasons 1-8 are the same as the Friday final shows for all seasons. But of course, the difference between the rewards are not that great in the weekday programs. Still, we have complemented the data on the Friday finals with roughly 50 weekday shows from season 4, which was arbitrarily chosen without any particular reasoning.

We treat each question as a single data point and collect full data for every single question. As a result, each data point consists of information regarding who the sender of the question is, who the receiver of the question is, the current standing in terms of points for all players, the current fraction of correct answers for each player. We also record whether the question was correctly answered,

wrongly answered or not answered at all. Furthermore, the demographic characteristics for every contestant is also recorded. These include *gender*, *age*, *place of residence* and *education level*.

While some characteristics, such as *gender* and place of residence, are very straightforward to observe, others, such as age and *education level*, are not always as obvious. Since there is no strict structure of how the contestants are presented in each show, there is a slight variation of what information is received for each contestant. Potential problems with the characteristics are discussed below. Lists of all variables used in the analysis can be found in the appendix in tables A17-A19 and basic descriptive statistics can be found in tables A20 and A21.

Place of residence is, as mentioned above, very straightforward. We can observe the place of residence for every contestant and if any doubt about the size of the city/town, the population number is readily available. Hence, we do not have any issues with classification errors due to subjective judgment regarding this variable.

Gender is also very easy to define. A potential issue in terms of gender would be if a trans-gender person would compete in a final. We have observed trans-gender persons compete in the show, but not advancing to the final. Hence, the classification in terms of gender is equally unproblematic as the classification of place of residence.

The exact *ages* are not presented. Instead, we have to use our own judgment based on the appearance and job status of the contestants. In most cases, this is not very difficult. Retired contestants are often presented as pensioners and students are often presented as students, which implies that they are young. There is of course a risk that a contestant that is, for example, a student and looks very young is older than 30. However, since the other contestants do not necessarily know the true age, they will also make a judgment based on the appearance and information given in the show. So even if a contestant's true age is different from what he or she appears, he or she is likely to be "treated" as a person of the age that he or she appears.

Furthermore, *education level* can also at some points be somewhat problematic. Since it is not clearly stated in the show whether the participants have a completed university degree or not, we have to make a judgement and evaluate the participants' professions (which is almost always revealed) and then base the definition on whether the profession generally requires a university degree or not. We are aware of the potential pitfalls here. First, a person might never have attended university, but still has a profession that normally does require a university degree. Second, a person with an academic degree might have a profession that typically does not require such a degree. However, since the actual education status is not officially presented for the TV-audience, the "true"

education status does not provide any information for any of the participants. Instead, it is simply the "impression" of someone's education status that can be taken into account, just as with the age of the contestants.

Also, before the final round, the three finalists are given a more detailed presentation and get a chance to talk to the TV-host about their personal life. Yet, there is no standard presentation, but more of an ad-hoc question session by the TV-host and hence there is still a chance that the contestants do not reveal any information about their jobs or educations. However, this will in most cases reveal more information about which type of job the contestants have and in some cases even if they went to university or not. After this presentation, it is very uncommon that we have any doubt about the contestants' education level.

Still, we had 83 cases where we were not entirely sure of a contestant's age and 65 where we were unsure of the education level of a contestant. This compares to a total data set including roughly 1,000 contestants. As a further robustness test, we modified the data set using the opposite value of initial classification. As an example, if we were unsure about whether a contestant was young or middle-aged and classified the person as young in our original data set, the person was classified as middle-age in the modified data set. [3]

Yet, a potential problem is that the contestants might have more information about each other than the TV-audience has. Even though no other information is "officially" presented, the contestants have moments to talk to each other before the show. During this time, the contestants can spend time together and potentially share information with each other that we cannot account for in this study. There is a chance that information is revealed that might affect the opinion of another contestant. This could for example be marital status, interests, earlier experiences, previous jobs or previous places of residence.

However, this interaction is more likely between some contestants than other. Since the weekly qualification round as well as the Friday final is recorded on the same day, there are a number of contestants in the studio at the same time and therefore it is unlikely that every contestant have time to share too much information with every other contestant. Furthermore, the producer of the show, Erik Hammar (2015), states that the contestants have a rather strict schedule and also show up at slightly different times. First, there is a mid-morning session with the 16 contestants (or 24 in earlier seasons) in the first two qualifications show, those broadcasted on Monday and Tuesday. Before the show starts, the production team provides information, description of the rules

[3]The hypotheses were then re-tested using this modified data set and there were no significant changes in our results, indicating that the impact of potential classification errors are low. All results were statistically significant at the same level and the changes in effect were minor. The results using the modified data set are shown in the appendix in tables A10, A15, and A16.

and general preparation for the TV show. During this time, the contestants can of course small-talk with each other, but this time is quite packed with practical information.

When these first two shows have finished recording, the contestants in the third and fourth qualification round, those broadcasted on Wednesday and Thursday, arrive to the recording site. The contestants in these latter groups receive the same information as the first two groups, while the contestants from the first two rounds have waiting time until the final. After the fourth round, the final starts. As a result, the contestants in round 3 and 4 have rather limited time to have longer conversations and hence communicate any information. Instead, it is mainly the contestants from round 1 and 2 who can communicate to a larger extent while they wait for the final.

4 Results

Before presenting the results, we want to clarify how we made the selection on which results to present in this section and which to leave in the appendix. In this text section, results for each hypothesis will be presented with two different regressions, one with and one without control variables. However, the data for the control variables themselves will not be presented here, but can be found in the various appendices.

All regressions use clustered standard errors by each unique episode. These are introduced to check for a potentially inflated sample, as several unique observations (questions) in the data set come from the same episode. A key assumption when calculating a non-clustered standard error is that each unique observation carries the same amount of information. In this case, this assumption is problematic. It is very reasonable to assume that decisions made actually affect decisions made later in the same show to a larger extent than a previous decision will affect a decision made in a completely different episode. This means that there might be some auto-correlation between observations and, hence, two decisions made in the same show do not contribute to the explanatory power to the same extent that two random decisions would do. However, both regressions with and without controls, but without any clustering, can be found in the appendix. A side benefit of doing a cluster analysis is that it introduces robust standard errors automatically. Naturally, robust standard errors are implemented also in the non-clustered regressions in the appendix.

As an additional robustness test, we have performed logistic regressions. None of our results change as we change the functional form, which is why these results are only to be found in the appendix in tables A9, A13, and A14. As mentioned in section 3.4, a further robustness test is by running the same regressions using

the modified data set. Doing this, it is possible to check for classification errors made when inputting the data from the shows. Just as with functional form, no results change significantly when testing the hypotheses with the modified data set. Again, we decide to only include the tables illustrating this in the appendix. These tables can be found in the appendix as tables A10, A15, and A16.

Another important consideration to take into account is that we will only denote the 0.1 %, 1 %, and 5 % levels of statistical significance in our tables. The reason why we excluded the 10 % levels that is commonly used as the least significant level presented is due to the vast number of hypotheses tested in this thesis. When running a high number of regressions, it is to be expected to find some non-results significant at the 10 % just by randomness. This is also a clear danger in our case on the 5 % level as well. Due to this risk of random positive results, we choose to disregard the 10 % level and to be cautious regarding results on the 5 %. However, to be transparent, we want to disclose that this decision, compared to everything presented in the method section, was made after we ran the regressions for the first time.

4.1 Discrimination

Table 2: Results for hypotheses on gender discrimination

	(1)	(2)	(3)	(4)
VARIABLES	incorr_sent	incorr_sent	incorr_sent	incorr_sent
fem_mal_disc	0.058	0.086		
	(0.563)	(0.433)		
mal_fem_disc			-0.116*	-0.128
			(0.022)	(0.106)
Constant	0.482***	0.590*	0.497***	0.457***
	(0.000)	(0.012)	(0.000)	(0.000)
Observations	183	183	385	385
Adjusted R-squared	-0.002	0.003	0.011	0.026
Controls		YES		YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

The results in table 2 do not support our first hypothesis 1.1 that females discriminate against men. There is no sign of women sending questions incorrectly with larger probability to a man than to a female. Instead, the insignificant effect is suggesting the opposite.

The second hypothesis on gender discrimination, hypothesis 2.1, gives us some interesting indications that men might discriminate against men, which is the opposite of the consensus in previous literature. However, it is not a robust result. When not employing the control variables the effect is both sizeable (-0.116) and statistically significant at the 5 % level. This would suggest that given that a man incorrectly sends a question, it is roughly 12 percentage points larger probability that the question was sent to a man, indicating discrimination.

However, when we add controls, the result is not significant anymore at any conventional level of significance. This weak finding might indicate discrimination between men, but we should not draw too strong conclusions. However, since the effect is fairly strong and goes in the opposite direction compared to our hypothesis, it would be unfair to ignore it completely.

Even if the signs of the coefficients were opposite of what was predicted, it is still possible to test hypothesis 4.3, as the values used are absolute values. However, a test does not yield any support for the hypothesis that men discriminate more than women in absolute terms. The Z-statistic found using the method described in section 3.2 is -0.312 (p-value: 0.9991). For full tables on gender discrimination, see tables A1 and A2 in the appendix.

Table 3: Results for hypotheses on age discrimination

	(1)	(2)	(3)	(4)
VARIABLES	incorr_sent	incorr_sent	incorr_sent	incorr_sent
youmid_old_disc	0.162	0.173		
	(0.101)	(0.086)		
midold_you_disc			0.059	0.033
			(0.569)	(0.735)
Constant	0.492***	0.697***	0.460***	0.431*
	(0.000)	(0.000)	(0.000)	(0.020)
Observations	263	263	271	271
Adjusted R-squared	0.023	0.028	-0.000	0.008
Controls		YES		YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

The results in table 3 show a vague indication that there might be some truth to hypothesis 2.1, i.e. that young and middle-aged individuals overestimate the ability of older individuals. Hence, indicating a positive discrimination in favor of the elders. The effect is strong around 16, 17 percentage points, but the results are not significant at the desired 5 %. However, the p-value is around

0.1 suggesting that a larger sample could be beneficial in this case.

Our second hypothesis on the topic of age discrimination, the results of which also are shown in table 3, is that elder contestants discriminate against younger contestants. However, we do not find any significant results for this hypothesis. Even before we include any control variables, the results are far from significant. Hence, we do not find any support regarding elder contestants' potential discrimination against their younger opponents. For full tables on education discrimination, see tables A3 and A4 in the appendix.

Table 4: Results for hypotheses on education discrimination

	(1)	(2)	(3)	(4)
VARIABLES	incorr_sent	incorr_sent	incorr_sent	incorr_sent
aca_work_disc	0.254**	0.236**		
	(0.001)	(0.002)		
work_work_disc			-0.175	-0.174
			(0.178)	(0.176)
Constant	0.340***	0.008	0.616***	0.655**
	(0.000)	(0.948)	(0.000)	(0.002)
Observations	332	332	193	193
Adjusted R-squared	0.061	0.110	0.024	0.057
Controls		YES		YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Furthermore, we find large and significant results suggesting that academics discriminate against workers, in the sense that they underestimate the ability of workers. The results in table 4 suggest that when a contestant with an academic background incorrectly sends a question, there is 23.6 percentage points larger probability that the question is sent to a contestant without an academic background. This is a very large effect that is not particularly affected by adding controls nor by having clustered standard errors. Results on the same significant level is also found when running a logistic regression (see table A9 in the appendix). The large and robust effect as well as the strong level of significance makes this one of the most distinguished findings of this thesis. When running the regression with the modified data set (see explanation in section 3.4), the effect drops slightly, see table A10 in appendix. However, it is still significant at the 1 % level, suggesting that the results are not driven by classification errors.

Table 4 also shows the lack of significant support for the second hypothesis on education discrimination, i.e. that workers discriminate against other work-

ers. If anything, we can see a slight indication that workers might discriminate against academics, something that would support theories regarding ingroup-outgroup effects. However, even though the direction of the effect is the negative both before and after adding controls, it is not close to statistically significant. Therefore, we do not find any support for hypothesis 3.2.

Of course, even if the sign for discrimination by the workers were negative, the opposite direction of what was predicted, we still test for hypothesis 3.3. This hypothesis proposes that academics discriminate more in absolute terms than workers. However, performing a test according the method laid out in section 3.2 does not lend any support for this. The Z-statistic for this test is 0.426 (p-value: 0.3372). For full tables on education discrimination, see tables A5 and A6 in the appendix.

Table 5: Results for hypotheses on residential discrimination

	(1)	(2)	(3)	(4)
VARIABLES	incorr_sent	incorr_sent	incorr_sent	incorr_sent
midlarg_sma_disc	-0.007	0.010		
	(0.943)	(0.906)		
smamid_larg_disc			0.026	0.034
			(0.752)	(0.668)
Constant	0.482***	0.658***	0.468***	0.512***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	284	284	324	324
Adjusted R-squared	-0.004	0.059	-0.002	0.008
Controls		YES		YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

The results for residential discrimination is presented in table 5. We are not close to finding any support neither for hypothesis 4.1 nor 4.2. Even though the number of observation is roughly the same as the number of observation in the regressions for the other hypothesis, the effects are hardly separate from zero and have very high standard errors. We conclude that there is no indications of any forms of residential discrimination.

The third hypothesis regarding this topic, hypothesis 4.3, advances a theory that people from relatively more urban cities would discriminate to a larger extent in absolute terms than people from relatively less urban cities. However, we fail to find any support for this theory using the method described in the method section. The Z-statistic found is -0.028 (p-value: 0.6103). This is not

surprising as both effects are not significantly different from zero. For full tables on residential discrimination, see tables A7 and A8 in the appendix.

4.2 Behavioral impact of gender group composition

In table 6 we present our results for the hypotheses on behavior shifts related to the gender group composition. Our first of two main findings is that when the group consist of three men, significantly more questions are being kept by the contestants compared to a single gender setting. With clustered standard errors and controls the effect is 0.045, meaning that men keep 4.5 percentage points more questions when the group consists of men only compared to when they face one other man and one woman. The positive result is significant for all robustness tests, including a logistic regression and using a modified data set to check for classification errors, please see appendix table A13 and A15 for detailed tables of these results. This result is in line with our predictions formulated in hypothesis 5.1.

Table 6: Results for hypotheses on gender group composition

	(1)	(2)	(3)	(4)
VARIABLES	kept_male	kept_male	kept_female	kept_female
gend_mmm	0.066**	0.045*		
	(0.003)	(0.035)		
gend_mff	-0.014	-0.018		
	(0.631)	(0.490)		
gend_fff			-0.115**	-0.124**
			(0.002)	(0.002)
gend_fmm			0.007	-0.017
			(0.805)	(0.567)
Constant	0.774***	0.760***	0.784***	0.695***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3,293	3,293	1,810	1,810
Adjusted R-squared	0.006	0.020	0.007	0.036
Controls		YES		YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

However, the data fails to find support for hypothesis 5.2, proposing that men would keep the question more when both of his opponents are females compared to the baseline case where he faces one male and one female. This is also clear in

table 6 as the *gend_mff* variable is not significantly different from zero. For full tables regarding hypotheses 5.1 and 5.2, please see table A11 in the appendix.

Therefore, our findings propose that men only keep questions to a larger extent when the group consist only of men. This is somewhat surprising when considering the literature on the subject (Eckel and Grossman 2002; Niederle and Vesterlund 2007). Previous literature would suggest a general behavioral difference between males and females, while lending limited support of differences of male behavior in various gender group compositions. However, we find support for a difference between the gender group compositions, but fail to show a general difference between the behavior of men and women.

Table 7 reveals the lack of difference between mens' behavior and womens' behavior in a mixed gender group, which would be found if there was a general behavioral difference between the genders. Instead, it is first when there is a single gender group we observe any difference. The table shows that the propensity to send the question away is 24 % for females and roughly 25 % for males when competing in a mixed gender group. As the difference between the behaviors in the two different categories of mixed gender groups for both genders is not significantly different from each other (see table 6), it feels perfectly natural to put them into one larger mixed gender group. Doing this, it is evident that the gender difference in preference regarding sending or keeping questions consists entirely of differences in behavior when competing in an all-male or all-female group. The regression output used to calculate table 7 can be found in the appendix as table A22.

Table 7: Percentage of questions sent away split by gender and gender group composition

		Group compo. in terms of gender	
		Single gender	Mixed gender
Gender of sender	Male	0.209	0.253
	Female	0.367	0.240

Further, our perhaps strongest result in this thesis can also be found in table 6. and concerns hypothesis 5.3, stating that females should send more questions and therefore behave less competitively in an all-female setting. Our results lend strong support for this hypothesis as the effect is not only very large (12.4 percentage units), but also significant at the 1 % level with the control variables and clustered standard errors included. A similar significant result is also found by running a logistic regression, see appendix for table A14 for details. The effect drops slightly to approximately 11.5 percentage points rather than 12.5 when checking for robustness using the modified data set, but the result is still significant at the 1 % level. Please see table A16 in the appendix for details.

Therefore, the results show that women in an all-female group chose to keep, on average, 12 percentage units fewer questions than women facing one man and one other woman. However, there is no significant difference between women's behavior when they compete against one man and one women and when they compete against two men, it is simply the single sex setting that causes a different behavior. Hence, we can reject hypothesis 5.4. This shows that women in a single gender group behave as they are more risk averse and less confident than women in a mixed gender group. For full tables regarding hypotheses 5.3 and 5.4, please see table A12 in the appendix.

Lastly, as previously discussed, it is important to distinguish between questions sent and questions incorrectly sent. In the previous hypotheses regarding discrimination, the dependent variable is a dummy variable taking different values based on whether the question was correctly or incorrectly sent. In the regressions regarding gender group behavior, our dependent variables are dummy variables taking different values based on whether the question was kept or sent and ignores whether receiver chosen was correctly chosen or not. Therefore we cannot draw any conclusion on whether the observed behavior is rational in an economic sense. Instead, our results show that men and women behave *differently* when the group consist only of men and women respectively.

5 Discussion and analysis

In this section, the results presented in the previous section are analyzed and discussed. First, the findings regarding the many hypotheses on discrimination are put up for discussion. Here, the main focus is the positive result suggesting that academics discriminate against, i.e. underestimate the ability of, workers. Yet, the non-significant results are also discussed to a lesser extent. Second, the strong findings regarding how behavior shifts with the gender group composition are considered. This part addresses both the very strong finding demonstrating that females choose to keep the question to a lesser extent in an all-female environment and the strong result showing that men keep the question more in an all-male group. Lastly, the limitations of our method and research idea are examined, followed by ideas for future research.

5.1 Discrimination

Our strongest finding regarding discrimination is a very strong and significant effect suggesting statistical discrimination against workers by academics. Our results indicate that, given all available information about the contestants' performance, it is more likely that an academic makes an incorrect decision by

sending a question to a worker than by sending a question to another academic. Since such incorrect decisions are clearly not based on the available information regarding the performance of the opponents (if they were, no incorrect decision should, by definition, be made), there must be other underlying factors affecting the decision of whom to send the question to. Our interpretation of these results is that academics believe workers to have a *relatively* lower ability than academics.

Another effect that might affect the result is the ingroup-outgroup effect. However, it is important to note that this effect might be stronger for other types of discrimination than education. Education is typically more directly related to knowledge and general ability than other characteristics, such as gender or age. Also, as Levitt (2004) found some evidence of discrimination in favor of well-educated individuals stemming not exclusively from other academics, but also from workers, there are reasons to believe that the ingroup-outgroup effect does play a relatively smaller role in the case of education discrimination. Furthermore, there are no significant results demonstrating that workers discriminate against academics. If the ingroup-outgroup effect indeed plays a major role in terms of discrimination between education groups, we would have expected clear-cut evidence showing that workers discriminate against individuals in the outgroup, i.e. against academics. However, as stated, no such results exist and the conclusion is therefore either that the ingroup-outgroup effect is only present among academics or that the effect is not the main driver or the results found for education discrimination.

Also, it is vital to point out that this study and its result adds significantly to the current state of the literature, where discriminatory behavior regarding education level has not been studied to a meaningful extent. This lack of previous empirical knowledge can probably be attributed to the difficulty of analyzing whether academics discriminate against workers in a field setting. By the nature of the labor market, academics and workers do not generally compete for the same type of jobs. Therefore, it is difficult to separate cases where an employee is subject to discrimination and cases where the worker or the academic simply does not have the skills necessary for the job. However, this unique game show setting where workers and academics repeatedly meet on fairly equal terms allows us to test for this type of discrimination. In this sense, the fact that the show uses more general trivia truly also facilitates the analysis.

On the same note, there are two other interesting features regarding labor market discrimination that the game show setting allows us to make parallels to, but that can be hard to analyze in a field setting. The two features are discrimination among equals in terms of power and hierarchy as well as discrimination in a less formal group environment. In our case, the contestants are more or less equals and the decisions are always made in groups of three. However, most major decisions in the labor market, such as employment, promotion, and salary

setting, is typically made by a well-defined boss. Hence, there is always a skewed relationship of power where one part does have a more powerful standing.

Still, less formal and less hierarchical decisions within an organization, such as the division of workload or informal responsibility within a group, is a just as important component in how individuals are being treated in the workplace. Unfortunately, these decisions are harder to quantify in order to perform economic analysis on. Since people can still, regardless of data availability, be subject to discrimination also in less formal situations, the literature needs to find different ways to observe and analyze such behavior in settings the division of power is more equal and less formal. Even though our findings are not directly applicable to a labor market setting, it still provides clues about statistical discrimination among equals in an informal group setting and how a certain group underestimate the relative ability of another group.

So far in this section, only the result regarding academics' discrimination of workers, the only significant result, and its potential implications have been discussed. Below we will briefly discuss the hypotheses for which no significant results are found.

First, no robust support for any of the hypotheses regarding gender discrimination could be found, even though there are some slight indications of men discriminating against men. Most previous research on the topic of gender discrimination have found that men, if they discriminate at all, discriminate against women. Also, in the light of the research on ingroup-outgroup effects, these indicative results are surprising. In terms of ingroup-outgroup effects, this could be partially explained by that the gender group is not the most "important" group that contestants identify themselves with. Instead, contestants might identify themselves with others based on other characteristics, which would explain the lack of gender ingroup-outgroup discrimination. Similarly, we do not find any support for the hypothesis that women discriminate against men. However, judging by previous literature, our non-result is not completely surprising. While a few researchers (e.g. Antonovics, Arcidiacono, and Walsh 2005) have found that women tend to discriminate against men, the literature on discrimination have not reached a verdict regarding this topic.

On the same note, the complete lack of clear results for gender discrimination is a bit surprising. As with other types of discrimination, it is harder to blame it on potential classification errors, as the classification of gender is a very clear-cut. Also, the gender of the contestant is very visible through-out the entire game, making the risk of forgetting the information impossible. On the other hand, the fact that the issue of gender discrimination has been very much on the agenda for the past years might have made people aware of the phenomenon, which potentially might make people able to safeguard themselves against gender discrimination. It could even be the case that people, being afraid to be regarded as

sexist or chauvinist, act in the opposite way, which would explain the indication of men-to-men discrimination. If true, this fact could be used to illustrate that an active debate on discrimination can have an impact on people's behavior. Another, perhaps less far-reaching, explanation of the lack of results would be that people do not view gender as a proxy for ability in the same way they view education. This belief could, however, also be affected by the aforementioned active public debate.

Moving on, we do not find significant support for the hypotheses regarding age discrimination. Before including clustered standard errors, our results regarding age discrimination show some indication of discrimination by young and middle-aged senders *in favor* of old contestants, which can be seen in appendix table A3. Individuals who have lived longer have had more time to obtain knowledge and experiences have, as a result, better potential of having good trivia knowledge. On the other hand, younger people might see elder people as "old and muddle-headed" and therefore underestimate the ability of the old. As a result, these two potential effects might take out each other and leave us with a non-result. We find no support for our other hypothesis on age discrimination, i.e. that older constants discriminate against younger. Even though younger individuals have less life experience, the results does not suggest that older contestants have beliefs that younger contestants should have a lower ability.

However, it is worth to point out the possibility that we would have found an effect using other age spans. Now, we put 18 year old secondary school students in the same category as young professional and these two groups might very well be treated differently by elder contestants. Still, to make more detailed judgments of a person's age would have been more problematic in being able to uphold an objective and consistent classification.

Lastly, there are no indications of any residential discrimination. This is not completely surprising as this notion is not supported extensively in the previous literature. Since large cities generally have higher educated individuals (Hammarström 2004), it is not unreasonable to believe that people would think that individuals from smaller towns would have a relatively lower ability. However, this is idea has no backing even before controlling for education. Yet, it is important to note that, even though the players' place of residence is revealed in the beginning of the show, it is not sure that all players remember the size of the hometowns of all their opponents throughout the game. Still, it is unreasonable to assume that every player forget about the others' place of residence and if there were any discrimination of this kind, we would have seen at least some indications of it.

At the same time, the same argument could be made regarding the education discrimination, which carries such a strong support in the data. However, it might very well be the case that the type of work someone does is easier to

remember than place of residence. Another reason for us finding strong support regarding education discrimination, but not residential discrimination may be that the type of education and job a person has is more telling in a person's physical appearance than where they live. It is very possible that a person's education attainment is more traceable in how the person dresses, talks or otherwise appears, at least more so than his or her place of residence.

5.2 Behavioral impact of gender group composition

There are undoubtedly differences in gender outcomes in many societal aspects, notably in the labor market. While there probably are separate explanations for these discrepancies, our findings support that there are behavioral differences between men and women in certain group settings. Yet, we can only claim that variations in gender composition seems to account for a small part of the total variation in decision-making, as our regressions consistently have a very small R-squared. However, this is not unexpected or a vital blow to our point, as it is unreasonable that the composition of the group should on average be a bigger factor than pure individual preferences and feelings, overall strategy, game situation, etc. Still, if we observe significant differences in behavior, it is not unreasonable to expect certain differences in outcomes, assuming that behavior in groups is a relevant factor in determining outcomes in e.g. the labor market.

As stated, our findings propose that men keep questions to a larger extent when the group consist only of men. This somewhat in line with much of the literature on the topic. It is fairly established that men are less averse to risk (Eckel and Grossman 2002), prefer competition to a larger extent (Price 2008) and are more confident (Niederle and Vesterlund 2007) than women. However, we do not find support for this in our results. Men and women do not behave differently in general, instead, it is first when they compete in single-gender groups we observe any difference.

Therefore, our new findings suggest that group composition might be of higher importance than earlier studies have indicated. The primary focus on research regarding gender differences in competitiveness and risk preferences has been studying how men and women respond to competition or risk taking in a mixed-gender environment (e.g. Gneezy, Niederle, and Rustichini 2003; Dreber, von Essen, and Ranehill 2011). While this has yielded some interesting results, showing that men (and boys) generally prefer competition to a larger extent than women does, it provides little information about *when* and *why* this is the case. This thesis does not address the root question *why* we observe differences between the genders, but it provides important clues regarding *when* these differences can be observed.

However, we propose a potential explanation for why we find different behaviors in single gender groups than in mixed gender groups. It could be the case that some kind of standard for how to behave within the group is set more easily when all group members have the same gender. If one male in an all-male group starts to behave in a certain way, he might set the standard for the group and the rest of the group would continue with a similar behavior. If a women that start to send questions to her opponents she might set a standard that indicates that this behavior is "the way to go", and hence it is more natural for the other contestant to mimic that behavior. And since there is no man to potentially break this pattern, it carries on.

However, in a mixed group, no clear consensus on how to behave in this group is set and contestants of both genders adapt to the behavior of the other gender, creating a general behavioral pattern for both genders in a mixed gender setting. This is assuming that the homogeneity of the group makes it easier to establish a consensus on "how to behave." A further assumption is that that men have an inherent desire to act more competitively, i.e. keep the question, and that women have the inherent desire to act less competitively by sending questions. As pointed out earlier, this notion is supported by earlier literature and would, in a sense, unite our findings with the current literature.

It is important to note that none of these behaviors can be considered *good* or *bad* in a competitive setting. As stated, keeping a question is simply associated with a preference for a riskier strategy or higher self-confidence since both opponent can receive points instead of just one opponent. In order to draw any conclusions on the objective quality of behavior, it would be necessary to map a complete optimal strategy for each decision, which would be very difficult in this comparatively complex game show. As a consequence, we cannot draw any direct parallels to research regarding individuals' *performance* in single-gender versus mixed-gender settings. This, however, is beyond the scope of this thesis, instead we show that there is a significant difference in *behavior* when individuals are put in a single-gender and mixed-gender environments, which is something that previous literature has paid less attention to. We argue that a better understanding for *how* men and women behave in single-gender groups compared to mixed-gender groups is necessary in order to understand any differences in performance and hence, differences in social and professional outcomes.

5.3 Limitations

A limitation of this design is of course that it is not purely a real world setting. While we argue that we have a well-diversified and large sample, strong incentives, and a setting different from that of a regular laboratory experiment, it is still important to point out that individuals' behavior could change due to the

fact that they are shown on television. However, even though there is no clear consensus regarding how the game show setting affects the risk behavior, we are aware of the potential pitfalls and that peoples' behavior on TV might differ from peoples' behavior in real life off-camera. Compared to a classic natural experiment, this design may have somewhat lower external validity. On the other hand, we claim that this rather isolated setting increase the internal validity to a more satisfying level than the typical natural experiment would have.

Furthermore, we have tested how individuals' preferences for risk and confidence change based on the group composition. While this has been the explicit goal, we cannot explain the underlying factors affecting this behavior. There are a few possible factors that might explain why people choose to keep or send a question, where the two main components arguably are risk aversion and self-confidence. Other potential effects could be a shame effect (i.e. that it is embarrassing to give wrong answers on television), preferences for equality (i.e. that all contestants should answer the same amount of questions) or altruism (i.e. you receive utility when giving other players a greater chance to win). In our thesis we have not tried to separate these effects, since they yield the same outcome. Even though the study of the underlying factors might be of interest in itself, it is beyond the scope of our thesis since we focus on *when* (i.e. in which group settings) this behavior occurs rather than *why* (i.e. what underlying factors that drive this behavior).

As noted in section 3.2, we would have wished to perform our discrimination analysis on one of the most studied forms of discrimination, race discrimination. Since race discrimination is one of the most debated forms of discrimination in our society, a study of this particular form of discrimination in a game show setting would clearly have been desirable. However, since the sample of non-Swedes is way too small, an analysis of this kind is simply not possible in this setting. We cannot exclude that race discrimination occurs in earlier stages of *Vem Vet Mest* (as well as other types of discrimination) but testing for that would require a completely different method. Furthermore, as the contestants have more options and different types of strategies, we claim that it would be more difficult to sufficiently isolate for discriminative behavior in the earlier stages.

A minor limitation with our data is that it consist of a mix between Friday final shows and weekday shows, as we added roughly 50 weekday shows. As discussed in section 3.4, the weekday shows before season 9 have the same incentive structure as the Friday final does, i.e the first place is the only position that matters and contestants should be indifferent between second and third place. However, the financial rewards are different between the Friday final and weekday shows. As a result, the incentives for winning the weekday shows might not be as strong as winning the Friday final and this could potentially have an effect on contestants' behavior. However, we would argue that, in the spirit of the moment,

people will probably make the same decisions in both weekday and Friday finals, striving to win the game and as well as answer every question correctly. This is also hampered by the victor being awarded an extra light bulb, which indirectly increases your chance of winning the money in the Friday final.

Furthermore, an ideal game show setting would not disclose any information on the topic of the question before the contestants were to decide whether to keep or send it. Now, even with vague topics, contestants might in some cases be able to adjust their decision based on the topic. For example, they might send a question on pop music to an old person, even though this would be classified as an incorrect question. However, the topics are often hard to infer any information from. Further, the impact of this is also limited by that any potential biases, e.g. sending questions on pop music to old people, would even out if not the elder contestants widely outperform the younger, as these types of questions would result in both incorrectly and correctly sent questions to an equal degree. Lastly, this would probably be most problematic for the hypotheses on discrimination, which we, with one exception, reject anyway.

Also, the fact that the contestant can interact with each other to some extent while off-camera is of course somewhat problematic. This issue is discussed in section 3.4 and, even though this is not a major concern, it is important to note that some contestants can in some cases have more information about their opponents than we as TV-audience have. However, the impact of this potential issue on our results are minor, even though it could exacerbate another potential issue in that we might have made errors when inputting the demographics data for our contestant. In some cases, the other contestants could possess the missing or vague information not apparent for us as TV-audience. Still, as is described in section 3.4 and shown in tables A10, A15, and A16 in the appendix, our results are robust when checking for any classification errors. Hence, there is no need to worry about this interaction.

Finally, another minor point is that this study is performed in a competitive setting. Therefore, we can only draw strong conclusions about peoples' behavior in a competitive setting. We cannot draw any conclusions about the behavior in cooperative settings. It is not unreasonable to believe that a similar behavior would occur in cooperative settings and further research on the topic would be highly interesting.

5.4 Future research

As stated above, we would find general studies expanding on the topics of discrimination and group behavior dependent on the gender composition in game shows of high interest. Our primary research recommendation would be to study

discrimination and group behavior in a cooperative setting, compared to the competitive setting in *Vem vet mest*. Also, in the spirit of our thesis, it would be interesting to see if the gender of the game show host affect the contestants' behavior. After all, the host might be regarded as a member of the group. Further, a study looking at learning effects, judging whether the amount of sub-optimal decisions diminishes both between and in the various stages in *Vem vet mest* would be intriguing. This has been done in other shows before (Berk, Hughson, and Vandezande 1996)

Another possible extension of our study would be to look at how the effect of knowing that you will be observed by hundreds of thousands of TV viewers affects your behavior, i.e. the audience effect. This could be done by staging a game of *Vem vet mest* with the only difference in that the show will be not be aired and then also possibly staging completely anonymous variants of the show in a laboratory setting. It is widely acknowledged in behavioral economics that the degree of anonymity has an impact on choices and behavior and in this sense it would be interesting to treat a behavior in a TV show as the absolute extreme level of non-anonymity.

As mentioned several times, our study does not address the question *why* we observe differences in competitiveness between the genders. Instead it only provides explanations for *when* these differences can be observed. As discussed, there are several effects, including risk preferences, overconfidence, shame effect, etc., that could explain why we observe differences. A natural enlargement of this study would therefore be perform a study designed to isolate which underlying effect that affect the difference in competitiveness. As mentioned in section 2.3, the literature regarding education discrimination is very limited. Since this type of discrimination is the only type detected in this study, there are reasons to believe that this discrimination exists also in other fora. Therefore, further studies on this topic in different settings could potentially lead to new findings regarding education. Particularly, it would be very interesting to delve deeper into what makes education so salient. It could either work as a clear proxy simply for the ability to answer trivia questions or in the sense that education level provides a signaling effect of higher ability in general. Any research trying to disentangle from where exactly the underestimation of workers by academics stem would be of interest.

Lastly, our strong results indicate a large possibility to use game shows in other types of studies. The wide variety in both topic and complexity of existing game shows indeed provide great opportunities for researchers. Yet, it still easier to isolate for the behavior the researcher wants to study in the most complex game show than in the true real world. Hence, game shows could be a useful middle-ground between lab experiments and pure field experiments, having a higher external validity than the former and a higher internal validity than the latter.

6 Concluding remarks

This thesis tries to take a broad approach on discrimination, testing a variety of types including gender, age, education, and residential discrimination. By analyzing advantageous empirical data, the results can confirm the existence of only one type of discrimination, namely that academics tend to discriminate against workers. However, we are not able to find any support for the other hypotheses regarding gender, age or residential discrimination. This implies that education is likely seen as a proxy for ability or as a signal of ability by academics.

While the results show no signs of gender discrimination, they provide clear evidence that both men and women behave differently in single-gender groups compared to mixed-gender groups. The strong indication that men become more competitive in an all-male setting while women become less competitive in an all-female setting suggests that behavioral differences might explain the gender gap to a larger extent than previous suggested by the literature. Further, this insight is also important to keep in mind when shaping policies and environments aiming to reduce the gender differences in societal and labor market outcomes.

7 References

Aigner, D.J. and Cain, G.G. 1977. Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review* 30:175-87.

Akerlof, G.A. and Kranton, R.E. 2000. Economics and identity. *Quarterly Journal of Economics* 115:715-53.

Altonji, J.G. and Blank, R.M. 1999. Race and gender in the labor market. *Handbook of Labor Economics* 3:3143-259.

Anderson, D.M. and Haupert, M.J. 1999. Employment and statistical discrimination: a hands-on experiment. *Journal of Economics* 25:85-102.

Andreoni, J. and Petrie, R. 2004. Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics* 88:1605-23.

Antonovics, K., Arcidiacono, P. and Walsh, R.P. 2005. Games and discrimination lessons from the weakest link. *Journal of Human Resources* 40:918-47.

Antonovics, K., Arcidiacono, P. and Walsh, R.P. 2003. Competing against the opposite sex. Working Paper, University of California at San Diego. Available at: https://escholarship.org/uc/item/0kx2f7xq.

Anwar, S. 2012. Testing for discrimination: Evidence from the game show Street Smarts. *Journal of Economic Behavior and Organization* 81:268-85.

Arch, E.C. 1993. Risk-taking: a motivational basis for sex differences. *Psychological Reports* 73:3-11.

Ariely, D., Bracha, A. and Meier, S. 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review* 99(1):544-555.

Arrow, K.J. 1973. The theory of discrimination. O. Ashenfelter and A. Rees, eds. *Discrimination in Labor Markets.* 1st edn., Cambridge, MA: Princeton University Press, pp. 3-33.

Arrow, K.J. 1998. What has economics to say about racial discrimination?. *Journal of Economic Perspectives* 12:91-100.

Arvey, R.D., Miller, H.E., Gould, R. and Burch, P. 1987. Interview validity for selecting sales clerks. *Personnel Psychology* 40:1-12.

Baldwin, M. and Johnson, W.G. 1994. Labor market discrimination against men with disabilities. *Journal of Human Resources* 29:1-19.

Ball, S.B. and Eckel, C.C. 1998. Stars upon thars: status and discrimination in ultimatum games. Working Paper, Virginia Polytechnic Institute and State University, Department of Economics. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.130.2412andrep=rep1andtype=pdf.

Baltussen, G., van den Assem, M.J. and van Dolder, D. 2014. Risky choice in the limelight. *Review of Economics and Statistics* (Forthcoming). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2057134.

Becker, G.S. 1957. The economics of discrimination. 1st edn., Chicago, IL: University of Chicago Press.

Beetsma, R.M. and Schotman, P.C. 2001. Measuring risk attitudes in a natural experiment: data from the television game show Lingo. *Economic Journal* 111:821-48.

Belot, M., Bhaskar, V. and van de Ven, J. 2010. Promises and cooperation: evidence from a TV game show. *Journal of Economic Behavior and Organization* 73:396-405.

Benson, P. 2001. The Hawthorne effect. W.E. Craighead and C.B. Nemeroff, eds. *The Corsini Encyclopedia of Psychology and Behavioral Science.* 3rd edn., New York City, NY: Wiley, pp. 427-28.

Berk, J.B., Hughson, E. and Vandezande, K. 1996. The price is right, but are the bids? An investigation of rational decision theory. *American Economic Review.* 86(4):954-70.

Bernhard, H., Fehr, E. and Fischbacher, U. 2006. Group affiliation and altruistic norm enforcement. *American Economic Review* 96(2):217-21.

Bertrand, M. 2011. New perspectives on gender. *Handbook of Labor Economics* 4:1543-90.

Blau, F.D. and Ferber, M.A. 1987. Discrimination: empirical evidence from the United States. *American Economic Review* 77(2):316-20.

Buser, T., Niederle, M. and Oosterbeek, H. 2014. Gender, competitiveness and career choices. *Quarterly Journal of Economics* 129:1409-47.

Byrnes, J.P., Miller, D.C. and Schafer, W.D. 1999. Gender differences in risk taking: a meta-analysis. *Psychological Bulletin* 125:367-83.

Ceraulo, S.C. 1999. Separated by Sex: A critical look at single-sex education for girls. *Journal of Chemical Education* 76:615-20.

Charness, G. and Gneezy, U. 2008. What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior and Organization* 68:29-35.

Charness, G., Rigotti, L. and Rustichini, A. 2007. Individual behavior and group membership. *American Economic Review* 97(4):1340-52.

Chen, Y. and Li, S.X. 2009. Group identity and social preferences. *American Economic Review* 99(1):431-57.

Clogg, C.C., Petkova, E., and Haritou, A. 1995. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology* 100:1261-93.

Croson, R. and Gneezy, U. 2009. Gender differences in preferences. *Journal of Economic Literature* 47:448-74.

Dasgupta, N. 2004. Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research* 17:143-69.

Datta Gupta, N., Poulsen, A. and Villeval, M.C. 2013. Gender matching and competitiveness: experimental evidence. *Economic Inquiry* 51:816-35.

Deck, C., Lee, J. and Reyes, J. 2008. Risk attitudes in large stake gambles: evidence from a game show. *Applied Economics* 40:41-52.

Dillingham, A.E., Ferber, M.A. and Hamermesh, D.S. 1994. Gender discrimination by gender: voting in a professional society. *Industrial and Labor Relations Review* 47:622-33.

Dreber, A., von Essen, E. and Ranehill, E. 2011. Outrunning the gender gap—boys and girls compete equally. *Experimental Economics* 14:567-82.

Duncan, C. and Loretto, W. 2004. Never the right age? Gender and agebased discrimination in employment. *Gender, Work and Organization* 11:95-115.

Eckel, C.C. and Grossman, P.J. 2008. Differences in the economic decisions of men and women: experimental evidence. *Handbook of Experimental Economics Results* 1:509-19.

Fershtman, C. and Gneezy, U. 2001. Discrimination in a segmented society: an experimental approach. *Quarterly Journal of Economics* 116:351-77.

Fox, C.R. and Tversky, A. 1995. Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics* 110:585-603.

Gertner, R. 1993. Game shows and economic behavior: risk-taking on "Card Sharks." *Quarterly Journal of Economics* 108:507-21.

Gibson, K.J., Zerbe, W.J. and Franken, R. 1993. The influence of rater and ratee age on judgments of work-related attributes. *The Journal of Psychology* 127:271-80.

Gneezy, U., Leonard, K.L. and List, J.A. 2009. Gender differences in competition: evidence from a matrilineal and a patriarchal society. *Econometrica* 77:1637-64.

Gneezy, U., Niederle, M. and Rustichini, A. 2003. Performance in competitive environments: gender differences. *Quarterly Journal of Economics* 118:1049-74.

Halpern, D.F., Eliot, L., Bigler, R.S., Fabes, R.A., Hanish, L.D., Hyde, J., Liben, L.S. and Martin, C.L. 2011. Education. The pseudoscience of single-sex schooling. *Science* 333:1706-7.

Hammar, E. 2015. Telephone interview conducted by Jesper Lundström, May 7 2015.

Hammarström, M. 2004. En högskola för alla?: Vägen dit för glesbygdens ungdomar. Jönköping, Sweden: Högskolan för lärande och kommunikation.

Hargreaves Heap, S.P. and Zizzo, D.J. 2009. The value of groups. *American Economic Review* 99(1):295-323.

Harker, R. 2000, Achievement, gender and the single-sex/coed debate. *Journal of Sociology of Education* 21:203-18.

Hartley, R., Lanot, G. and Walker, I. 2006. Who really wants to be a millionaire? Estimates of risk aversion from gameshow data. *Journal of Applied Econometrics* 29:861-79.

Healy, P. and Noussair, C. 2004. Bidding behavior in the price is right game: an experimental study. *Journal of Economic Behavior and Organization* 54:231-47.

Hensvik, L.E. 2014. Manager impartiality: worker-firm matching and the gender wage gap. *Industrial and Labor Relations Review* 67:395-421.

Lee, V.E. and Bryk, A.S. 1986. Effects of single-sex secondary schools on student achievement and attitudes. *Journal of Education Psychology* 78:381-95.

Lee, V.E. and Marks, H.M. 1990. Sustained effects of the single-sex secondary school experience on attitudes, behaviors, and values in college. *Journal of Education Psychology* 82:578-92.

Levitt, S.D. 2004. Testing theories of discrimination: Evidence from "Weakest Link." *Journal of Law and Economics* 47:431-53.

Lindquist, G.S. and Säve-Söderbergh, J. 2011. "Girls will be girls", especially among boys: risk-taking in the "daily double" on Jeopardy. *Economics Letters* 112:158-60.

List, J.A. 2006. Friend or foe? A natural experiment of the prisoner's dilemma. *Review of Economics and Statistics* 88:463-71.

Locke, E.A. 1973. Satisfiers and dissatisfiers among white-collar and blue-collar employees. *Journal of Applied Psychology* 58:67-76.

Lundberg, S.J. and Startz, R. 1983. Private discrimination and social intervention in competitive labor market. *American Economic Review* 73(3):340-7.

Manning, A. and Saidi, F. 2010. Understanding the gender pay gap: what's competition got to do with it?. *Industrial and Labor Relations Review* 63:681-98.

Marianne, B. 2011. New perspectives on gender. *Handbook of Labor Economics* 4:1543-90.

McGoldrick, A.E. and Arrowsmith, J. 1993. Recruitment advertising: discrimination on the basis of age. *Employee Relations* 15:54-65.

Metrick, A. 1995. A natural experiment in "Jeopardy!". *American Economic Review* 85(1):240-53.

Niederle, M. and Vesterlund, L. 2007. Do women shy away from competition? Do men compete too much?. *Quarterly Journal of Economics* 122:1067-101.

Nowell, C. and Tinkler, S. 1994. The influence of gender on the provision of a public good. *Journal of Economic Behavior and Organization* 25:25-36.

Oaxaca, R.L. and Ransom, M.R. 1994. On discrimination and the decomposition of wage differentials. *Journal of Econometrics* 61:5-21.

Ortmann, A. and Tichy, L.K. 1999. Gender differences in the laboratory: evidence from prisoner's dilemma games. *Journal of Economic Behavior and Organization* 39:327-39.

Paternoster, R., Brame, R., Mazerolle, P., and Piquero, A. 1998. Using the correct statistical test for equality of regression coefficients. *Criminology* 36:859-66.

Perry, E.L., Kulik, C.T. and Bourhis, A.C. 1996. Moderating effects of personal and contextual factors in age discrimination. *Journal of Applied Psychology* 81:628-47.

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.

Powell, M. and Ansic, D. 1997. Gender differences in risk behaviour in financial decision-making: an experimental analysis. *Journal of Economic Psychology* 18:605-28.

Price, J. 2008. Gender differences in the response to competition. *Industrial and Labor Relations Review* 61:320-33.

Prince, M. 1993. Women, men and money styles. *Journal of Economic Psychology* 14:175-82.

Riach, P.A. 2015. A field experiment investigating age discrimination in four European labour markets. *International Review of Applied Economics* (Forthcoming). Available at: http://www.tandfonline.com/doi/full/10.1080/02692171.2015.1021667.

Riach, P.A. and Rich, J. 2010. An experimental investigation of age discrimination in the English labor market. *Annals of Economics and Statistics* 99:169-85.

Schwab, S. 1986. Is statistical discrimination efficient?. *American Economic Review* 76(6):228-34.

Singer, M. and Sewell, C. 1989. Applicant age and selection interview decisions: effect of information exposure on age discrimination in personnel selection. *Personnel Psychology* 42:135-54.

Snape, E. and Redman, T. 2003. Too old or too young? The impact of perceived age discrimination. *Human Resource Management Journal* 13:78-89.

Stiglitz, J.E. 1973. Approaches to the economics of discrimination. *American Economic Review* 63(2):287-95.

Sutter, M., Lindner, P. and Platsch, D. 2009. Social norms, third-party observation and third-party reward, Working Paper. University of Innsbruck, Department of Public Finance. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.380.8704rep=rep1type=pdf.

Tajfel, H. 1970. Experiments in intergroup discrimination. *Scientific American* 223:96-102.

Tenorio, R. and Cason, T.N. 2002. To spin or not to spin? Natural and laboratory experiments from The price is right. *Economic Journal* 112:170-95.

Vaughan, G.M., Tajfel, H. and Williams, J. 1981. Bias in reward allocation in an intergroup and an interpersonal context. *Social Psychology Quarterly* 44:37-42.

Wickström, G. and Bendix, T. 2000. The "Hawthorne effect"—what did the original Hawthorne studies actually show?. *Scandinavian Journal of Work, Environment and Health* 26:363-7.

Wright, R.E. and Ermisch, J.F. 1991. Gender discrimination in the British labour market: a reassessment. *Economic Journal* 101:508-22.

8 Appendix

The items in the appendix appears in the following order with one table per page:

- Extended regression tables for discrimination hypotheses
  - Table A1: Hypothesis 1.1 - Females against males
  - Table A2: Hypothesis 1.2 - Males against females
  - Table A3: Hypothesis 2.1 - Young/middle-aged against old
  - Table A4: Hypothesis 2.2 - Old/middle-aged against young
  - Table A5: Hypothesis 3.1 - Academics against workers
  - Table A6: Hypothesis 3.2 - Workers against academics
  - Table A7: Hypothesis 4.1 - Small/mid cities against large cities
  - Table A8: Hypothesis 4.2 - Mid/large cities against small cities
- Robustness check tables for discrimination hypotheses
  - Table A9: Logistic regressions
  - Table A10: Classification error regressions
- Extended regression tables for group behavior hypotheses
  - Table A11: Hypothesis 5.1 and 5.2 - Male behavior
  - Table A12: Hypothesis 5.3 and 5.4 - Female behavior
- Robustness check tables for group behavior hypotheses
  - Table A13: Logistic regression for male behavior
  - Table A14: Logistic regression for female behavior
  - Table A15: Classification error regression for male behavior
  - Table A16: Classification error regression for female behavior
- Descriptive statistics and variable definitions
  - Table A17: Dependent variables for discrimination hypotheses
  - Table A18: Control variables for discrimination hypotheses
  - Table A19: Variables for group behavior hypotheses
  - Table A20: Descriptive statistics for discrimination variables
  - Table A21: Descriptive statistics for group behavior variables
- Other
  - Table A22: Interaction variable regression on gender composition
  - List A23: Example questions

Table A1: Hypothesis 1.1 - Females' discrimination against males

VARIABLES	(1) incorr_sent	(2) incorr_sent	(3) incorr_sent	(4) incorr_sent	(5) inc_sent_type1	(6) inc_sent_type2	(7) inc_sent_type1	(8) inc_sent_type2
fem_mal_disc	0.058	0.058	0.086	0.086	-0.005	0.063	0.004	0.082
	(0.437)	(0.563)	(0.327)	(0.433)	(0.940)	(0.280)	(0.969)	(0.220)
educ_send_d			-0.047	-0.047			0.069	-0.116
			(0.564)	(0.541)			(0.403)	(0.110)
young_send_d			0.085	0.085			0.076	0.009
			(0.525)	(0.452)			(0.453)	(0.938)
midage_send_d			0.040	0.040			0.077	-0.038
			(0.732)	(0.718)			(0.438)	(0.700)
smallres_send_d			-0.170	-0.170			0.043	-0.212*
			(0.129)	(0.090)			(0.678)	(0.048)
midres_send_d			-0.101	-0.101			0.067	-0.168
			(0.332)	(0.298)			(0.451)	(0.089)
educ_rec_d			-0.115	-0.115			-0.153	0.038
			(0.161)	(0.173)			(0.125)	(0.604)
young_rec_d			-0.028	-0.028			0.074	-0.102
			(0.826)	(0.833)			(0.619)	(0.418)
midage_rec_d			-0.069	-0.069			0.101	-0.170
			(0.559)	(0.550)			(0.412)	(0.176)
smallres_rec_d			-0.068	-0.068			0.059	-0.126
			(0.563)	(0.617)			(0.627)	(0.218)
midres_rec_d			0.033	0.033			0.121	-0.088
			(0.724)	(0.737)			(0.218)	(0.249)
educ_static_d			0.174*	0.174*			0.003	0.171**
			(0.045)	(0.042)			(0.975)	(0.005)
young_static_d			-0.109	-0.109			-0.077	-0.032
			(0.399)	(0.476)			(0.629)	(0.815)
midage_static_d			-0.074	-0.074			-0.015	-0.059
			(0.541)	(0.621)			(0.915)	(0.605)
smallres_static_d			-0.019	-0.019			-0.069	0.050
			(0.881)	(0.890)			(0.589)	(0.619)
midres_static_d			0.086	0.086			0.083	0.003
			(0.361)	(0.397)			(0.409)	(0.973)
Constant	0.482***	0.482***	0.590**	0.590*	0.325***	0.157***	0.127	0.463
	(0.000)	(0.000)	(0.005)	(0.012)	(0.000)	(0.000)	(0.482)	(0.053)
Observations	183	183	183	183	183	183	183	183
Adjusted R-squared	-0.002	-0.002	0.003	0.003	-0.005	0.001	-0.027	0.082
Cluster		YES		YES			YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A2: Hypothesis 1.2 - Males' discrimination against females

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	incorr_sent	incorr_sent	incorr_sent	incorr_sent	inc_sent_type1	inc_sent_type2	inc_sent_type1	inc_sent_type2
mal_feml_disc	-0.116*	-0.116	-0.128*	-0.128	-0.081	-0.035	-0.095	-0.033
	(0.022)	(0.156)	(0.014)	(0.106)	(0.084)	(0.315)	(0.205)	(0.324)
educ_send_d			0.076	0.076			0.018	0.058
			(0.174)	(0.166)			(0.722)	(0.186)
young_send_d			-0.189*	-0.189*			-0.224**	0.036
			(0.023)	(0.027)			(0.002)	(0.669)
midage_send_d			-0.057	-0.057			-0.087	0.030
			(0.393)	(0.366)			(0.188)	(0.636)
smallres_send_d			-0.033	-0.033			-0.006	-0.027
			(0.662)	(0.671)			(0.922)	(0.665)
midres_send_d			0.096	0.096			0.077	0.019
			(0.104)	(0.108)			(0.155)	(0.694)
educ_rec_d			-0.024	-0.024			0.008	-0.032
			(0.651)	(0.716)			(0.901)	(0.477)
young_rec_d			0.056	0.056			0.071	-0.015
			(0.470)	(0.604)			(0.481)	(0.760)
midage_rec_d			0.108	0.108			0.047	0.061
			(0.115)	(0.199)			(0.565)	(0.188)
smallres_rec_d			-0.008	-0.008			0.077	-0.084
			(0.910)	(0.926)			(0.310)	(0.103)
midres_rec_d			0.030	0.030			0.087	-0.056
			(0.597)	(0.669)			(0.164)	(0.234)
educ_staticsend_d			0.049	0.049			0.009	0.040
			(0.397)	(0.511)			(0.899)	(0.350)
young_staticsend_d			-0.061	-0.061			-0.047	-0.014
			(0.470)	(0.602)			(0.678)	(0.804)
midage_staticsend_d			-0.126*	-0.126			-0.150	0.024
			(0.046)	(0.150)			(0.099)	(0.605)
smallres_staticsend_d			-0.030	-0.030			0.009	-0.039
			(0.668)	(0.731)			(0.896)	(0.550)
midres_staticsend_d			0.043	0.043			0.150*	-0.107*
			(0.471)	(0.553)			(0.030)	(0.024)
Constant	0.497***	0.497***	0.457***	0.457***	0.343***	0.154***	0.329**	0.129
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.007)	(0.155)
Observations	385	385	385	385	385	385	385	385
Adjusted R-squared	0.011	0.011	0.026	0.026	0.005	0.000	0.054	0.017
Cluster		YES		YES			YES	YES [b]

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A3: Hypothesis 2.1 - The discrimination by young/middle-aged people against old people

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	incorr_sent	incorr_sent	incorr_sent	incorr_sent	inc_sent_type1	inc_sent_type2	inc_sent_type1	inc_sent_type2
youmid_old_disc	0.162**	0.162	0.173**	0.173	0.219***	-0.058	0.246**	-0.073
	(0.008)	(0.101)	(0.007)	(0.086)	(0.000)	(0.239)	(0.010)	(0.121)
gend_send_d			-0.115	-0.115			-0.055	-0.060
			(0.072)	(0.080)			(0.463)	(0.464)
educ_send_d			0.051	0.051			0.032	0.019
			(0.442)	(0.418)			(0.670)	(0.797)
smallres_send_d			-0.076	-0.076			-0.063	-0.013
			(0.366)	(0.371)			(0.459)	(0.886)
midres_send_d			-0.094	-0.094			-0.029	-0.065
			(0.220)	(0.263)			(0.730)	(0.476)
gend_rec_d			-0.090	-0.090			-0.065	-0.026
			(0.150)	(0.250)			(0.434)	(0.673)
educ_rec_d			-0.008	-0.008			0.012	-0.019
			(0.909)	(0.912)			(0.875)	(0.755)
smallres_rec_d			-0.019	-0.019			0.150	-0.169*
			(0.828)	(0.852)			(0.152)	(0.034)
midres_rec_d			0.054	0.054			0.177*	-0.123
			(0.479)	(0.551)			(0.044)	(0.122)
gend_static_d			-0.016	-0.016			0.055	-0.071
			(0.801)	(0.829)			(0.497)	(0.250)
educ_static_d			-0.043	-0.043			-0.048	0.006
			(0.554)	(0.596)			(0.578)	(0.922)
smallres_static_d			-0.123	-0.123			0.029	-0.152
			(0.149)	(0.276)			(0.793)	(0.075)
midres_static_d			0.022	0.022			0.118	-0.095
			(0.777)	(0.807)			(0.294)	(0.222)
Constant	0.492***	0.492***	0.697***	0.697***	0.269***	0.223***	0.152	0.545**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.443)	(0.002)
Observations	263	263	263	263	263	263	263	263
Adjusted R-squared	0.023	0.023	0.028	0.028	0.047	0.002	0.050	0.017
Cluster		YES		YES			YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A4: Hypothesis 2.2 - The discrimination by old/middle-aged people against young people

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	incorr_sent	incorr_sent	incorr_sent	incorr_sent	inc_sent_type1	inc_sent_type2	inc_sent_type1	inc_sent_type2
midold_you_disc	0.059	0.059	0.033	0.033	0.065	-0.006	0.056	-0.022
	(0.341)	(0.569)	(0.607)	(0.735)	(0.261)	(0.896)	(0.529)	(0.600)
gend_send_d			-0.023	-0.023			0.130	-0.154*
			(0.727)	(0.757)			(0.066)	(0.031)
educ_send_d			-0.085	-0.085			-0.058	-0.027
			(0.190)	(0.218)			(0.408)	(0.665)
smallres_send_d			-0.057	-0.057			0.050	-0.106
			(0.514)	(0.450)			(0.447)	(0.124)
midres_send_d			0.079	0.079			0.030	0.049
			(0.296)	(0.309)			(0.676)	(0.495)
gend_rec_d			0.087	0.087			0.108	-0.021
			(0.180)	(0.318)			(0.166)	(0.749)
educ_rec_d			0.015	0.015			-0.029	0.044
			(0.822)	(0.865)			(0.741)	(0.391)
smallres_rec_d			-0.160	-0.160			0.044	-0.204**
			(0.074)	(0.108)			(0.599)	(0.004)
midres_rec_d			-0.013	-0.013			0.117	-0.129*
			(0.859)	(0.882)			(0.174)	(0.040)
gend_static_d			0.061	0.061			0.082	-0.021
			(0.356)	(0.515)			(0.342)	(0.741)
educ_static_d			0.021	0.021			0.006	0.014
			(0.780)	(0.805)			(0.938)	(0.811)
smallres_static_d			-0.038	-0.038			-0.005	-0.033
			(0.666)	(0.711)			(0.959)	(0.603)
midres_static_d			0.100	0.100			0.081	0.019
			(0.167)	(0.293)			(0.399)	(0.746)
Constant	0.460***	0.460***	0.431**	0.431*	0.283***	0.177***	0.038	0.393*
	(0.000)	(0.000)	(0.006)	(0.020)	(0.000)	(0.000)	(0.802)	(0.013)
Observations	271	271	271	271	271	271	271	271
Adjusted R-squared	-0.000	-0.000	0.008	0.008	0.001	-0.004	0.006	0.062
Cluster		YES		YES			YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A5: Hypothesis 3.1 - Academics' discrimination against workers

VARIABLES	(1) incorr_sent	(2) incorr_sent	(3) incorr_sent	(4) incorr_sent	(5) inc_sent_type1	(6) inc_sent_type2	(7) inc_sent_type1	(8) inc_sent_type2
aca_work_disc	0.254***	0.254**	0.236***	0.236**	0.199***	0.056	0.210**	0.026
	(0.000)	(0.001)	(0.000)	(0.002)	(0.000)	(0.191)	(0.002)	(0.504)
gend_send_d			-0.008	-0.008			0.063	-0.071
			(0.880)	(0.891)			(0.319)	(0.286)
young_send_d			0.166	0.166*			-0.028	0.194**
			(0.054)	(0.045)			(0.724)	(0.009)
midage_send_d			0.201**	0.201**			0.049	0.152*
			(0.003)	(0.005)			(0.515)	(0.019)
smallres_send_d			0.096	0.096			0.038	0.058
			(0.216)	(0.221)			(0.614)	(0.443)
midres_send_d			0.095	0.095			0.046	0.049
			(0.110)	(0.104)			(0.377)	(0.420)
gend_rec_d			0.101	0.101			0.009	0.092
			(0.060)	(0.116)			(0.889)	(0.071)
young_rec_d			-0.015	-0.015			0.064	-0.078
			(0.859)	(0.894)			(0.550)	(0.234)
midage_rec_d			0.097	0.097			0.111	-0.015
			(0.160)	(0.214)			(0.125)	(0.790)
smallres_rec_d			-0.034	-0.034			-0.029	-0.005
			(0.648)	(0.692)			(0.721)	(0.930)
midres_rec_d			0.129*	0.129			0.044	0.085
			(0.038)	(0.079)			(0.554)	(0.109)
gend_static_d			-0.041	-0.041			-0.110	0.069
			(0.460)	(0.524)			(0.081)	(0.187)
young_static_d			-0.145	-0.145			-0.170	0.025
			(0.078)	(0.162)			(0.123)	(0.757)
midage_static_d			-0.009	-0.009			-0.042	0.033
			(0.887)	(0.899)			(0.608)	(0.564)
smallres_static_d			0.006	0.006			0.052	-0.046
			(0.935)	(0.943)			(0.463)	(0.506)
midres_static_d			0.099	0.099			0.040	0.059
			(0.125)	(0.182)			(0.575)	(0.237)
Constant	0.340***	0.340***	0.008	0.008	0.190***	0.150***	0.096	-0.087
	(0.000)	(0.000)	(0.949)	(0.948)	(0.000)	(0.000)	(0.324)	(0.490)
Observations	332	332	332	332	332	332	332	332
Adjusted R-squared	0.061	0.061	0.110	0.110	0.043	0.002	0.061	0.042
Cluster		YES		YES			YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A6: Hypothesis 3.2 - Workers' discrimination against workers

VARIABLES	(1) incorr_sent	(2) incorr_sent	(3) incorr_sent	(4) incorr_sent	(5) inc_sent_type1	(6) inc_sent_type2	(7) inc_sent_type1	(8) inc_sent_type2
work_work_disc	-0.175*	-0.175	-0.174	-0.174	-0.128	-0.047	-0.158	-0.015
	(0.018)	(0.178)	(0.059)	(0.176)	(0.068)	(0.407)	(0.206)	(0.826)
gend_send_d			-0.159*	-0.159*			-0.084	-0.075
			(0.041)	(0.041)			(0.285)	(0.323)
young_send_d			0.185	0.185			0.079	0.106
			(0.125)	(0.102)			(0.436)	(0.430)
midage_send_d			0.187	0.187			0.111	0.076
			(0.068)	(0.080)			(0.255)	(0.399)
smallres_send_d			-0.020	-0.020			0.066	-0.086
			(0.855)	(0.868)			(0.518)	(0.429)
midres_send_d			0.083	0.083			-0.012	0.095
			(0.327)	(0.347)			(0.901)	(0.266)
gend_rec_d			0.001	0.001			-0.040	0.042
			(0.984)	(0.990)			(0.686)	(0.584)
young_rec_d			-0.031	-0.031			-0.048	0.016
			(0.790)	(0.827)			(0.739)	(0.890)
midage_rec_d			-0.087	-0.087			-0.024	-0.064
			(0.397)	(0.489)			(0.860)	(0.399)
smallres_rec_d			-0.020	-0.020			0.096	-0.115
			(0.875)	(0.882)			(0.396)	(0.341)
midres_rec_d			0.047	0.047			0.149	-0.103
			(0.592)	(0.599)			(0.068)	(0.143)
gend_static_d			0.196**	0.196			0.175	0.022
			(0.009)	(0.112)			(0.119)	(0.760)
young_static_d			-0.109	-0.109			-0.171	0.061
			(0.355)	(0.485)			(0.212)	(0.681)
midage_static_d			-0.271**	-0.271*			-0.187	-0.084
			(0.002)	(0.014)			(0.073)	(0.299)
smallres_static_d			0.039	0.039			-0.003	0.042
			(0.712)	(0.751)			(0.983)	(0.671)
midres_static_d			-0.079	-0.079			-0.096	0.017
			(0.364)	(0.384)			(0.329)	(0.833)
Constant	0.616***	0.616***	0.655***	0.655**	0.411***	0.205***	0.434	0.221
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.069)	(0.215)
Observations	193	193	193	193	193	193	193	193
Adjusted R-squared	0.024	0.024	0.057	0.057	0.012	-0.002	0.030	0.002
Cluster		YES		YES			YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A7: Hypothesis 4.1 - The discrimination by people from small/mid cities against people from large cities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	incorr_sent	incorr_sent	incorr_sent	incorr_sent	inc_sent_type1	inc_sent_type2	inc_sent_type1	inc_sent_type2
smamid_larg_disc	0.026	0.026	0.034	0.034	-0.028	0.054	-0.033	0.067
	(0.640)	(0.752)	(0.558)	(0.668)	(0.582)	(0.215)	(0.651)	(0.163)
gend_send_d			-0.100	-0.100			-0.059	-0.041
			(0.081)	(0.084)			(0.285)	(0.509)
educ_send_d			-0.009	-0.009			-0.018	0.009
			(0.881)	(0.867)			(0.739)	(0.885)
young_send_d			-0.010	-0.010			-0.077	0.066
			(0.915)	(0.915)			(0.383)	(0.510)
midage_send_d			-0.014	-0.014			-0.098	0.084
			(0.860)	(0.863)			(0.177)	(0.277)
gend_rec_d			0.107	0.107			0.029	0.078
			(0.082)	(0.143)			(0.670)	(0.210)
educ_rec_d			-0.136*	-0.136*			-0.107	-0.029
			(0.028)	(0.045)			(0.122)	(0.614)
young_rec_d			0.002	0.002			0.080	-0.079
			(0.986)	(0.989)			(0.474)	(0.391)
midage_rec_d			0.038	0.038			0.117	-0.080
			(0.645)	(0.699)			(0.174)	(0.279)
gend_static_d			0.027	0.027			-0.097	0.124*
			(0.666)	(0.710)			(0.160)	(0.045)
educ_static_d			0.056	0.056			0.027	0.029
			(0.396)	(0.463)			(0.676)	(0.578)
young_static_d			-0.034	-0.034			0.036	-0.070
			(0.731)	(0.771)			(0.778)	(0.459)
midage_static_d			-0.033	-0.033			-0.026	-0.006
			(0.682)	(0.738)			(0.776)	(0.929)
Constant	0.468***	0.468***	0.512***	0.512***	0.308***	0.160***	0.445***	0.067
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.631)
Observations	324	324	324	324	324	324	324	324
Adjusted R-squared	-0.002	-0.002	0.008	0.008	-0.002	0.002	0.004	0.012
Cluster		YES		YES			YES	YES

Robust p-value in parentheses]

*** p< 0.001,** p< 0.01, * p< 0.05

Table A8: Hypothesis 4.2 - The discrimination by people from mid/large cities against people from small cities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	incorr_sent	incorr_sent	incorr_sent	incorr_sent	inc_sent_type1	inc_sent_type2	inc_sent_type1	inc_sent_type2
midlarg_sma_disc	-0.007	-0.007	0.010	0.010	0.074	-0.080	0.116	-0.106*
	(0.910)	(0.943)	(0.871)	(0.906)	(0.164)	(0.096)	(0.154)	(0.034)
gend_send_d			-0.050	-0.050			-0.102	0.052
			(0.415)	(0.414)			(0.065)	(0.386)
educ_send_d			-0.117	-0.117			-0.128	0.010
			(0.078)	(0.096)			(0.061)	(0.898)
young_send_d			-0.073	-0.073			0.065	-0.138
			(0.482)	(0.527)			(0.495)	(0.312)
midage_send_d			-0.022	-0.022			0.060	-0.082
			(0.799)	(0.803)			(0.497)	(0.450)
gend_rec_d			0.103	0.103			0.067	0.036
			(0.090)	(0.163)			(0.312)	(0.581)
educ_rec_d			-0.035	-0.035			0.051	-0.086
			(0.567)	(0.660)			(0.498)	(0.199)
young_rec_d			-0.234**	-0.234			-0.166	-0.068
			(0.009)	(0.091)			(0.126)	(0.376)
midage_rec_d			-0.102	-0.102			-0.122	0.019
			(0.194)	(0.354)			(0.202)	(0.778)
gend_static_d			-0.140*	-0.140			-0.047	-0.093
			(0.026)	(0.074)			(0.459)	(0.126)
educ_static_d			0.032	0.032			-0.026	0.058
			(0.616)	(0.697)			(0.712)	(0.393)
young_static_d			-0.081	-0.081			-0.143	0.063
			(0.418)	(0.572)			(0.217)	(0.370)
midage_static_d			0.167*	0.167			0.010	0.157**
			(0.033)	(0.142)			(0.920)	(0.007)
Constant	0.482***	0.482***	0.658***	0.658***	0.234***	0.248***	0.413**	0.245
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)	(0.139)
Observations	284	284	284	284	284	284	284	284
Adjusted R-squared	-0.004	-0.004	0.059	0.059	0.003	0.006	0.040	0.039
Cluster		YES		YES			YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A9: Robustness check: logistic regressions for discrimination hypotheses

VARIABLES	(1) incorr_sent	(2) incorr_sent	(3) incorr_sent	(4) incorr_sent	(5) incorr_sent	(6) incorr_sent	(7) incorr_sent	(8) incorr_sent
fem_mal_disc	0.370							
	(0.410)							
mal_fem_disc		-0.555						
		(0.095)						
aca_work_disc			1.059**					
			(0.002)					
work_work_disc				-0.798				
				(0.177)				
midlarg_sma_disc					0.041			
					(0.914)			
smamid_larg_disc						0.148		
						(0.654)		
youmid_old_disc							0.755	
							(0.084)	
midold_you_disc								0.143
								(0.725)
gend_send_d			-0.036	-0.727*	-0.239	-0.422	-0.515	-0.105
			(0.898)	(0.039)	(0.366)	(0.080)	(0.077)	(0.736)
educ_send_d	-0.203	0.322			-0.529	-0.033	0.221	-0.356
	(0.531)	(0.171)			(0.079)	(0.877)	(0.418)	(0.213)
young_send_d	0.395	-0.843*	0.840*	0.836	-0.345	-0.048		
	(0.405)	(0.028)	(0.031)	(0.098)	(0.493)	(0.905)		
midage_send_d	0.181	-0.244	0.997**	0.869	-0.102	-0.060		
	(0.700)	(0.355)	(0.003)	(0.065)	(0.793)	(0.859)		
smallres_send_d	-0.744	-0.155	0.453	-0.060			-0.343	-0.232
	(0.077)	(0.640)	(0.210)	(0.911)			(0.349)	(0.454)
midres_send_d	-0.455	0.406	0.452	0.410			-0.422	0.336
	(0.260)	(0.110)	(0.097)	(0.303)			(0.254)	(0.289)
gend_rec_d			0.473	0.022	0.453	0.444	-0.407	0.364
			(0.103)	(0.963)	(0.161)	(0.134)	(0.238)	(0.303)
educ_rec_d	-0.500	-0.104			-0.187	-0.567*	-0.032	0.067
	(0.150)	(0.710)			(0.595)	(0.040)	(0.913)	(0.855)
young_rec_d	-0.148	0.262	-0.096	-0.170	-1.078	0.005		
	(0.789)	(0.578)	(0.855)	(0.789)	(0.084)	(0.992)		
midage_rec_d	-0.334	0.488	0.454	-0.401	-0.457	0.161		
	(0.485)	(0.191)	(0.203)	(0.458)	(0.340)	(0.689)		
smallres_rec_d	-0.287	-0.035	-0.135	-0.108			-0.077	-0.675
	(0.614)	(0.921)	(0.728)	(0.851)			(0.857)	(0.102)
midres_rec_d	0.159	0.130	0.611	0.203			0.238	-0.056
	(0.691)	(0.666)	(0.075)	(0.600)			(0.539)	(0.874)
gend_static_d			-0.191	0.881	-0.617	0.116	-0.074	0.253
			(0.526)	(0.118)	(0.063)	(0.698)	(0.819)	(0.512)
educ_static_d	0.752*	0.214			0.136	0.235	-0.188	0.087
	(0.032)	(0.500)			(0.700)	(0.453)	(0.590)	(0.802)
young_static_d	-0.503	-0.253	-0.671	-0.496	-0.420	-0.147		
	(0.432)	(0.605)	(0.168)	(0.478)	(0.518)	(0.760)		
midage_static_d	-0.338	-0.541	-0.026	-1.255*	0.733	-0.139		
	(0.585)	(0.141)	(0.936)	(0.017)	(0.150)	(0.732)		
smallres_static_d	-0.070	-0.132	0.040	0.155			-0.547	-0.163
	(0.900)	(0.724)	(0.916)	(0.781)			(0.255)	(0.702)
midres_static_d	0.366	0.186	0.465	-0.361			0.084	0.419
	(0.383)	(0.542)	(0.166)	(0.375)			(0.831)	(0.285)
Constant	0.425	-0.190	-2.356***	0.703	0.758	0.045	0.894	-0.288
	(0.659)	(0.721)	(0.000)	(0.442)	(0.222)	(0.928)	(0.236)	(0.701)
Observations	183	385	332	193	284	324	263	271
Cluster	YES	YES	YES	YES	YES	YES	YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A10: Robustness check: classification errors for discrimination hypotheses

VARIABLES	(1) incorr_sent	(2) incorr_sent	(3) incorr_sent	(4) incorr_sent	(5) incorr_sent	(6) incorr_sent	(7) incorr_sent	(8) incorr_sent
fem_mal_disc	0.108							
	(0.330)							
mal_fem_disc		-0.148						
		(0.075)						
aca_work_disc			0.184**					
			(0.008)					
work_work_disc				-0.123				
				(0.355)				
midlarg_sma_disc					0.006			
					(0.942)			
smamid_larg_disc						0.034		
						(0.674)		
youmid_old_disc							0.132	
							(0.213)	
midold_you_disc								0.056
								(0.564)
gend_send_d			0.072	-0.177*	-0.047	-0.095	-0.099	0.001
			(0.283)	(0.030)	(0.466)	(0.106)	(0.118)	(0.987)
educ_send_d	-0.090	0.056			-0.099	0.002	0.086	-0.100
	(0.236)	(0.308)			(0.175)	(0.975)	(0.172)	(0.159)
young_send_d	0.137	-0.153	0.175*	0.224	-0.064	-0.003		
	(0.227)	(0.081)	(0.048)	(0.069)	(0.562)	(0.971)		
midage_send_d	0.042	-0.060	0.139	0.185	-0.029	-0.032		
	(0.702)	(0.353)	(0.055)	(0.108)	(0.746)	(0.706)		
smallres_send_d	-0.188	-0.027	0.032	-0.012			-0.116	-0.029
	(0.052)	(0.728)	(0.721)	(0.925)			(0.147)	(0.684)
midres_send_d	-0.100	0.097	0.055	0.138			-0.090	0.094
	(0.295)	(0.121)	(0.415)	(0.150)			(0.286)	(0.257)
gend_rec_d			0.063	-0.029	0.109	0.103	-0.104	0.110
			(0.363)	(0.795)	(0.152)	(0.159)	(0.193)	(0.215)
educ_rec_d	-0.084	-0.040			-0.040	-0.074	0.064	0.011
	(0.343)	(0.555)			(0.610)	(0.302)	(0.412)	(0.908)
young_rec_d	-0.017	0.057	-0.034	-0.065	-0.193	-0.040		
	(0.905)	(0.613)	(0.774)	(0.667)	(0.155)	(0.756)		
midage_rec_d	-0.097	0.047	0.091	-0.126	-0.123	0.026		
	(0.434)	(0.565)	(0.301)	(0.349)	(0.253)	(0.791)		
smallres_rec_d	-0.058	-0.017	-0.030	-0.166			-0.009	-0.171
	(0.670)	(0.837)	(0.745)	(0.244)			(0.936)	(0.096)
midres_rec_d	0.063	0.004	0.166*	-0.007			0.077	-0.038
	(0.519)	(0.956)	(0.036)	(0.931)			(0.425)	(0.668)
gend_static_d			-0.045	0.164	-0.166*	0.028	-0.066	0.076
			(0.527)	(0.181)	(0.043)	(0.699)	(0.387)	(0.438)
educ_static_d	0.142	0.037			0.018	0.134	0.027	0.034
	(0.107)	(0.611)			(0.834)	(0.076)	(0.757)	(0.685)
young_static_d	-0.062	-0.013	-0.063	-0.037	-0.059	-0.073		
	(0.686)	(0.910)	(0.586)	(0.816)	(0.686)	(0.549)		
midage_static_d	-0.110	-0.070	0.009	-0.205	0.171	-0.080		
	(0.456)	(0.410)	(0.912)	(0.076)	(0.116)	(0.421)		
smallres_static_d	-0.004	-0.036	0.086	0.078			-0.046	0.009
	(0.975)	(0.689)	(0.371)	(0.584)			(0.690)	(0.928)
midres_static_d	0.081	0.046	0.112	-0.085			0.085	0.079
	(0.439)	(0.528)	(0.145)	(0.367)			(0.326)	(0.414)
Constant	0.591*	0.493***	-0.000	0.598**	0.664***	0.463***	0.551**	0.364
	(0.012)	(0.000)	(0.999)	(0.004)	(0.000)	(0.000)	(0.002)	(0.052)
Observations	183	385	314	196	284	324	267	263
R-squared	0.104	0.053	0.115	0.109	0.100	0.051	0.064	0.060
Cluster	YES	YES	YES	YES	YES	YES	YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A11: Hypotheses 5.1 and 5.2 - Male behavior in groups with various gender composition

VARIABLES	(1) kept_male	(2) kept_male	(3) kept_male	(4) kept_male
gend_mmm	0.066***	0.045**	0.066**	0.045*
	(0.000)	(0.006)	(0.003)	(0.035)
gend_mff	-0.014	-0.018	-0.014	-0.018
	(0.503)	(0.403)	(0.631)	(0.490)
age_yym		-0.021		-0.021
		(0.721)		(0.737)
age_yyo		-0.034		-0.034
		(0.791)		(0.822)
age_ymm		-0.009		-0.009
		(0.847)		(0.874)
age_ymo		-0.003		-0.003
		(0.955)		(0.959)
age_yoo		-0.043		-0.043
		(0.756)		(0.659)
age_mmy		0.035		0.035
		(0.437)		(0.490)
age_myy		0.074		0.074
		(0.189)		(0.267)
age_mmm		0.074		0.074
		(0.078)		(0.133)
age_mmo		0.080		0.080
		(0.063)		(0.113)
age_myo		-0.018		-0.018
		(0.736)		(0.797)
age_moo		0.001		0.001
		(0.985)		(0.987)
age_ooy		0.181*		0.181
		(0.015)		(0.078)
age_oyy		-0.364		-0.364***
		(0.193)		(0.000)
age_ooo		0.001		0.001
		(0.994)		(0.995)
age_oom		-0.046		-0.046
		(0.401)		(0.513)
age_oym		0.012		0.012
		(0.828)		(0.864)
age_omm		0.026		0.026
		(0.615)		(0.654)
res_ssm		-0.039		-0.039
		(0.521)		(0.626)
res_ssb		0.070		0.070
		(0.225)		(0.371)
res_smm		-0.026		-0.026
		(0.732)		(0.759)
res_smb		-0.077		-0.077
		(0.214)		(0.383)
res_sbb		0.009		0.009
		(0.883)		(0.916)
res_mms		-0.088		-0.088
		(0.139)		(0.294)
res_mss		0.056		0.056
		(0.435)		(0.498)
res_mmm		0.008		0.008
		(0.884)		(0.924)
res_mmb		0.010		0.010
		(0.860)		(0.896)
res_msb		0.025		0.025
		(0.656)		(0.744)
res_mbb		-0.005		-0.005
		(0.928)		(0.947)
res_bbs		-0.001		-0.001
		(0.984)		(0.987)
res_bss		-0.041		-0.041
		(0.574)		(0.648)
res_bbb		0.052		0.052
		(0.349)		(0.483)
res_bbm		-0.000		-0.000
		(0.994)		(0.996)
res_bsm		0.037		0.037
		(0.500)		(0.635)
res_bmm		0.042		0.042
		(0.451)		(0.588)
educ_aww		0.013		0.013
		(0.625)		(0.671)
educ_wwa		-0.036		-0.036
		(0.164)		(0.294)
educ_waa		-0.044		-0.044
		(0.077)		(0.128)
educ_aaw		-0.019		-0.019
		(0.296)		(0.434)
educ_www		-0.021		-0.021
		(0.675)		(0.778)
Constant	0.774***	0.760***	0.774***	0.760***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3,293	3,293	3,293	3,293
Adjusted R-squared	0.006	0.020	0.006	0.020
Cluster			YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A12: Hypotheses 5.3 and 5.4 - Female behavior in groups with various gender composition

VARIABLES	(1) kept_female	(2) kept_female	(3) kept_female	(4) kept_female
gend_fff	-0.115**	-0.124**	-0.115**	-0.124**
	(0.001)	(0.002)	(0.002)	(0.002)
gend_fmm	0.007	-0.017	0.007	-0.017
	(0.736)	(0.483)	(0.805)	(0.567)
age_yym		0.001		0.001
		(0.989)		(0.988)
age_yyo		0.296**		0.296**
		(0.002)		(0.001)
age_ymm		-0.009		-0.009
		(0.910)		(0.877)
age_ymo		-0.011		-0.011
		(0.909)		(0.921)
age_yoo		0.054		0.054
		(0.739)		(0.597)
age_mmy		0.109		0.109*
		(0.125)		(0.032)
age_myy		0.186*		0.186*
		(0.037)		(0.037)
age_mmm		0.184**		0.184***
		(0.009)		(0.000)
age_mmo		0.118		0.118*
		(0.106)		(0.037)
age_myo		0.177*		0.177**
		(0.049)		(0.001)
age_moo		0.186*		0.186***
		(0.033)		(0.001)
age_ooy		0.238**		0.238***
		(0.003)		(0.000)
age_oyy		0.237*		0.237
		(0.043)		(0.062)
age_ooo		0.140		0.140
		(0.233)		(0.130)
age_oom		0.070		0.070
		(0.411)		(0.373)
age_oym		0.261***		0.261***
		(0.001)		(0.000)
age_omm		0.146		0.146*
		(0.063)		(0.017)
res_ssm		-0.007		-0.007
		(0.936)		(0.932)
res_ssb		0.027		0.027
		(0.775)		(0.772)
res_smm		-0.023		-0.023
		(0.815)		(0.842)
res_smb		-0.047		-0.047
		(0.586)		(0.544)
res_sbb		-0.085		-0.085
		(0.450)		(0.349)
res_mms		-0.068		-0.068
		(0.473)		(0.421)
res_mss		-0.035		-0.035
		(0.733)		(0.718)
res_mmm		0.076		0.076
		(0.385)		(0.371)
res_mmb		0.003		0.003
		(0.968)		(0.967)
res_msb		-0.056		-0.056
		(0.514)		(0.499)
res_mbb		0.013		0.013
		(0.889)		(0.894)
res_bbs		0.047		0.047
		(0.591)		(0.561)
res_bss		-0.062		-0.062
		(0.655)		(0.553)
res_bbb		0.174*		0.174*
		(0.039)		(0.032)
res_bbm		0.077		0.077
		(0.350)		(0.354)
res_bsm		0.071		0.071
		(0.416)		(0.401)
res_bmm		0.008		0.008
		(0.925)		(0.919)
educ_aww		-0.095*		-0.095*
		(0.045)		(0.030)
educ_wwa		-0.049		-0.049
		(0.148)		(0.238)
educ_waa		-0.084*		-0.084*
		(0.012)		(0.036)
educ_aaw		-0.025		-0.025
		(0.380)		(0.454)
educ_www		-0.028		-0.028
		(0.687)		(0.572)
Constant	0.784***	0.695***	0.784***	0.695***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1,810	1,810	1,810	1,810
Adjusted R-squared	0.007	0.036	0.007	0.036
Cluster			YES	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A13: Robustness check: logistic regression for male group behavior hypotheses

VARIABLES	(1) kept_male
gend_mmm	0.302*
	(0.039)
gend_mff	-0.099
	(0.515)
age_yym	-0.120
	(0.728)
age_yyo	-0.206
	(0.784)
age_ymm	-0.034
	(0.919)
age_ymo	-0.016
	(0.964)
age_yoo	-0.291
	(0.581)
age_mmy	0.198
	(0.502)
age_myy	0.473
	(0.281)
age_mmm	0.477
	(0.109)
age_mmo	0.518
	(0.097)
age_myo	-0.054
	(0.886)
age_moo	0.030
	(0.933)
age_ooy	1.326
	(0.206)
age_oyy	-1.560***
	(0.000)
age_ooo	-0.010
	(0.988)
age_oom	-0.222
	(0.550)
age_oym	0.090
	(0.829)
age_omm	0.155
	(0.638)
res_ssm	-0.258
	(0.617)
res_ssb	0.476
	(0.389)
res_smm	-0.168
	(0.752)
res_smb	-0.434
	(0.423)
res_sbb	0.041
	(0.943)
res_mms	-0.474
	(0.362)
res_mss	0.411
	(0.468)
res_mmm	0.047
	(0.932)
res_mmb	0.033
	(0.945)
res_msb	0.130
	(0.798)
res_mbb	-0.041
	(0.936)
res_bbs	-0.036
	(0.944)
res_bss	-0.244
	(0.656)
res_bbb	0.370
	(0.466)
res_bbm	-0.013
	(0.979)
res_bsm	0.217
	(0.679)
res_bmm	0.258
	(0.621)
educ_aww	0.063
	(0.765)
educ_wwa	-0.214
	(0.304)
educ_waa	-0.265
	(0.137)
educ_aaw	-0.123
	(0.452)
educ_www	-0.145
	(0.754)
Constant	1.162*
	(0.032)
Observations	3,293
Cluster	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A14: Robustness check: logistic regression for female group behavior hypotheses

VARIABLES	(1) kept_female
gend_fff	-0.698***
	(0.001)
gend_fmm	-0.134
	(0.456)
age_yym	-0.027
	(0.938)
age_ymm	-0.100
	(0.745)
age_ymo	0.008
	(0.988)
age_yoo	0.202
	(0.737)
age_mmy	0.565*
	(0.033)
age_myy	1.114
	(0.092)
age_mmm	1.049***
	(0.000)
age_mmo	0.638*
	(0.030)
age_myo	1.007**
	(0.002)
age_moo	1.101***
	(0.001)
age_ooy	1.343***
	(0.000)
age_oyy	1.566
	(0.063)
age_ooo	0.774
	(0.266)
age_oom	0.378
	(0.333)
age_oym	1.673***
	(0.000)
age_omm	0.782*
	(0.020)
res_ssm	-0.032
	(0.949)
res_ssb	0.227
	(0.704)
res_smm	-0.107
	(0.869)
res_smb	-0.217
	(0.625)
res_sbb	-0.418
	(0.407)
res_mms	-0.373
	(0.450)
res_mss	-0.108
	(0.838)
res_mmm	0.477
	(0.352)
res_mmb	0.030
	(0.950)
res_msb	-0.306
	(0.524)
res_mbb	0.097
	(0.864)
res_bbs	0.306
	(0.535)
res_bss	-0.382
	(0.541)
res_bbb	1.653**
	(0.006)
res_bbm	0.505
	(0.321)
res_bsm	0.468
	(0.385)
res_bmm	0.082
	(0.867)
educ_aww	-0.609*
	(0.014)
educ_wwa	-0.278
	(0.265)
educ_waa	-0.475*
	(0.036)
educ_aaw	-0.138
	(0.516)
educ_www	-0.181
	(0.563)
Constant	0.848
	(0.089)
Observations	1,808
Cluster	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A15: Robustness check: classification errors for male group behavior hypotheses

VARIABLES	(1) kept_male
gend_mmm	0.045*
	(0.034)
gend_mff	-0.022
	(0.423)
age_yym	-0.020
	(0.758)
age_yyo	-0.038
	(0.807)
age_ymm	-0.010
	(0.879)
age_ymo	0.002
	(0.981)
age_yoo	0.078
	(0.623)
age_mmy	0.034
	(0.548)
age_myy	0.087
	(0.198)
age_mmm	0.066
	(0.236)
age_mmo	0.079
	(0.164)
age_myo	-0.001
	(0.992)
age_moo	-0.037
	(0.563)
age_ooy	0.080
	(0.542)
age_oyy	-0.367***
	(0.000)
age_ooo	0.004
	(0.976)
age_oom	-0.045
	(0.523)
age_oym	0.046
	(0.528)
age_omm	0.035
	(0.585)
res_ssm	-0.037
	(0.642)
res_ssb	0.077
	(0.331)
res_smm	-0.026
	(0.762)
res_smb	-0.078
	(0.375)
res_sbb	0.016
	(0.848)
res_mms	-0.091
	(0.276)
res_mss	0.040
	(0.629)
res_mmm	0.001
	(0.987)
res_mmb	0.008
	(0.916)
res_msb	0.020
	(0.796)
res_mbb	-0.010
	(0.895)
res_bbs	0.001
	(0.994)
res_bss	-0.031
	(0.728)
res_bbb	0.056
	(0.452)
res_bbm	-0.002
	(0.978)
res_bsm	0.023
	(0.771)
res_bmm	0.042
	(0.581)
educ_aww	0.016
	(0.613)
educ_wwa	-0.039
	(0.247)
educ_waa	-0.053
	(0.072)
educ_aaw	-0.012
	(0.655)
educ_www	-0.007
	(0.913)
Constant	0.759***
	(0.000)
Observations	3,293
Adjusted R-squared	0.018
Cluster	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A16: Robustness check: classification errors for female group behavior hypotheses

VARIABLES	(1) kept_female
gend_fff	-0.114**
	(0.007)
gend_fmm	-0.012
	(0.696)
age_yym	0.011
	(0.893)
age_yyo	0.303**
	(0.001)
age_ymm	-0.019
	(0.773)
age_ymo	0.009
	(0.936)
age_yoo	0.045
	(0.674)
age_mmy	0.095
	(0.086)
age_myy	0.117
	(0.174)
age_mmm	0.162**
	(0.002)
age_mmo	0.109
	(0.078)
age_myo	0.148*
	(0.014)
age_moo	0.113
	(0.181)
age_ooy	0.243***
	(0.001)
age_oyy	0.232
	(0.089)
age_ooo	0.125
	(0.180)
age_oom	0.087
	(0.291)
age_oym	0.226***
	(0.000)
age_omm	0.119
	(0.065)
res_ssm	-0.018
	(0.831)
res_ssb	0.008
	(0.933)
res_smm	-0.046
	(0.700)
res_smb	-0.060
	(0.434)
res_sbb	-0.052
	(0.603)
res_mms	-0.088
	(0.298)
res_mss	-0.066
	(0.483)
res_mmm	0.067
	(0.434)
res_mmb	-0.006
	(0.937)
res_msb	-0.064
	(0.433)
res_mbb	-0.007
	(0.938)
res_bbs	0.041
	(0.597)
res_bss	-0.051
	(0.621)
res_bbb	0.158*
	(0.045)
res_bbm	0.047
	(0.568)
res_bsm	0.055
	(0.512)
res_bmm	-0.006
	(0.942)
educ_aww	-0.061
	(0.153)
educ_wwa	-0.043
	(0.323)
educ_waa	-0.086*
	(0.041)
educ_aaw	0.005
	(0.877)
educ_www	0.002
	(0.972)
Constant	0.706***
	(0.000)
Observations	1,810
Adjusted R-squared	0.031
Cluster	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

Table A17: Definitions of dependent variables for discrimination hypotheses

Variable name	Values and conditions
fem_male_disc	0 iff female sender, male receiver, female static 1 iff female sender, female receiver, male static
mal_fem_disc	0 iff male sender, male receiver, female static 1 iff male sender, female receiver, male static
youmid_old_disc	0 iff young/middle sender, old receiver, young/middle static 1 iff young/middle sender, young/middle receiver, old static
midold_you_disc	0 iff old/middle sender, old/middle receiver, young static 1 iff old/middle sender, young receiver, old/middle static
aca_work_disc	0 iff academic sender, academic receiver, worker static 1 iff academic sender, worker receiver, academic static
work_work_disc	0 iff worker sender, academic receiver, worker static 1 iff worker sender, worker receiver, academic static
midlarg_sma_disc	0 iff large/middle sender, large/middle receiver, small static 1 iff arge/middle sender, small receiver, large/middle static
smamid_larg_disc	0 iff small/middle sender, small/middle receiver, large static 1 iff small/middle sender, large receiver, small/middle static

For each variable, all other observations not covered in the definition, both closed and open questions, are left out as missing values

Table A18: Definitions of control variables for discrimination hypotheses

Variable name	Takes value 1 if and only if:	
	Player role	Opponents
gend_send_d	Sender	Male
educ_send_d	Sender	Academic
young_send_d	Sender	Young
midage_send_d	Sender	Middle-aged
old_send_d	Sender	Old
smallres_send_d	Sender	From a small town
midres_send_d	Sender	From a medium town
bigres_send_d	Sender	From a large city
gend_rec_d	Receiver	Male
educ_rec_d	Receiver	Academic
young_rec_d	Receiver	Young
midage_rec_d	Receiver	Middle-aged
old_rec_d	Receiver	Old
smallres_rec_d	Receiver	From a small town
midres_rec_d	Receiver	From a medium town
bigres_rec_d	Receiver	From a large city
gend_static_d	Static	Male
educ_static_d	Static	Academic
young_static_d	Static	Young
midage_static_d	Static	Middle-aged
old_static_d	Static	Old
smallres_static_d	Static	From a small town
midres_static_d	Static	From a medium town
bigres_static_d	Static	From a large city

All other observations for each variable have value of 0, unless they are open questions. As no decision is being made in open questions, they are classified as missing values.

Table A19: Definitions of variables for group behavior hypotheses

Variable name	Variable group	Takes value 1 if and only if:	
		Decision maker	Opponents
gend_fff	Gender	Female	Female, Female
gend_fmm	Gender	Female	Male, Male
gend_ffm	Gender	Female	Female, Male
gend_mmm	Gender	Male	Male, Male
gend_mff	Gender	Male	Female, Female
gend_mmf	Gender	Male	Male, Female
age_yyy	Age	Young	Young, Young
age_yym	Age	Young	Young, Middle-aged
age_yyo	Age	Young	Young, Old
age_ymm	Age	Young	Middle-aged, Middle-aged
age_ymo	Age	Young	Middle-aged, Old
age_yoo	Age	Young	Old, Old
age_mmy	Age	Middle-aged	Middle-aged, Young
age_myy	Age	Middle-aged	Young, Young
age_mmm	Age	Middle-aged	Middle-aged, Middle-aged
age_mmo	Age	Middle-aged	Middle-aged, Old
age_myo	Age	Middle-aged	Young, Old
age_moo	Age	Middle-aged	Old, Old
age_ooy	Age	Old	Old, Young
age_oyy	Age	Old	Young, Young
age_ooo	Age	Old	Old, Old
age_oom	Age	Old	Old, Middle-aged
age_oym	Age	Old	Young, Middle-aged
age_omm	Age	Old	Middle-aged, Middle-aged
res_sss	Residence	Small town	Small town, Small town
res_ssm	Residence	Small town	Small town, Medium town
res_ssb	Residence	Small town	Small town, Large city
res_smm	Residence	Small town	Medium town, Medium town
res_smb	Residence	Small town	Medium town, Large city
res_sbb	Residence	Small town	Large city, Large city
res_mms	Residence	Medium town	Medium town, Small town
res_mss	Residence	Medium town	Small town, Small town
res_mmm	Residence	Medium town	Medium town, Medium town
res_mmb	Residence	Medium town	Medium town, Large city
res_msb	Residence	Medium town	Small town, Large city
res_mbb	Residence	Medium town	Large city, Large city
res_bbs	Residence	Large city	Large city, Small town
res_bss	Residence	Large city	Small town, Small town
res_bbb	Residence	Large city	Large city, Large city
res_bbm	Residence	Large city	Large city, Medium town
res_bsm	Residence	Large city	Small town, Medium town
res_bmm	Residence	Large city	Medium town, Medium town
educ_aww	Education	Academic	Worker, Worker
educ_wwa	Education	Worker	Worker, Academic
educ_waa	Education	Worker	Academic, Academic
educ_aaw	Education	Academic	Academic, Worker
educ_www	Education	Worker	Worker, Worker
educ_aaa	Education	Academic	Academic, Academic

All other observations for each variable have value of 0, unless they are open questions. As no decision is being made in open questions, they are classified as missing values.

Table A20: Descriptive statistics for variables used in discrimination hypotheses

Variable	Obs	Average	Std. Dev	Min	Max
gend_send_d	1099	0.622	0.485	0	1
educ_send_d	1099	0.596	0.491	0	1
young_send_d	1099	0.234	0.423	0	1
midage_send_d	1099	0.575	0.495	0	1
old_send_d	1099	0.192	0.388	0	1
smallres_send_d	1099	0.242	0.426	0	1
midres_send_d	1099	0.440	0.497	0	1
bigres_send_d	1099	0.318	0.466	0	1
gend_rec_d	1099	0.526	0.500	0	1
educ_rec_d	1099	0.596	0.491	0	1
young_rec_d	1099	0.240	0.426	0	1
midage_rec_d	1099	0.550	0.497	0	1
old_rec_d	1099	0.210	0.403	0	1
smallres_rec_d	1099	0.230	0.418	0	1
midres_rec_d	1099	0.411	0.492	0	1
bigres_rec_d	1099	0.359	0.480	0	1
gend_static_d	1099	0.578	0.494	0	1
educ_static_d	1099	0.678	0.467	0	1
young_static_d	1099	0.200	0.394	0	1
midage_static_d	1099	0.577	0.493	0	1
old_static_d	1099	0.223	0.416	0	1
smallres_static_d	1099	0.231	0.419	0	1
midres_static_d	1099	0.415	0.493	0	1
bigres_static_d	1099	0.354	0.478	0	1
fem_mal_disc	183	0.546	0.499	0	1
mal_fem_disc	385	0.545	0.499	0	1
aca_work_disc	332	0.557	0.497	0	1
work_work_disc	193	0.622	0.486	0	1
midlarg_small_disc	284	0.504	0.501	0	1
smamid_large_disc	324	0.519	0.500	0	1
youmid_old_disc	263	0.506	0.501	0	1
midold_young_disc	271	0.583	0.494	0	1

Observations listed above are closed questions
from the final stage that were sent.

Table A21: Descriptive statistics for variables used in group behavior hypotheses

Variable	Obs	Mean	Std. Dev	Min	Max
kept_male	3293	0.792	0.406	0	1
kept_female	1810	0.773	0.419	0	1
gend_mmm	5103	0.198	0.399	0	1
gend_mff	5103	0.119	0.312	0	1
gend_ffm	5103	0.165	0.371	0	1
gend_fmm	5103	0.145	0.352	0	1
gend_mmf	5103	0.330	0.472	0	1
gend_fff	5103	0.043	0.202	0	1
age_yyy	5103	0.033	0.168	0	1
age_yym	5103	0.038	0.191	0	1
age_yyo	5103	0.009	0.056	0	1
age_ymm	5103	0.069	0.253	0	1
age_ymo	5103	0.035	0.183	0	1
age_yoo	5103	0.009	0.064	0	1
age_mmy	5103	0.139	0.346	0	1
age_myy	5103	0.030	0.161	0	1
age_mmm	5103	0.224	0.428	0	1
age_mmo	5103	0.147	0.354	0	1
age_myo	5103	0.046	0.208	0	1
age_moo	5103	0.037	0.190	0	1
age_ooy	5103	0.022	0.146	0	1
age_oyy	5103	0.008	0.064	0	1
age_ooo	5103	0.009	0.080	0	1
age_oom	5103	0.047	0.211	0	1
age_oym	5103	0.043	0.202	0	1
age_omm	5103	0.055	0.229	0	1
res_sss	5103	0.020	0.121	0	1
res_ssm	5103	0.047	0.211	0	1
res_ssb	5103	0.037	0.188	0	1
res_smm	5103	0.024	0.154	0	1
res_smb	5103	0.062	0.241	0	1
res_sbb	5103	0.031	0.172	0	1
res_mms	5103	0.056	0.231	0	1
res_mss	5103	0.023	0.143	0	1
res_mmm	5103	0.084	0.277	0	1
res_mmb	5103	0.100	0.301	0	1
res_msb	5103	0.078	0.269	0	1
res_mbb	5103	0.066	0.248	0	1
res_bss	5103	0.019	0.129	0	1
res_bbb	5103	0.055	0.229	0	1
res_bbm	5103	0.115	0.319	0	1
res_bsm	5103	0.064	0.246	0	1
res_bmm	5103	0.068	0.252	0	1
educ_aaa	5103	0.243	0.429	0	1
educ_aww	5103	0.081	0.274	0	1
educ_wwa	5103	0.150	0.357	0	1
educ_waa	5103	0.160	0.367	0	1
educ_aaw	5103	0.324	0.468	0	1
educ_www	5103	0.043	0.153	0	1

Observations listed above are closed questions from the final stage.

Table A22: Interaction variable regression regarding gender and sending decisions

VARIABLES	(1) send
samegend	0.127***
	(0.000)
gend_send_d	0.013
	(0.398)
samegend_gend_send_d	-0.171***
	(0.000)
Constant	0.240***
	(0.000)
Observations	5,103
Adjusted R-squared	0.024
Controls	YES

Robust p-value in parentheses
*** p< 0.001,** p< 0.01, * p< 0.05

List A23 - Example questions

Below are ten example questions from the final stage presented. First is the category presented, then the question itself, followed by the correct answer. As the show is presented in Swedish, the original categories and questions are in Swedish. As a result, the English questions presented here are the authors translations. All questions are from season 14 episode 66, aired on the 20th April 2015.

1. Dieter / Diets

   Vilken brittisk Michael är journalist och medicinsk reporter och har tillsammans med Mimi Spencer skrivit boken *5:2 dieten?* / Which British Michael is journalist and medicine reporter and has, together with Mimi Spencer, written the book called *The 5:2 diet?*

   (Answer: Michael Mosley)

2. Hos frisören / At the hair dresser

   Vad kallas, efter franskans ord för ljus, metoden då man bleker håret för att få en ljusare färg? / What is the name, after the French word for blonde, of the method where you bleach the hair to get a lighter color?

   (Answer: Blondera / Bleach)

3. Kanada / Canada

   I vilken kanadensisk provins, där franska är det dominerande språket. ligger staden Montreal? / In which Canadian province, where French is the primary language, is the city Montreal located?

   (Answer: Quebec)

4. I kylskåpet / In the fridge

   Vilken kemisk reaktion motsvarar förruttnelse och innebär att fett eller smör bryts ner så att det luktar och smakar illa? / Which chemical reaction equals decay and implies that fat or butter is catabolized which gives away a disgusting taste and smell?

   (Answer: Härskning / Rancidification)

5. Volym / Volume

   Hur utläses volymmåttet som i matlagningsreceptet ofta förkortas l? / How is the volume measurement that is often abbreviated as l deducted?

   (Answer: Liter)

6. Medeltida städer / Medieval cities

   I vilket landskap finns tätorten Lödöse som på 1100-talet var en viktig hamn? / In which county is Lödöse, which was an important harbor in the 12th century, located?

   (Answer: Västergötland)

7. Förnamn / Forenames

   Vilket finskt kvinnonamn betyder ”den enda” och bars av formgivaren och arkitekten Aalto? / Which Finnish female name means “the only” and was carried by the designer and architect Aalto?

   (Answer: Aino)

8. Julklappar / Christmas gifts

   Vad utsågs till årets julklapp 2013 av Handelns utredningsinstitut: råsaftcentrifug, spikmatta eller glassmaskin? / What was chosen by Swedish Institute of Retail as Christmas gift of the year in 2013: juicer, acupressure mat or ice cream machine?

   (Answer: Råsaftcentrifug / Juicer)

9. Organisationer / Organizations

   För vilka två engelska ord står UN i namnet på flyktingorganisationen UNHCR? / What two English words does UN stand for in the name of the refugee agency UNHCR?

   (Answer: United Nations)

10. Nobelpristagare / Nobel prize winners

    Vilken amerikan har skrivit *”Den gamle och havet”* och *”Farväl till vapnen”*? / Which American has written *“The old man and the sea”* and *“Farewell to arms”*?

    (Answer: Ernest Hemingway)