Erasmus School of Economics

Master Thesis Behavioural Economics

Gender bias and the Ultimatum Game: do agentic women face a backlash?

Name author: Dirk van den Beuken
Student ID number: 430613

Supervisor: Dr. Marine Hainguerlot

Second assessor: Dr. Tong Wang

Date final version: 25-07-2021

Abstract

This paper examines the effect of implicit and explicit gender bias on negotiation behaviour, which is important in career advancement and helps to explain the glass ceiling effect and the gender wage gap. Using the Ultimatum Game, the Implicit Association Test and the Gender Stereotype Index, this research isolates the backlash effect and tests for a correlation with gender bias to explain why women obtain worse outcomes in negotiations. Results show a positive correlation between the implicit gender associations and the rejection levels in the Ultimatum Game in games with female proposers. This finding suggests that individuals who associate females with "communal" terms administer backlash to agentic female proposers, even at personal costs. The explicit gender associations showed no significant correlation with the rejection levels. The results should be interpreted with caution due to several limitations, arising from the data collection, design choices and limited explanatory power of the models.

Keywords: Gender bias, Negotiation, Implicit associations, Explicit associations, Backlash effect

List of Tables

List of Figures

Acknowledgements

Compliance statements

Non-plagiarism statement

By submitting this thesis, the author declares to have written this thesis completely by himself, and not to have used sources or resources other than the ones mentioned. All sources used, quotes and citations that were literally taken from publications, or that were in close accordance with the meaning of those publications, are indicated as such.

Copyright statement

Contents

1 Introduction

1.1 Research question

This paper studies the effect of implicit and explicit gender associations on negotiation behaviour. Previous studies often indicate that women obtain worse outcomes in negotiations than men. The current research focuses on the cause of these outcomes. A central idea is the so-called "backlash effect". Backlash refers to the economic or social punishment of a person acting in a stereotype-incongruent way by another person. Specifically, the current study examines if gender bias can predict the backlash effect that is administered to women acting in a stereotype-incongruent manner.

Gender stereotypes in negotiations align with the broader stereotypes on agentic and communal behaviour (Bowles, 2013), in which men are often seen as agentic and women as communal (A. Eagly, 1987; Rudman & Glick, 2001; A. Eagly & Wood, 2012). Men are expected to be competitive and self-interested, while women on the other hand are expected to be cooperative and concerned about others (Kray & Thompson, 2004; C. Eckel, De Oliveira, & Grossman, 2008; A. Eagly & Wood, 2012). These stereotypes can be problematic since the agentic and masculine traits are often associated with being a good negotiator, and good negotiation skills can be important for salary discussions and general career advancement. In this way, the term negotiation itself is not gender-neutral (Small, Gelfand, Babcock, & Gettman, 2007). The current research examines if women experience economic backlash when they behave agentic in negotiations.

In order to capture the backlash effect rather than the overall negotiation outcome, the current study must be able to identify if stronger stereotypes or gender associations correlate with more backlash behaviour. For this purpose, I include measurements of gender associations in this study. A distinction can be made between implicit and explicit associations. Explicit associations are conscious and deliberate associations between multiple things or groups, these are expressed explicitly. Alternatively, implicit associations refer to a person's internal or even subconscious associations. Implicit and explicit associations may not always align, either because individuals do not want to share their (socially undesirable) opinions or because they are not aware of these themselves (Fazio, Jackson, Dunton, & Williams, 1995). The topic of implicit associations is particularly interesting because it tests if we can use normally unobserved bias to predict real negotiation behaviour. Including both implicit and explicit association tests in the analysis offers the opportunity to gain more insights into what factors drive the behaviour in a negotiation

setting. This is relevant, because different causes of gender-biased behaviour require different approaches to address the issue.

If the study finds that gender bias predicts higher rejection levels to offers from women, this would be an indication that worse outcomes by women in negotiations are not (solely) caused by their own limited skills or by their lack of agentic behaviour, but rather by being treated differently when they do not act according to the communal stereotype. Therefore, the goal of the current paper is to answer the following research question:

What is the influence of implicit and explicit gender bias on rejection levels in negotiations?

The current research uses a combination of social psychology and experimental economics. It uses the Ultimatum Game to capture backlash behaviour and implicit and explicit association tests to measure gender bias. The methodology further explains these three methods.

1.2 Relevance

Even though gender inequality has generally decreased in the western world, a lot of inequality and gender bias still exists. Specifically, the current paper focuses on the gender bias in a basic negotiation setting, which is relevant for several reasons.

Firstly, being a strong negotiator can be important for furthering one's position in a company. If a person is more capable or more enabled to perform well in negotiations, the chances of thriving and reaching the top of that company improve. Therefore, the current research is relevant for explaining why there are fewer women than men in top positions. If the research finds that individuals expect communal behaviour from women and reject more offers from them, this could indicate that the backlash effect contributes to the glass ceiling effect, which holds women back from reaching top positions in a company.

Secondly, the research is relevant for the discussion about a gender wage gap. In many companies, wages differ among employees and are a result of wage negotiations between the employer and the employee. According to Kolb (2009), one of the ways in which the compensation gap can be decreased is through pro-active and effective negotiation. However, if the research finds that agentic women experience backlash from their counterparts in negotiations, acting pro-actively does not help them and this backlash can partially

explain the gender gap in terms of salary.

Thirdly, the current paper may have some external validity in other business domains. The underlying logic in this research is that implicit and explicit associations about gender-attributes can predict backlash behaviour. If the paper finds that backlash to agentic women arises in a basic negotiation setting, the same might be found in other business domains, such as sales or leadership.

The current paper also furthers our understanding of the impact of implicit associations on negotiation behaviour. The current combination of the Implicit Association Test from the field of social psychology and the Ultimatum Game from the field of experimental economics has been used in the context of racial bias. Here, the offers of black people were rejected more often when there was a stronger racial bias in the counterpart (Kubota, Li, Bar-David, Banaji, & Phelps, 2013). The current research expands the combination of these methods to the context of gender bias.
Recent research has shown that women's lower performance in a negotiation game is predicted by their counterparts' implicit stereotypes on negotiation skills (Pardal, Alger, & Latu, 2020). However, this paper focused on the outcome of the negotiation rather than on the cause of this outcome. As was mentioned by Dannals et al. (2021), the gender gap can either be caused by women behaving in a stereotype congruent way out of a fear of backlash, or by women experiencing actual backlash. The current paper focuses solely on the responses of the counterparts to female and male offers. By doing this, it isolates the backlash effect. Additionally, the current paper considers agentic versus communal behaviour rather than perceived negotiation skills to capture the backlash effect.

Finally, by including both the implicit and the explicit gender associations, the current paper provides new insights on the interplay between these two metrics. Previous studies find that implicit and explicit associations do not always align (Fazio & Olson, 2003; Greenwald, McGhee, & Schwartz, 1998; Rudman & Kilianski, 2000; Kawakami & Dovidio, 2001; Monteith, Voils, & Ashburn-Nardo, 2001; Ottaway, Hayden, & Oakes, 2001). The current paper tests if this holds for gender bias on communal versus agentic behaviour and considers if implicit or explicit associations have a bigger influence on negotiation behaviour.

2 Theoretical framework

The current research combines several areas of literature. The theoretical framework is divided into four subtopics: Ultimatum Game and gender, gender bias in negotiations, backlash effect and the self-advocacy dilemma and implicit and explicit associations. The existing literature is used to formulate suitable hypotheses which help to answer the research question.

2.1 Ultimatum Game and gender

The current research uses the Ultimatum Game to study first responses in a basic negotiation setting. Previous research on gender bias using game theory show some mixed results.

Aguiar et al. (2009) lets people choose if they want a male or female dictator in the dictator game. In the dictator game, player 1 (the dictator) is endowed with a certain amount of money and can then decide to give any part of that to player 2. The majority of the subjects choose a female dictator, indicating they expect higher or more generous offers from women. This finding shows an associations of women with generosity or being nice. In general, behaviour that increases economic negotiation outcomes are more congruent with male compared to female gender roles (Meyerson & Fletcher, 2000).

Two important papers on gender differences in the Ultimatum Game find opposing results. Both Solnick (2001) and Eckel and Grossman (2001) find that offers to women are on average lower than offers to men. This could indicate that people expect women to accept more offers and men to be less willing to accept, which is consistent with the agentic and communal stereotypes. However, Solnick finds that offers from men are significantly less likely to be rejected, while Eckel and Grossman find the opposite. When considering explanations, Solnick indicates that women are not content with less than men, since female responders indicate higher minimum acceptable amounts than male responders. This finding is not in line with the previous research on expected compensation and entitlement to higher rewards (Major, McFarlin, & Gagnon, 1984; Jackson, Gardner, & Sullivan, 1992; Kaman & Hartel, 1994; Desmarais & Curtis, 1997; Jost, 1997). Alternatively, Solnick states that other players seem to expect women to be content with less, causing lower offers to them and higher minimum acceptable offers. Eckel and Grossman explain that their results support the chivalry hypothesis that is in line with previous research by Eagly and Crowley (1986), and helping behaviour can arise in male-female dyads. There are some important design differences between the two studies, as recognized by the authors. Firstly, Solnick uses a one-shot game under two treatments while

Eckel and Grossman let all subjects play both roles in the game four times. Differences can arise due to limited understanding or "noise" in the Solnick game or due to learning effect or perhaps even signalling in the Eckel and Grossman setting. Secondly, Solnick uses the strategy method, while Eckel and Grossman use a sequential game method. The strategy method seems to lead to more rejections, this was also found in a meta-analysis on cultural differences in the Ultimatum Game (Oosterbeek, Sloof, & Van De Kuilen, 2004). Thirdly, subjects in the Solnick game are placed in different rooms, while the subjects in the Eckel and Grossman game face each other across the room. On the one hand, one could argue that individuals may not believe they are actually matched with someone if they cannot see their counterpart in the negotiation. On the other hand, it could be that physical presence of the counterpart enhances the observed chivalry and solidarity effects. These effects could be caused by the manners of individuals (Camerer & Thaler, 1995), which may be stronger in a face-to-face setting. Finally, Eckel and Grossman highlight that there may be some differences due to the subject pool, because these are taken from different universities and different study backgrounds.
A third research, by McGee and Constantinides (2013), aims to clarify some of the differences. Still, it is neither able to completely confirm nor deny one of the papers and stresses the need for repetition. Another research by Fabre et al. (2016) tries to measure a potential backlash effect for women with stereotypical 'male' jobs in the Ultimatum Game. In this Ultimatum Game, the responders see a gender-identifiable name along with either a typically male or female profession of their proposer. They find no proof for the backlash effect in this instance. It should be noted that the potential backlash effect in the current study is only based on potential incongruence between expected communal behaviour and the amount offered by women in a proposer role, instead of on additional characteristics such as the job in Fabre et al. Some previous research on the Ultimatum Game and gender also finds that risk attitudes can play role in the behaviour of subjects (García-Gallego, Georgantzís, & Jaramillo-Gutiérrez, 2012). However, since the current research only considers rejection levels by the responder, the risk attitudes do not play a role. Overall, the current research resembles the research by Solnick more closely; the online survey ensures anonymity from the counterpart, there is a one-shot design and the strategy method is used as well. Therefore, I expect to find results that more closely resemble Solnicks' findings.

Following the previous research on gender in negotiations and backlash effects, I formulate the first hypothesis. The prediction is that women are expected to act more communal and because of it, in line with Solnick (2001), there will be higher minimum acceptance levels for female proposers than for male proposers. This results in the following hypoth-

esis:

Hypothesis 1: Individuals show higher rejection levels to offers from women than to offers from men

2.2 Gender bias in negotiations

Studies about gender differences in negotiations are often fueled by concerns about the gender wage gap and the glass ceiling effects. However, only highlighting differences puts responsibility for change on "fixing the women" instead of on creating cultures in which both can thrive (Meyerson & Fletcher, 2000; Kolb, 2009).

Women are found to obtain worse bargaining outcomes in an employer-employee situation when they bargain as the employee (Dittrich, Knabe, & Leipold, 2014). At every stage of their career, women seem to be less effective in gaining access to powerful positions, worse negotiation outcomes may contribute to the observed glass-ceiling for women in the workplace (Stuhlmacher & Walters, 1999).

The question arises if women have worse negotiation skills than men. Interestingly, women can advocate effectively on behalf of others, but much less so when they have to advocate for themselves (Wade, 2001). This indicates that it may not be a lack of negotiation skills that is holding them back. Even though men and women may not perform identically, there is no factual basis for saying women perform weaker or worse in negotiations (Craver & Barnes, 1998). If the bargaining skills of women are not worse than those of men, it seems other reasons than the lack of skills play a role.

One could argue that part of the slower advancement of women in companies and the gender pay gap are caused by differences in willingness to negotiate and a lack of assertive or agentic behaviour by women. Previous literature finds this could be due to women not asking for more, avoidance of gender incongruent behaviour and different expectations. Trying to get higher compensation is in line with the stereotypical agentic male behaviour, but it contradicts the expectations of behaviour from communal women (Wade, 2001). The authors of the book "women don't ask" indicate that women are reluctant to ask for more because they fear that asking can be damaging to the personal relationship or because society may react badly to them being assertive (Babcock & Laschever, 2009). In one example, investment bankers are allowed to enter a requested amount for a gift card after completing a survey. Only 76 percent of females compared to 90 percent of males

choose to fill in an amount (Greig, 2008). The authors use this example to explain that fewer women are willing to negotiate, and this causes them to advance more slowly than men.

Gender stereotypes, and specifically gender-incongruency, can lead to avoidance (J. Bear, 2011). In her research, Bear finds that women are more likely to avoid negotiation about compensations than men, whereas men avoid negotiations about the access to a lactation room. Kugler et al. (2018) confirm that women are less likely to initiate negotiations than men in certain contexts. The differences are smaller when situational ambiguity about the appropriateness to negotiate is low and when the situational cues are more consistent with female gender roles than with male gender roles. When the situational appropriateness to negotiate is perceived as higher, in the female gender-congruent situation of hiring a babysitter, women's perceived self-efficacy is a more positive predictor of their outcomes (Miles & LaSalle, 2008). Additionally, the framing of negotiations can play a role. Opportunities for asking instead of opportunities for negotiating over the same demands suit the stereotypical female gender roles better and are less intimidating for women (Small et al., 2007). Men performed better when negotiating over a typically masculine object while these differences are not found when negotiating over a typically feminine object (J. B. Bear & Babcock, 2012). These findings indicate that gender differences may not be absolute but situation dependent, and that women show avoidance when the stereotypical appropriateness for them to negotiate is more questionable.

Differences in expectations may also influence likelihood to negotiate. Women seem to expect less compensation for their work (Major et al., 1984; Jackson et al., 1992; Kaman & Hartel, 1994) or do not feel entitled to higher rewards (Desmarais & Curtis, 1997; Jost, 1997). In a study among business students, men have higher expectations and are more likely to actively negotiate their salary (Kaman & Hartel, 1994). Research among MBA students shows that women negotiate lower salaries, but this effect disappears when controlling for their pre-determined goals (Stevens, Bavetta, & Gist, 1993). Reservation salary, being the minimum salary someone is willing to accept, can therefore partly explain why females ask less salary than males under stereotype threat (Tellhed & Björklund, 2011).

2.3 Backlash effect and the self-advocacy dilemma

The differences in willingness to participate in negotiations still do not explain why women often obtain worse outcomes when they are already involved in a negotiation. The social role theory explains how differences between men and women arise from their social roles in a society (A. H. Eagly & Wood, 2016). There is often a general consensus about these

roles in cultures. Gender-typical behaviour is encouraged by other people and individuals tend to conform to their gender identities (Wood & Eagly, 2012). Prescriptive stereotypes indicate how we think men and women should act (Bowles, 2013). Stereotype policing, in which we protect and maintain stereotypes by punishing deviations from this behaviour, is also termed backlash.

This idea about stereotype policing, along with previous findings, sets the stage for the ideas about backlash. A model of cultural stereotypes shows how certain behaviour can be explained in the context of expectancy violation through justification, backlash and fear of backlash (Rudman & Fairchild, 2004). The perceiver of the expected violation seeks justification and gives backlash in the form of social or economic sanctions. The actor of the violation has a fear of backlash and uses recovery strategies like hiding the deviant behaviour. According to Rudman and Fairchild, the probability of backlash increases when the perceivers think it is justified. Backlash and fear thereof are not exclusive to the field of gender inequality. For example, backlash has been studied in the situation of ethnic discrimination against Arab Americans after 9/11 (Nassar-McMillan, Lambert, & Hakim-Larson, 2011) and in the context of casual sexual offers (Conley, Ziegler, & Moors, 2013). In 1972, the phrase "fear of success" is used in the context of achievement related conflicts for women (Horner, 1972). The fear of success seems to play a role in both men and women in stereotype-incongruent roles (Costrich, Feinstein, Kidder, Marecek, & Pascale, 1975; Cherry & Deaux, 1978).

A recent study argues that the gender gap can either be caused by women behaving in a stereotype-congruent way or by women receiving backlash for acting in a stereotype-incongruent way (Dannals et al., 2021). Women who succeed at a stereotypical male task can receive penalties (Heilman, Wallen, Fuchs, & Tamkins, 2004) and the current study is interested in this backlash effect. In order to measure this, it must first be established what the stereotype gender expectations are in the context of negotiations. The following parts highlight prescriptive stereotypes and stereotype-incongruent behaviour in several business settings in order to find the relevant stereotypical expectations for the current study. Specifically, the focus here is on gender stereotypes in leadership, self-advocacy and finally negotiations.

Firstly, when women in leadership positions carry out their tasks in typically masculine styles, they are devalued relative to their male counterparts (A. H. Eagly, Makhijani, & Klonsky, 1992). Drawing from previous papers, the authors describe masculine and feminine styles as agentic and communal respectively (Bakan, 1966; Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; A. H. Eagly & Steffen, 1984). Additionally,

women using intimidation techniques are seen as less likeable while men who use the same techniques are not (Bolino & Turnley, 2003). The men using these techniques are then seen as better performers, while the women are not.

Secondly, when competent or agentic women apply for a job, the job criteria seem to shift from competence to social skills (Phelan, Moss-Racusin, & Rudman, 2008). This situation leads to the self-advocacy or self-promotion dilemma (Janoff-Bulman & Wade, 1996; Rudman, 1998; Wade, 2001). Self-promotion may be necessary for leaving a competent impression, yet women who decide to do so violate gender prescriptions and can therefore face social reprisals (Rudman, 1998). Women experience social and economic penalties or backlash for self-promotion and the fear of this backlash interferes with self-promotion success (Moss-Racusin & Rudman, 2010). In some contexts, women hedge their assertiveness and use fewer competing tactics, since this is gender-incongruent for them and they anticipate negative feedback or backlash (Amanatullah & Morris, 2010). The backlash effect seems to be the most evident in a situation in which the communal behaviour of women is activated but suppressed by the woman (Tinsley, Cheldelin, Schneider, & Amanatullah, 2009). Self-advocating or assertive women suffer social backlash and non-assertive women suffer leadership backlash, as they are perceived as being less competent (Amanatullah & Tinsley, 2013). Janoff-bulman and Wade (1996) find that past experiences have taught women not to burn bridges, while men do not face the same consequences for self-promotion, as this is congruent with the male stereotype of agentic, rather than communal, behaviour.

Thirdly, in the context of negotiations, Kulik and Olekalns (2012) state that "best practice" negotiation behaviour can lead to expectancy violations for women and generate backlash. Male negotiators punish female negotiators more than other males for starting a negotiation, while females penalized both men and women (Bowles, Babcock, & Lai, 2007). Additionally, the anticipated backlash they reported was negatively correlated with the likelihood of asking in a negotiation, once again confirming the avoidance finding recognized by Bear (2011). It is important to note here that the self-advocacy or negotiation dilemma does not just influence women who are currently involved in negotiations, but it can play a role in the willingness to negotiate as well. If women in negotiations are confronted with a backlash based on stereotypes, this may decrease the likelihood of them participating in future negotiations.

Much of the relevant literature indicates gender expectations can be described as either agentic or communal behaviour. Agentic behaviour can be seen as individualistic, determined and assertive. Communal behaviour on the other hand is about compassion, caring and cooperation. Females are often associated with communal rather than agentic

words, and agentic women are seen as less likeable (A. Eagly, 1987; Rudman & Glick, 2001). Males are often associated with agentic characteristics and females with communal characteristics (A. Eagly & Wood, 2012). Importantly, these agentic or stereotypical masculine traits are often seen as important for negotiation success. The terms to describe successful negotiators overlap with the terms describing stereotypical men because they are agentic terms. In this way, 'negotiation' itself is not gender neutral (Small et al., 2007). One can expect communal proposers to offer higher amounts to the responders, as they are less competitive and more social. Since women are associated more strongly with communal behaviour, it could be the case that deviations from this social stereotype cause a backlash.

2.4 Implicit and explicit associations

A study from 1995 tests race and gender discrimination in the bargaining for a new car (Ayres & Siegelman, 1995). In this study, prospective car buyers with different genders and ethnicities use identical bargaining scripts at several car-dealers. The dealers offer significantly lower prices to white males than to black or female test buyers. This particular study does not present one overall theory of discrimination, but one could argue the car-dealers make assumptions about their counterpart's know-how or negotiation skills. Kray and Thompson (2004) offer several theoretical perspectives on gender bias in negotiations. They argue that most negotiators have some implicit links between gender stereotypes and negotiation performance. Therefore, even if the car-dealer are not aware of this, implicit associations may influence their offers.

Measuring implicit beliefs and the corresponding expectations can be difficult, and this field of research has seen some developments. Female leaders who make the same arguments as male leaders, according to a set script, experience more negative non-verbal response (Butler & Geis, 1990; Koch, 2005). This non-verbal response seems to be an indicator of the implicit associations within the subjects. Another study finds that in a racial setting, explicit prejudice predicts verbal friendliness and implicit prejudice predicts non-verbal friendliness (Dovidio, Kawakami, & Gaertner, 2002). Additionally, women with identical resumes are hired less often for a research position than males (Steinpreis, Anders, & Ritzke, 1999) and blind-auditions for an orchestra result in more female hires than regular auditions (Goldin & Rouse, 2000). Either knowingly or unknowingly, the females are disadvantaged due to the associations of the assessors. Steele (1997) recognizes that individuals are more likely to fulfill gender stereotypes when these stereotypes "hang in the air", so when individuals are aware of the stereotype relevance but are not

directly confronted with it. Interestingly, by using stereotype priming it was found that when women are confronted with negative explicit stereotypes about female performance, they tend to counter this prediction (Kray, Thompson, & Galinsky, 2001). Men fear not fulfilling the positive stereotype, causing self-doubt and lower performance after explicit priming. In the case of priming implicit stereotypes, the men still seem to have the upper hand.
Implicit beliefs also influence how people perform in negotiations. By measuring the implicit negotiation beliefs using scales adapted from Chiu, Hong and Dweck (1997), Kray and Haselhuhn (2007) find that negotiators who believe negotiation attributes are easily influenced perform better than those who think these attributes are fixed.

Most recent researches on implicit gender stereotypes use the Implicit Association Test (IAT). This test uses the reaction time of a subject to assess how strongly they associate certain concepts with each other (Greenwald et al., 1998). Since its introduction, the test has been widely used to measure the impact of racial, gender or other biases on groups of individuals.
In the field of gender bias research, the IAT shows that both men and women associate men more strongly with authority than women (Rudman & Kilianski, 2000). Additionally, men interacting with female superiors show negative associations with women, while men interacting with female equals or subordinated do not show this, stressing the importance of the situation when measuring implicit associations (Richeson & Ambady, 2001). In the hiring process, feminization of management, requiring managers to be both agentic and communal, may actually lead to more discrimination against women (Rudman & Glick, 1999, 2001).
Previous research on the Ultimatum Game and implicit associations shows that a physical description of the proposer changes acceptance rates (Marchetti, Castelli, Harlé, & Sanfey, 2011). The term "beauty premium" is used to describe how attractive people received higher offers (Solnick & Schweitzer, 1999). A 2013 study on racial bias uses a combination of the Ultimatum Game and the Implicit Association Test (Kubota et al., 2013), just like the current paper does. In this paper, racial bias, deducted from the IAT, is a predictor of the acceptance rates in the Ultimatum Game. The offers of black people are rejected more often isf their counterpart has a stronger racial bias. An important difference to the study of Kubota et al. is that the Implicit Association Test used in this research does not consider good or bad for males and females (black and white people in Kubota et al.), but rather checks for a specific stereotype on expected behaviour.

Using an IAT, Rudman and Glick (2001) confirm that females are associated more strongly

with communal words than with agentic words. As was mentioned before, these terms are used in the current research as well. A recent study also finds that women's lower performance in a negotiation game is predicted by their male counterparts' implicit beliefs about negotiation skills (Pardal et al., 2020). This effect is the strongest in situations in which the counterparts of the women were males with strong implicit but weak explicit biases. This is a clear example of how individuals do not always communicate their implicit associations, either because they do not want to or because they might not be aware of them themselves. To observe potential differences between the impact of explicit and implicit associations on negotiation behaviour, both implicit and explicit measures are used.
The current research is similar to the research by Pardal et al. (2020) in the sense that it is also interested in the impact or predictive power of implicit and explicit associations. However, there are a couple of major differences. First of all, the current study uses the Ultimatum Game rather than an extensive negotiation exercise and therefore captures just the first response data to an ultimatum instead of a back and forth negotiation. Secondly, the current research is primarily focused on the responder data, rather than the overall outcome. By doing this, it aims to isolate a potential backlash effect, without the interference of fear of backlash. Finally, the current study looks at associations with agentic and communal behaviour, instead of negotiation skills, because the aim is to find out if individuals follow their traditional gender stereotype rather than perceived skills.
Therefore, if a correlation is found between the associations and the rejection levels in the Ultimatum Game, similar to Kubota et al., this could be an indication that individuals with stronger gender associations assert backlash to women who do not act according to their communal stereotype.

In line with the research of Pardal et al. (2020) and Kubota et al. (2013), the current research is interested in the predictive power of implicit and explicit gender associations on negotiation behaviour. This results in the following general hypothesis, which can help to answer the research question presented in the introduction:

Hypothesis 2: Stronger implicit and explicit associations between women and communal words or men and agentic words predict higher rejection levels to offers from women

Hypothesis 2 is divided into hypothesis 2a and 2b, for the influence of implicit and explicit associations respectively, because the implicit and explicit associations do not always align (Fazio & Olson, 2003; Greenwald et al., 1998; Rudman & Kilianski, 2000; Kawakami &

Dovidio, 2001; Monteith et al., 2001; Ottaway et al., 2001). To disentangle the effects of the implicit compared to the explicit associations, hypothesis 2c is about the relative predictive power of implicit versus explicit associations. Fazio and Olson (2003) discuss that there is no ultimate truth in what is to be considered the "real" belief, and both implicit and explicit beliefs can predict behaviour in certain situations. According to them, when motivation and/or opportunity to deliberate are low, the implicit measure is likely to be more predictive. Additionally, the implicit measure may be more predictive when the questions concern stereotypes or sensitive social topics (Kawakami & Dovidio, 2001). Because the current research concerns a rather sensitive topic and individuals may have limited motivation to deliberate, I expect the implicit associations to be stronger predictors of rejection levels than the explicit associations.

3 Methodology

This study aims to answer the research question: "what is the influence of implicit and explicit gender bias on rejection levels in negotiations?". The following sections explain the methods used to be able to answer the research question.

3.1 Experimental design

The current research has an experimental approach with an online survey to generate the necessary data. Depending on reading speed, the survey takes approximately 10-15 minutes to complete and could only be done on a device with a keyboard.
The subjects could only advance on the correct device and the instructions read that this is part of a master thesis and more information on the topic is provided at the end of the survey. Next to this, the subjects had to consent to the general rules of the survey and the use of their anonymous data for the purpose of this master thesis study. The introduction also stated that they can leave the survey at any time without consequences, they can contact the researcher for questions, the results are only shared in group form and cannot be linked to an individual and the data is stored anonymously using password protection for a maximum of five years after publication. After giving their consent, subjects could choose to take the survey in Dutch or in English, in order to assist some Dutch speaking participants. However, to prevent any differences in associations or interpretations between Dutch and English words, all participants saw the same English words in the implicit and explicit association tests.

The experiment then consisted of four parts: (1) general questions about subject demographics, (2) the Ultimatum Game, (3) the Implicit Association Test (IAT) and (4) explicit association questions including the Gender Sterotype Index (GSI). The following sections further elaborate on the details of these four parts and all the survey questions can be found in Appendix A.

All subjects completed the Ultimatum Game only once as a responder and the study uses a between-subjects approach. Randomization took place at various stages of the experiment. Firstly, the survey randomly allocated subjects to play against a male or female proposer in the Ultimatum Game. Secondly, it randomly ordered parts (3) and (4) in the survey, the Implicit Association Test (IAT) and the Gender Stereotype Index (GSI), to counter any potential effects of having one before the other. Thirdly, within the GSI, subjects got a random question order to mediate any anchoring or priming effects.

The IAT randomly allocated subjects to one of four variants, with each variant having different categories left and right or a different order of the questions.
Figure 1 below portrays the general structure of the survey. Both the Ultimatum Game with the male and female responder also let the subject play in a proposer role afterwards and the IAT randomly allocated subjects to one of its four variants.

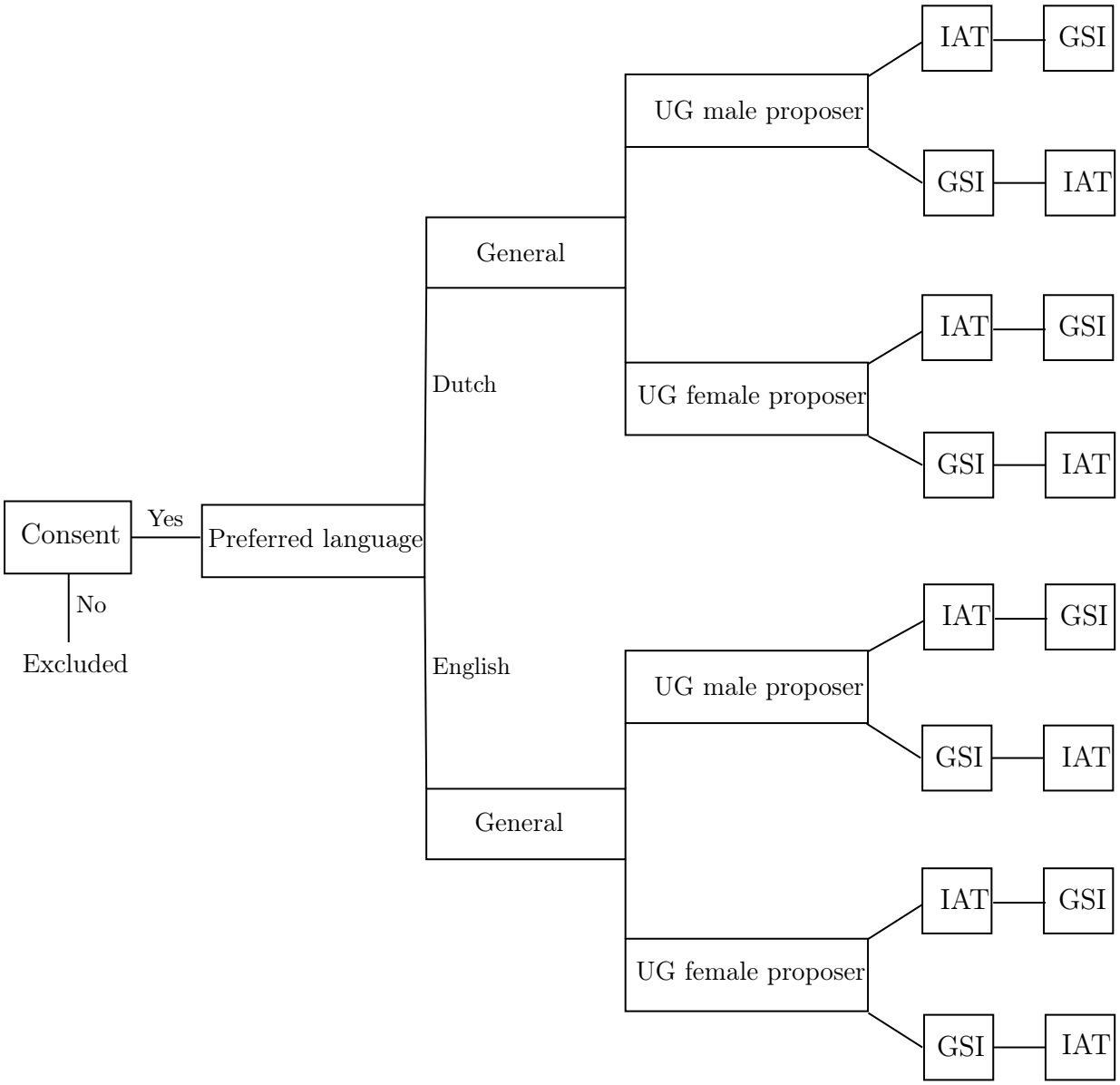

Figure 1: Survey structure diagram

Individuals had a chance of winning up to 10 euros, based on a random lottery selecting five couples and based on the outcome of their Ultimatum Game.

3.2 Recruitment

Before the data collection, I performed power calculations based on data collected by the research of Solnick (2001) on gender in the Ultimatum Game. This study is the most suitable reference point for the current study because it uses a somewhat similar approach, uses the same offer amount and also focuses on male and female proposers. The Eckel and Grossman (2001) research is not suitable because it only presents standard deviation for the proposals, not for the rejection levels. Both power calculations can be found in Appendix B. The first calculation considered the ideal division of participants to the subsamples of male and female proposers. Ideally, approximately 52% of the participants is allocated to the sample with male proposers. Due to the small difference obtained here and in order to increase the efficiency of the matching process, the survey allocated the subjects evenly across the two samples.

The second power calculation was aimed at finding the required number of participants in the experiment. For this purpose, the G*Power software by Faul et al. was used (2007). This is an a priori power analysis to compute the sample size, specifically for the Mann-Whitney U test that is performed after data collection. Information was needed on whether it concerns a one-tailed or two-tailed test, on the effect size d, on the type 1 and type 2 errors and on the allocation ratio. Hypothesis 1 predicts a one-sided direction of the effect, based on the findings of Solnick. However, because research in this domain has been ambiguous and I could not exclude the possibility that the opposite effect was found, a two-tailed test was performed. The effect size was calculated using the mean of the minimum acceptable offers to male and to female proposers and the standard deviations from the research by Solnick. The alpha included was 0.05, based on a 95% confidence interval, and the power of the test was set to 0.80. This resulted in an ideal sample size of 46 for both groups, or 92 in total. No initial distinction was made between male and female responders, since this is not the primary focus of the current study.

No individuals were excluded from participating in the survey. The survey was spread among the author's social networks through direct messages and LinkedIn and the survey was closed after 1,5 weeks.

3.3 Materials

3.3.1 General questions

The first part of the survey consists of general questions that also inquire about the demographics. The questions are about the age, gender, country of origin and marital status of the subjects. Additionally, a control question inquires how subjects got in contact with the survey.

3.3.2 Ultimatum Game

The Ultimatum Game is a widely used game in experimental economics. The game consists of two periods and has two players; a proposer and a responder. The proposer is endowed with a certain amount of money, which is 10 euros in the current study. The proposer can offer any part of the money to the responder. This offer is sent as an ultimatum, meaning that the responder can choose to either accept or reject the offer. In the case a responder rejects the offer, neither player gets any reward. If the responder accepts, both players get the money that is allocated in the game. Therefore, completely rational responders would accept any offer greater than null to maximize their own payoff, yet this is not always observed in practice.

All subjects first played in the role of responder. I used a one-shot design to prevent spillovers or learning, just like the design of Solnick (2001). Before the start of the game, the text explained the procedure and the probabilities of winning monetary amounts to the subjects. The subjects indicated they understood the rules of the game, their role as responders and the potential payoff. The subjects could see the gender of their proposer, and in order to prevent the individuals making up their mind about their rejection level before seeing the gender of their counterpart, they saw the gender during the entire explanation.
A choice list presented subjects with possible offers of the proposer, ranging from zero to six euros in steps of 50 cents. Afterwards, I deducted the rejection level using the Becker DeGroot Marschak design (Becker, DeGroot, & Marschak, 1964). The responders indicated at what point they would no longer accept the offer, if the male or female proposer offered this. Subjects were told they are matched with a proposer after the experiment. If the offer of the proposer was higher than their rejection level, and they were among the random lottery winners, I payed out the allocated amount to both subjects.
There was a timer for the answer, just like in Kubota et al. (2013). This time constraint could help to create a more instinctive real-life response. Because the strategy design of the current research takes more time to answer than the yes or no question in Kubota et

al., the timer was set to 10 seconds. However, to prevent problems with the data collection, the question did not auto-advance after 10 seconds. Instead, the timer acted as a nudge to increase the answer speed. Additionally, the response time for the Ultimatum Game was recorded to investigate if underlying evaluations or deliberate considerations were more likely to drive decisions.

The following screenshots from the survey show the manipulation and the choice list design. Figure 2 shows a screenshot that explains part of the Ultimatum Game and portrays the manipulation. Figure 3 shows a screenshot that captures the choice list design, along with the timer.

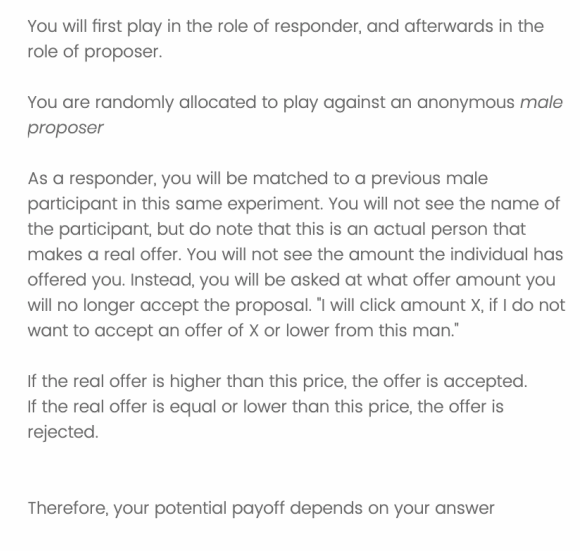

Figure 2: Survey screenshot: part of Ultimatum Game explanation

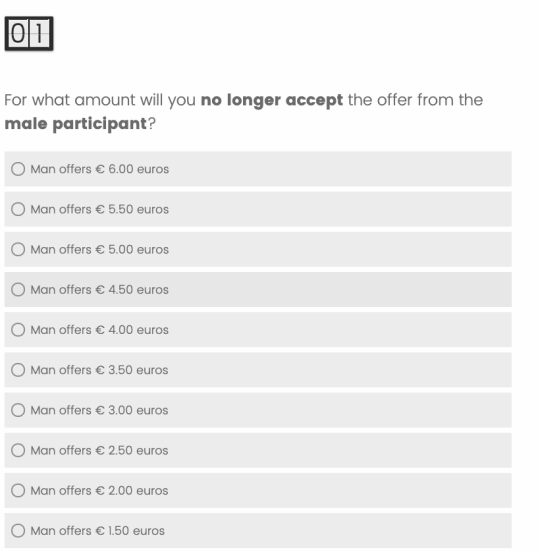

Figure 3: Survey screenshot: strategy design with timer

To avoid deception of the subjects, the experiment also let the subjects offer an amount to a random future participant. After the experiment closed, I made random couples of proposers and responders using a random number generator in excel, while considering whether this was supposed to be a male or female proposer. If the number of participants was uneven, or more people have answered to a male/female proposer than there were male/female participants in the sample, two responders were matched to just one proposer. I treated all of the responses to the same proposer as separate couples and included their outcomes separately in the random selection of winners. I then randomly selected five couples using another random number generator. These couples received their winnings if the proposed amount was higher than the responder's rejection level. No participant could win more than one game. At the end of the experiment, individuals who wanted to win the prize had to leave their contact details. Therefore, the research was conducted in a single-blind manner.

3.3.3 Implicit Association Test (IAT)

The Implicit Association Test uses reaction time to measure how strongly a subject associates certain *attributes* with certain *groups*. The attributes used in this test are 'agentic' and 'communal' and the groups are 'male' and 'female'. For the group words, I used the commonly-used gender words listed below. The attribute terms to distinguish communal versus agentic associations were the same terms used by Rudman and Glick (2001). Listed below are the terms for the attributes and groups. I composed the IAT with the help of the IATgen tool by Carpenter et al. (2019).

Gender
Male: man, son, father, boy, uncle, grandfather, husband
Female: woman, daughter, mother, girl, aunt, grandmother, wife

Agentic versus communal
Agentic: independent, competitive, autonomous, individual, hierarchical, and self-sufficient
Communal: communal, attached, cooperative, together, kinship, and commitment

The entire test consisted of seven blocks. In block one, one of the groups was placed in the top left corner of the screen, the other group was placed in the top right. Words that can be associated with either the left or right group appeared in the middle, and the subject had to allocate the word to the right or left by using the 'e' and 'i' keys on the keyboard respectively. If the subjects made a mistake, a red cross was shown in the screen and the subjects had to provide the correct entry. The second block was the same as the first block, except now it placed one of the attributes on the left and the other attribute on the right. The third block was a final practice round. Here, a combination of the group and attribute shown before was shown in the top left, and the other combination in the top right. Again, the individual had to allocate words that arrive in the middle as quickly as possible. The fourth block of the test was the same as the third, but now the results were recorded. Blocks five, six and seven of the test were similar to blocks one/two, three and four, with the exception that now the attributes or groups changed places.

Ultimately, the test compares the response speed in blocks four and seven. The relative response speed to one of the combinations of gender and agentic or communal compared to the other combination is an indication of an individual's associations. For example, if a person answered faster when female was matched with communal than when female

was matched with agentic, this is a sign that the person associates female more strongly with communal. However, the IAT results should be interpreted on a group level rather than for specific individuals.

One example of the test subjects do is illustrated in Table 1 below. With the two groups (male and female) and two attributes (agentic and communal), there are four alternative test orders. The IAT randomly allocated subjects to the different alternatives.

Table 1: Overview of one out of four possible versions of the Implicit Association Test

Block	**Left**	**Right**	**Type**	**Number of Trials**
1	Male	Female	Practice	20
2	Agentic	Communal	Practice	20
3	Male + agentic	Female + communal	Practice	20
4	*Male + agentic*	*Femlale + communal*	*Recorded*	*40*
5	Female	Male	Practice	40
6	Female + agenic	Male + communal	Practice	20
7	*Female + agentic*	*Male + communal*	*Recorded*	*40*

Due to the potential language gap with the subjects of this research and the somewhat difficult terms to describe communal or agentic attributes, I provided the division between agentic and communal words, as well as the meanings and Dutch translations of the words beforehand.

3.3.4 Explicit associations

To test hypothesis 2b, subjects completed the Gender Stereotype Index (GSI) as used by Rudman and Glick (2001). The GSI let subjects rate communal and agentic terms on a scale from -3 (truer of women) to 3 (truer of men). These traits were the same as the ones used in the IAT and the order in which they appeared was randomized to prevent priming or anchoring from the first question. Additionally, this part directly asked subjects about their associations of the overarching terms ‘communal’ and ‘agentic’ with gender.

3.4 Measures

Using the materials above, a lot of information was gathered. Firstly, the general questions delivered several demographic variables, in nominal, ordinal or ratio form. Most

important here is the gender of the subject, which is nominal data.

The Ultimatum Game obtained the rejection level for each subject, this is ratio data. The rejection level is the amount for which the subjects would no longer accept an offer from the male or female proposer in the game. A higher rejection level is less social, less rational, and it increases the chance that an offer is not accepted. This amount is used as an indicator for negotiation behavior. Additionally, the timer showed how long the subjects considered the possible proposals before continuing. For hypothesis 1, I use the data to compare rejections levels to male and to female proposers. In hypothesis 2, I consider if the implicit and explicit associations can predict the rejection levels.

The Implicit Association Test (IAT) delivered D-scores per participant, resulting from the D effect size algorithm proposed by Greenwald et al. (2003), this is ordinal data. These D-scores range from -2 to 2 and they indicate the association of individuals between gender and the communal or agentic words. A D-score of 0 indicates there is no implicit association at all. Higher D-scores indicate stronger associations between females and communal words or males and agentic words and negative D-scores indicate the opposite. I use the D-scores to test the predictive power of implicit associations for hypothesis 2.

The Gender Stereotype Index (GSI) was found by subtracting the mean of the communal traits from the mean of the agentic traits. This subtraction resulted in scores ranging from -6 to 6. Similar to the IAT scores, a higher score indicates stronger explicit associations between female and communal or male and agentic terms.

3.5 Planned analyses

From the literature review, two main hypotheses arose, and hypothesis 2 is split into three parts.

Hypothesis 1
H0: Individuals show similar rejection levels to offers from women as to offers from men
H1: Individuals show higher rejection levels to offers from women than to offers from men

Hypothesis 2a
H0: Implicit associations between gender and agentic or communal words cannot predict rejection levels to offers from women
H1: Implicit associations between women and communal words or men and agentic words

predict higher rejection levels to offers from women

Hypothesis 2b

H0: Explicit associations between gender and agentic or communal words cannot predict rejection levels to offers from women

H1: Explicit associations between women and communal words or men and agentic words predict higher rejection levels to offers from women

Hypothesis 2c

H0: Implicit associations between women and communal words or men and agentic words are not stronger predictors for rejection levels than explicit associations

H1: Implicit associations between women and communal words or men and agentic words are stronger predictors for rejection levels than explicit associations

3.5.1 Tests

After collecting the data, I needed several analyses to review the hypotheses and answer the research question. The next paragraphs present the planned analyses per hypothesis.

3.5.1.1 Hypothesis 1

Hypothesis 1 is about the rejection levels of the subjects to offers from male and female proposers. Hypotheses 1 is tested with the non-parametric Mann-Whitney U test. This test is chosen because the underlying distribution of the population is not known, and therefore I do not know if the necessary assumptions for parametric tests hold. For the Mann-Whitney U tests to give valid results, several assumptions must hold.
Firstly, it is assumed that the observations are independent. This is the case in the current research, as the subjects do not interact with each other and there is a one-shot design. Secondly, data is needed at least at an ordinal level. This assumption is also satisfied in the current research, as the rejection levels are on a ratio level. Thirdly, there must be one dependent variable with two independent groups. In the current research, the dependent variable is the Rejection level and the groups are "male proposer" and "female proposer".

The Mann-Whitney U test compares the distribution of the rejection levels to female offers to the Rejections levels to male offers. Firstly, the test ranks all the rejection levels from the two samples from low to high. Afterwards, it calculates the sum of the ranks per category (female offer/male offer) and considers if Rejection levels to female proposers differ significantly from Rejection levels to male proposers. The software STATA delivers the results. I reject the null-hypothesis if the test shows that the Rejection levels differ

significantly, depending on the gender of the proposer.

3.5.1.2 Hypothesis 2

Similar to Pardal et al. (2020), multiple linear regression help to answer hypotheses 2a, 2b and 2c. The dependent variable in these OLS regressions is the Rejection level from the Ultimatum Game. The regressions include either the D-score from the Implicit Association Test (IAT) or the Gender Stereotype Index (GSI) score from the explicit association questions. I include the gender of the counterpart as a dummy variable "Femaleproposer", that takes on value 1 if the proposer is female and 0 if the proposer is male. All regressions include the control variables Age, Education, gender (in the form of a dummy variable "Femaleresponder"), Timing and a dummy variable for subjects who indicated they are single. I choose the dummy variable "Single" rather than the variable marital status because the variable marital status itself has limited meaning. In the sample, 65 out of the 99 individuals who answered the question are single, 19 are married or living together, 1 is widowed and 14 choose "other". Finally, the regressions include an interaction term between the Female proposer dummy variable and the D-score from the IAT or the GSI-score.

Regression including D-score and interaction effect

$$Rejectionlevel = \beta_0 + \beta_1 \text{Age} + \beta_2 Education + \beta_3 Femaleresponder + \beta_4 Timing + \beta_5 Single + \beta_6 Dscore + \beta_7 Femaleproposer + \beta_8 DscoreXFemaleproposer + \varepsilon$$

Regression including GSI and interaction effect

$$Rejectionlevel = \beta_0 + \beta_1 Age + \beta_2 Education + \beta_3 Femaleresponder + \beta_4 Timing + \beta_5 Single + \beta_6 GSI + \beta_7 Femaleproposer + \beta_8 GSIXFemaleproposer + \varepsilon$$

The first formula includes the interaction term between the D-scores and the Female proposer dummy variable and can therefore test hypothesis 2a. The second formula does the same for the GSI and tests hypothesis 2b. Based on the outcomes of the regressions, I consider the validity of hypothesis 2c. This can be somewhat more arbitrary, as the strength of the predictive power of the D-score and the GSI depend on the coefficients of the variables with their significance in their respective models and on the predictive power of these models. If the results do not show unanimously that the implicit associations are stronger predictors, I do not reject the null-hypothesis.

Besides the two regressions above, I perform three additional regressions. Two of these

consider the effect of the D-score and GSI solely and exclude the interaction term from the regressions above. The fifth regression includes the D-score, the GSI and both their interaction terms.

4 Data

After the data collection, the sample consists of 100 respondents who have at least completed the Ultimatum Game and can therefore be included in the analysis for hypothesis 1. This number is higher than the necessary target that was calculated using power calculations. Of these 100 subjects, 95 completed the Gender Stereotype Index and 94 completed the entire survey, including the Implicit Association Test (IAT). Finally, four respondents were excluded from the IAT results due to excessive speed in the test.

4.1 Sample characteristics

Table 2 below illustrates some general demographics of the participants in the survey.

Table 2: Sample characteristics

	Mean	**Sd**	**Min**	**Max**	**Count**
Age	26.29	8.26	17	60	100
Gender	1.38	0.51	1	3	100
Femaleresponder	0.36	0.48	0	1	99
Education	3.28	0.70	2	5	100
Single	0.65	0.48	0	1	100
GSI	1.53	0.92	-0.33	4.17	95
Dscore	0.37	0.33	-0.39	1.17	90
Rejection	3.10	1.80	0	6	100
N	100				

The Age of the subjects ranges from 17 until 60 years old. The somewhat low mean age is mostly caused by the survey being spread among the social network of the author. The Gender ranges from 1 to 3, indicating it contains males, females and one person who indicated "other". The Female responder dummy variable below includes only the males and females and the low mean shows the gender in the sample is not evenly distributed. More specifically, 63 subjects are male, 36 subjects are female and 1 subject indicated "other". Education ranges from those who completed high-school to those who obtained a PhD. The biggest group consists of those who completed a bachelor degree (53%) and those who obtained a master degree (33%). This high level of education in the sample makes sense since the survey was mostly spread among (former) students. The dummy variable "Single" is included rather than the marital status because the marital status mean and standard deviation have limited explanatory value and 65% of the sample indicated they

are single.

The Gender Stereotype Index (GSI) has a mean of 1.53 on a scale of -3 to 3, and 0 indicates a neutral explicit association between gender and agentic or communal traits. On average in the sample, there is a stronger explicit association between males and agentic traits and females and communal traits. However, this is not true for the entire sample, as the minimum is slightly negative. The same holds for the D-scores. The D-score can range from -2 until 2, and an implicit association between males and agentic terms or females and communal terms results in positive scores. The mean score in the sample is positive, but this does not hold for all the subjects, as there are some negative scores as well.
One could argue that women score lower D-scores and/or GSI-scores because of an in-group effect or because the relatively highly educated women associate themselves with agentic terms and with female, causing a stronger association between the two. However, the mean D-scores for men and women are 0.38 and 0.37 respectively. The mean GSI for men and women are 1.60 and 1.46 respectively. These differences are not sufficiently large to assume that men and women in the sample have different explicit and implicit associations. Finally, the Rejection levels consist of the entire available range from 0 until 6 and the mean is 3.10. On average in the sample, individuals no longer accept the offer from the proposer when it is below the option of 3.50 euros.

One of the interests of this study is the potential difference between the implicit and explicit associations. If individuals are aware of their implicit associations and choose to answer honestly, I would expect their correlation to be equal or close to 1. The Pearson correlation coefficient is 0.364 and this is significant at a 0.1% significance level, clearly indicating the two results differ significantly in the sample. The correlations of control variables are further discussed in the Results section.

4.2 Subsamples

Table 3 below shows the results of t-tests on potential differences between the two subsamples based on the gender of the proposer. This is important to consider because differences between the two groups could maybe explain part of the results. Table 3 shows there are no significant differences between the characteristics of the subjects who interacted with a female proposer and of those who interacted with a male proposer, as none of the mean differences are significant at a 5% or 10% significance level.

Table 3: T-tests on differences between the subsamples on proposer gender

		Responder	
	Mean difference	**T-statistic**	**P-value**
Age	0.76	0.46	0.646
Education	-0.02	-0.13	0.900
Gender	0.11	1.09	0.279
Femaleresponder	0.08	0.79	0.430
Maritalstatus	0.48	1.39	0.168
Single	-0.13	-1.34	0.183
Timing	-0.83	-0.83	0.409
Dscore	-0.09	-1.28	0.205
GSI	-0.28	-1.51	0.134

5 Results

5.1 Hypothesis 1

Hypothesis 1 investigates if there are significant differences between the Rejection levels to proposals from male and from female proposers. The Ultimatum Game delivers the following mean Rejection levels for the different interactions.

Table 4: Mean Rejection levels in the Ultimatum Game

	Responder		
	Male	**Female**	**Average**
Proposer			
Male	€2.91	€3.26	€3.04
	(0.31)	(0.43)	(0.25)
Female	€3.27	€2.88	€3.14
	(0.31)	(0.48)	(0.26)
Average	€3.11	€3.08	€3.10
	(0.22)	(0.32)	(0.18)

Note. Standard error in parentheses.

Table 4 shows that the average mean over the whole sample is 3.10, with a standard error of 0.18. From this table alone, not much can be said about potential differences. The mean Rejection level seems to be lower for same-sex proposers, but it needs to be verified if these differences are significant. For females, this effect would be in line with the solidarity effect described by Eckel and Grossman (2001) and it is further investigated in section "5.3.2.3. Chivalry and solidarity effects".

The Mann-Whitney U test is conducted for the entire sample and for male and female subjects separately. Table 5 contains the results of these tests.

Table 5: Mann-Whitney U tests on the Rejection levels to male versus female proposers

	All subjects	**Males**	**Females**
Prob > |z|	0.693	0.363	0.523
Z	-0.395	-0.911	0.639
N	100	63	36

Note. H0: Rejection level to female proposer is equal to Rejection level to male proposers

None of the groups in Table 5 show a significant difference between the Rejection levels to male and to female proposers. The Mann-Whitney U test does not provide any evidence to reject the null-hypothesis that subjects show similar Rejection levels to male and female proposers.

5.2 Hypothesis 2

5.2.1 Raw data

The second hypothesis aims to test a potential correlation between the implicit and explicit associations of individuals and their Rejection levels to female proposers. The following figures 4 until 7 show the collected data points, with Rejection levels on the y-axis and the D-scores or the GSI on the x-axis. The figures include the linear fitted value lines along with 95% confidence intervals as well.

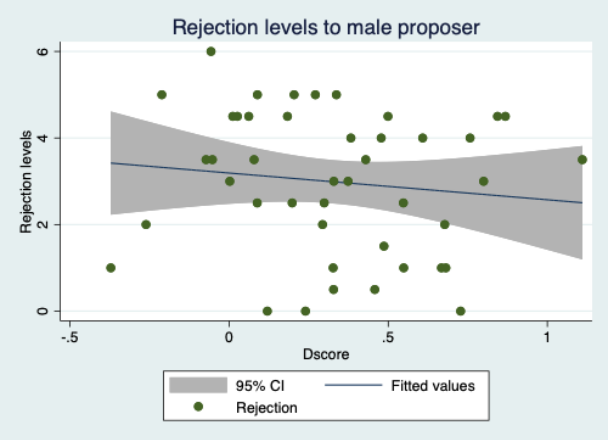

Figure 4: Rejection levels to male proposers and D-scores

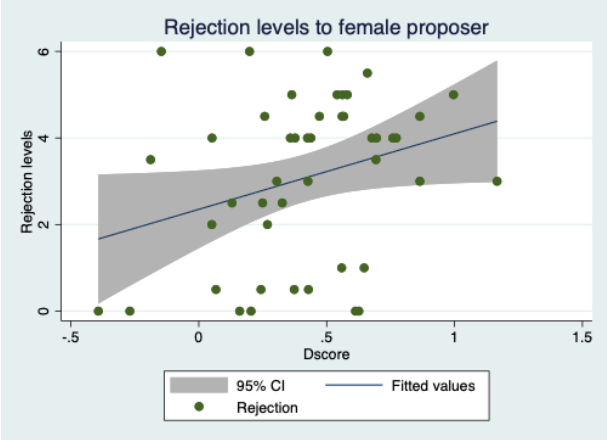

Figure 5: Rejection levels to female proposers and D-scores

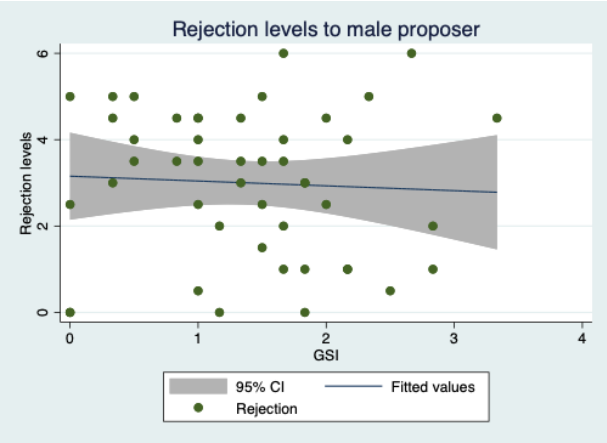

Figure 6: Rejection levels to male proposers and GSI

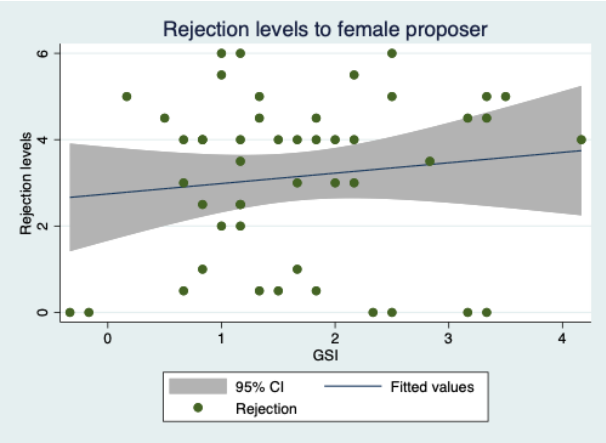

Figure 7: Rejection levels to female proposers and GSI

The top figures, Figure 4 and Figure 5, contain the D-scores from the Implicit Association Test while the figures below focus on the explicit associations. Figure 4 shows a slightly negative relationship between the D-scores of individuals and their Rejection levels to male responders, and Figure 5 shows a stronger positive relationship between the D-scores and Rejection levels to female proposers. On average in the sample, it seems individuals who implicitly associate males more strongly with agentic words and females with communal words show higher Rejection levels to female proposers. If this is indeed true, it is in line with hypothesis 2a. Looking at Figure 6 and Figure 7, the fitted value lines are tilted similarly as in Figure 4 and Figure 5 respectively, but the relationship seems to be weaker.

Table 6 below shows the Pearson correlation coefficients for the effects above, namely the correlations between the Rejection levels in the Ultimatum Game and the D-scores or GSI.

Table 6: Pearson correlation coefficients for D-scores and GSI with Rejection levels

	D-score		**GSI**	
	Male prop	Female prop	Male prop	Female prop
Corr. coefficient	-0.123	0.295	-0.053	0.127
P-value	0.425	0.047	0.722	0.390
N	44	46	47	48

Only the correlation between the D-score and the Rejection level when the proposer is female is significant at a 5% significance level. This finding is in line with hypothesis 2a, which states that a higher D-score predicts higher Rejection levels in negotiations with a female proposer. Additionally, Table 6 seems to confirm hypothesis 2c, which states that

the implicit associations are stronger predictors for Rejection levels than explicit associations. The following section looks more closely at these relationships using multiple linear regressions.

5.2.2 Multiple linear regressions

Table 7 displays the outcomes of five OLS regressions. The dependent variable in all of these models is the Rejection level obtained from the Ultimatum Game. Recall that the Rejection level is the level for which a subjects will no longer accept the offer from the other person. A higher Rejection level is therefore less social and less rational.
Model (1) contains the D-score as an independent variable, and Model (2) adds the interaction term with the Female proposer dummy variable. Model (3) and Model (4) do the same for the GSI. Finally, Model 5 contains both interaction effects.

Table 7: Multiple linear regressions on Rejection levels with interaction effects

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Age	0.052*	0.046*	0.051*	0.050*	0.046
	(0.028)	(0.027)	(0.028)	(0.028)	(0.028)
Education	0.274	0.310	0.318	0.317	0.296
	(0.322)	(0.317)	(0.315)	(0.316)	(0.322)
Femaleresponder	-0.046	0.013	-0.006	0.013	0.023
	(0.394)	(0.388)	(0.387)	(0.389)	(0.394)
Single	0.849*	0.821*	0.866*	0.859*	0.823*
	(0.468)	(0.459)	(0.453)	(0.455)	(0.464)
Timing	0.014	0.021	0.017	0.019	0.022
	(0.040)	(0.039)	(0.038)	(0.038)	(0.040)
Dscore	0.123	-1.021			-0.873
	(0.616)	(0.835)			(0.874)
Femaleproposer	-0.094	-0.959*	-0.066	-0.506	-1.279
	(0.385)	(0.576)	(0.383)	(0.736)	(0.782)
Femaleproposer x Dscore		2.294**			2.092*
		(1.154)			(1.243)
GSI			0.061	-0.117	-0.226
			(0.206)	(0.327)	(0.343)
Femaleproposer x GSI				0.295	0.272
				(0.421)	(0.446)
Constant	0.074	0.408	-0.137	0.119	0.704
	(1.216)	(1.206)	(1.236)	(1.292)	(1.305)
N	89	89	94	94	89
R^2	0.095	0.138	0.094	0.099	0.143
Adj. R^2	0.017	0.052	0.020	0.014	0.033
F	1.220	1.601	1.268	1.164	1.302

Note. Standard errors in parentheses; * $p \leq 0.10$, ** $p \leq 0.05$.

The most important finding is that the interaction terms between the Female proposer dummy variable and the D-scores are positive and significant in models (2) and (5) at a 5% and 10% significance level respectively. The interaction term allows me to measure the predictive power of the D-score purely in negotiations with female proposers. In the case of Model (2), when the proposer is female, a one-point increase in the D-score increases the Rejection level with 2.29 euros. This finding is in line with hypothesis 2a, which states that stronger implicit associations between women and communal words or men and agentic words (higher D-score), predict higher rejection levels to offers from women. The D-score itself is insignificant in all the models. It makes sense that if the prediction of hypothesis 2a holds, the negotiations with the male proposers would cancel out the potential predictive power of the D-scores in negotiations with female proposers. More specifically, I expected a higher D-score to predict lower Rejection levels to male proposers and higher Rejection levels to female proposers. If this holds for the sample, they can balance out the influence of D-scores on total Rejection levels and therefore leave the variable insignificant.

GSI and its interaction terms with the Female proposer dummy are insignificant in both instances they are included. Because of this, there is no evidence that explicit associations can predict responder behaviour and there is no evidence to reject the null-hypothesis of hypothesis 2b. This finding is however in line with hypothesis 2c, which states that the implicit associations have a stronger explanatory power for Rejection levels than the explicit associations. Besides the insignificance of the interaction term between the Female proposer dummy variable and the GSI, Model (4) also has a lower explanatory power than Model (2), as indicated by the R-squared, adjusted R-squared and the F-statistic.

Figure 8 and Figure 9 present a visual representation of the linear predictions of Rejection levels for different D-scores and GSI-scores. These figures show similar trends as the figures 4 until 7, but they extrapolate as well. In both cases, the linear predictions show a positive trend when the proposer is female. The linear predictions of Figure 8 clearly illustrate the general trend that a higher D-score predicts higher Rejection levels to female proposers (red line) and lower Rejection levels to male proposers (blue line). The trend suggests that stronger associations between women and communal terms or men and agentic terms in the counterpart predict higher Rejection levels to female proposers. The same seems to hold for the GSI in Figure 9, but the relationship is weaker and as we can see in Table 7, this effect is not significant in the current models. Importantly, the trends concern linear predictions, the confidence intervals are quite large and the multiple linear regressions indicate limited significance for some of these correlations.

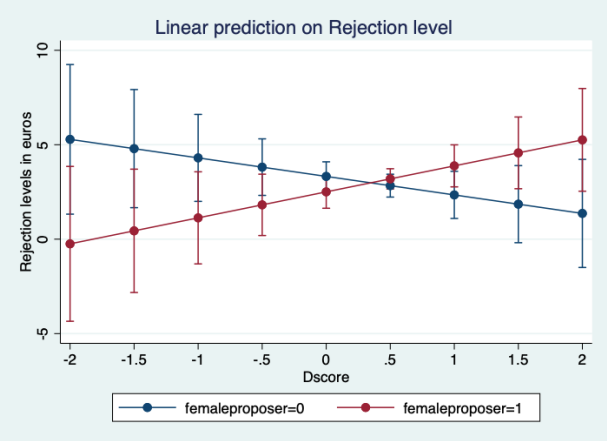

Figure 8: Interaction effect D-scores with Female proposer dummy

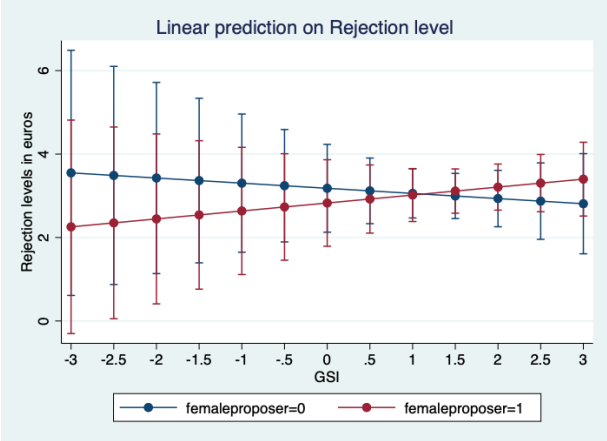

Figure 9: Interaction effect GSI with Female proposer dummy

It must be noted from Table 7 that the R-squared, adjusted R-squared and F-statistic are low for all the available regressions, meaning the models have limited explanatory power. Additionally, most of the variables are insignificant at a 10% significance level and only the interaction effect of the dummy variable Female proposer and D-score is significant at a 5% significance level in one instance.
The variable Age is positive and significant at a 10% significance level in all but the last model, indicating older individuals on average have higher Rejection levels. However, due to the uneven distribution of Age in the sample this finding must be interpreted with caution. The coefficient of the dummy variable Single is also significant at a 10% significance level in all the models, suggesting that individuals who indicate they are single in the sample show higher Rejection levels.

5.3 Further analyses

5.3.1 Hypothesis 1

The findings of the Mann-Whitney U tests do not support hypothesis 1 and they are not consistent with the research by Solnick (2001). Several robustness tests help me to consider alternative explanations for the findings. Firstly, the response latency could influence the results. Kubota et al. (2013) use a timer to give the subjects time-pressure. The current research uses the response time purely to impose a feeling of time-pressure, but it does not automatically advance to the next question. Perhaps slower responders answer less intuitively and therefore deliver different results.

I performed several additional Mann-Whitney U tests for subsamples based on the response latencies. The results from these tests can be found in Table 11 in Appendix C. The results are insignificant for all the subsamples based on maximum response time. It could still be the case that a shorter maximum response time with automatic advancement alters the results and this may be valuable in future research. However, the current number of quick responders is relatively small and the results show no clear indication that certain subsamples based on time might find significant differences.

Secondly, perhaps the Mann-Whitney U test was not the most suitable test for this sample. As was explained in the methodology, the Mann-Whitney U is less suitable when there are many duplicate answers in the sample. Table 12 in Appendix C below shows the possible options for Rejection levels, along with the number of subjects who chose this. As Table 12 shows, there seems to be a decent spread for the Mann-Whitney U test to give meaningful results.

Alternatively, a t-test was performed to consider if this test found any significant differences in Rejections levels to male and female proposers. However, the t-test is a parametric test and this requires four general assumptions to be true. Firstly, it assumes that the observations are independent. This seems to be the case in the current research, as the subjects do not interact with each other and it is one-shot design. Secondly, subjects must be drawn from a normally distributed population. The distribution of the population is not known and this needs to be assumed when interpreting the results of the test. Thirdly, the two groups in the comparison must have the same variance. Based on the previous research by Solnick (2001), this is likely to hold in the current research. Finally, the variables must be measured in a continuous or ordinal level to interpret the results. This is also satisfied in the current research, as the Rejection levels are indeed on an ordinal level.
The second assumption causes the most uncertainty to the validity of the t-test and this must be kept in mind when interpreting the results. Table 8 below displays the results of the t-test for the sample as a whole and for subsamples based on the gender of the responder.

Table 8: T-tests for difference in Rejection levels to male and female proposers for subsamples on gender

	All subjects	**Males**	**Females**
Mean difference	-0.10	-0.36	0.38
T-statistic	-0.28	-0.81	0.59
Two-sided p-value	0.777	0.424	0.557
N	100	63	36

The t-tests in Table 8 provide no evidence to suggest that the Rejection levels to male and female responders differ significantly. Therefore, I fail to reject the null hypothesis for hypothesis 1 in this research. Individuals in the sample do not show different Rejection levels to offers from women than to offers from men.

5.3.2 Hypothesis 2

5.3.2.1 Correlation of control variables

The control variables in Table 7 show limited significance. Possibly, collinearity plays a role in the model. If this is the case, the coefficients and the significance could deviate greatly when one of the variables is excluded. Table 9 below displays the pairwise Pearson correlation coefficients for all independent variables in the model.

Table 9: Linear correlations between explanatory variables

	Age	**Edu**	**F_resp**	**Single**	**Timing**	**Dscore**	**F_prop**	**GSI**
Age	1							
Edu	0.439***	1						
F_resp	0.027	0.023	1					
Single	-0.378***	-0.268*	-0.129	1				
Timing	-0.022	-0.038	-0.102	0.314**	1			
Dscore	0.248*	0.307**	-0.018	-0.032	0.055	1		
F_prop	-0.006	0.086	-0.025	0.143	0.132	0.130	1	
GSI	0.119	0.038	-0.063	0.009	-0.001	0.368***	0.156	1

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Note. F_resp and F_prop represent Femaleresponder and Femaleproposer respectively

Even though none of the values are dangerously high, there might be some concerns about

the higher coefficients in Table 9 still. The highest five correlations, which are the only ones that exceed |0.3| and are significant at a 1% or a 0.1% significance level, are the couples Age-Education, Single-Age, Timing-Single, D-score-Education and Dscore-GSI.

Additional regressions provide a better insight in the effect of these correlations. These regressions are performed for the models (2), (4) and (5) of Table 7, in order to capture any changes to the effects I am interested in. For each of these three models, the variables Age, Education, Single and Timing are each excluded in one alternative model.
Table 13 in the Appendix D shows that the significance of some independent variables of Model (2) change based on inclusion or exclusion of other variables, as was predicted by their correlation. Age becomes significant at a 5% significance level when Education is excluded and the dummy variable Single is significant at a 5% significance level when Timing is excluded. The D-score is still not significant in any of the regressions. Most importantly, the interaction term between the D-scores and the Female proposer dummy is significant in all of the scenarios, either at a 5% or 10% significance level.
When doing the same for Model (4) of Table 7, including the interaction term between the Female proposer dummy variable and GSI, Table 14 shows this interaction term remains insignificant in all five regressions.
The alternative regressions for Figure (5) in Table 15 show similar results with regards to the variables Age, Education and Single. Strikingly, when excluding the variable Education or Timing from Model (2), the significance of the interaction term between D-score and Rejection level decreases from 5% to 10%. For Model (5), the significance of the interaction term in models 5C and 5D likewise decrease to become insignificant at a 10% significance level. Even though there is no strong reason to exclude these variables from the model, it is important to keep in mind that the main effect strongly depends on the model and the variables included in that model, and can even disappear completely.

5.3.2.2 Correlation of implicit and explicit associations

Table 7 shows that the D-score is significantly correlated with Rejection levels in Ultimatum Games with a female proposer, but the Gender Stereotype Index (GSI) score is not. Additionally, while the correlation between the D-score and the GSI is significant at a 0.1% significance level, the Pearson correlation coefficient of 0.368 from Table 9 is rather low. Before drawing any conclusions about this level of correlation, it must be considered if the correlation is linear. Figure 10 below illustrates the correlation between the D-scores and GSI of the subjects in the sample.

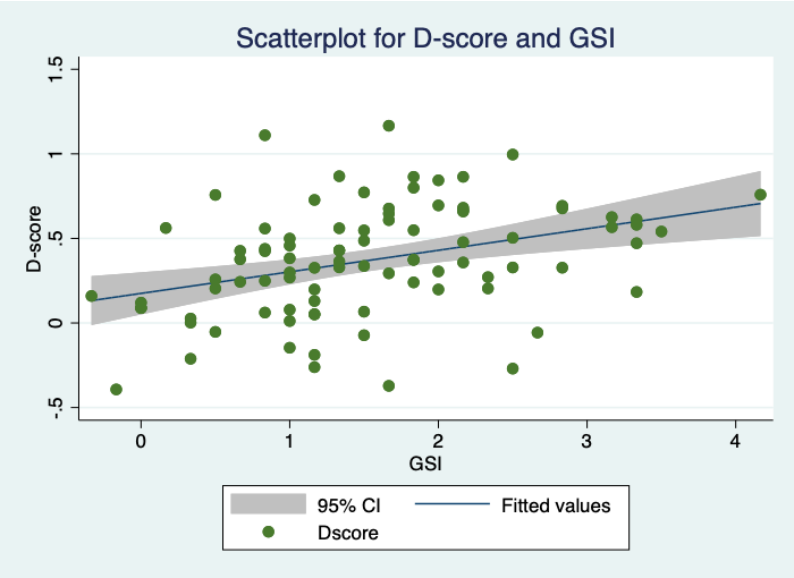

Figure 10: Scatterplot showing correlation of D-scores and GSI

Based on Figure 10, there is no reason to assume there is a non-linear relationship between the two variables. Assuming the Implicit Association Test is a reliable method for deducting the subjects' implicit associations, this limited correlation can thus have two reasons. Either individuals give answers that do not reflect their true associations in the explicit part of the survey on purpose, for example to give socially desirable offers, or they are unaware of their implicit associations. The discussion further elaborates on this limited correlation.

5.3.2.3 Chivalry and solidarity effects

The seemingly lower Rejection levels for female responders to female proposers, as identified in Table 4, is in line with the research by Eckel and Grossman (2001), who find the opposite of Solnick (2001). In their research, the authors find that offers from female proposers are significantly more likely to be accepted by both male and female responders. They label these effects the chivalry and the solidarity effect respectively. In the current study, the Mann-Whitney U tests and most of the regressions do not show these effects.

To test if the current research shows signs of the solidarity and/or the chivalry effect, I conduct the same regressions (1) and (3) from Table 7, including the D-score and the GSI respectively, for subsamples of male and female responders. These two models must be used, rather than than the models including interaction effects, because the main effect of the Female proposer dummy variable cannot be properly interpreted when the interaction

term is present. Table 16 in Appendix D illustrates the results.

The solidarity effect and the chivalry effect predict that in subsamples of females and males respectively, the dummy variable Female proposer is negative and significantly correlated with Rejection levels. This way, the sample would indicate that females and males respectively show lower and more social Rejection levels to female proposers. However, Table 16 shows that this effect is not significant in either of the models at a 10% significance level. Even though Table 4 seemed to show a same-sex effect, the Mann-Whitney U tests and Table 16 do not show any significance for either the solidarity or the chivalry effect. Potentially, the limited number of women in the sample cause these results.

5.3.2.4 Robustness tests

To check the normality of the residuals, linearity of the data, homoscedacity of the residuals and multicollinearity, I performed several tests. Some tests were done for all models in Table 7, while some are done for specific models to gain a deeper insight. The results from the tests can be found in Appendix E.

For hypothesis testing, normality of residuals is required to ensure the p-values are valid. the Shapiro-Wilk W test shows significant results for models (1), (3) and (4) of Table 7 at a 5% significance level. For these three models, I reject the null-hypothesis that the data are normally distributed. This hurts the interpretation of the variables in these models because the amount of predictability of the independent variables is not constant for all values of the dependent variable. Model (2) and Model (5) show insignificant results and do not seem to violate the assumption of normality.
To gain further insights into the distribution of Model (2) and Model (5), I perform additional tests. The Kernel density estimation is a non-parametric test that aims to estimate the probability density function. In this case, the test shows limited deviations from the normal density for models (2) and (5) of Table 7. Next to this, the cumulative density function of the normal distribution and the quantile function of the normal distribution show limited deviations from the normal distribution. Therefore, I do not reject the null hypothesis that there is no normal distribution.

The current research uses linear regressions, hereby assuming the relationship between the dependent and independent variables is linear. Component-plus-residual and augmented component-plus-residual plots can identify potential non-linearities for the variables of interest D-score and GSI. These plots aim to model the residuals of one predictor against

the dependent variable in the regression models. The line of best fit does not deviate much from the smoothed line in any of the figures, indicating there are no major issues with the linear fit.

The Breusch-Pagan test considers if heteroscedacity might play a role in any of the models of Table 7. It tests the null-hypothesis that the error variances are all equal. In none of the models, heteroscedacity seems to be an issue, as the Breusch-Pagan test shows insignificant p-values for all models. Additionally, the White test can sometimes be useful for identifying the type of hetereoscedacity if it occurs in the model. Models (2) and (4) show positive and significant results for the skewness at a 10% and 5% significance level respectively, suggesting the results here are skewed right. The Kurtosis measure suggests that the results of some of the models are more heavily tailed or have more outliers. However, the White test is an asymptotic test and is less reliable for smaller samples like in the current study. Looking at the Breusch-Pagan test and considering the limitations of the White test in this instance, the heteroscedacity is not a major reason for concern in the models.

The variance inflation factor test indicates that multicollinearity does not seem to be a problem in these models. An often used rule of thumb is that a variance inflation factor higher than 10 is problematic. The interaction terms with GSI show the highest variance inflation factors, yet the highest score of 7.13 is still below the rule of thumb and the multicollinearity is not considered a major threat to the models.

The described robustness tests do not highlight one major concern for the validity of the main models (2) and (5) from Table 7. Most striking are the results of the Skewness and Kurtosis as indicated by the White test, yet the validity of these findings is questionable in for the current data set and in the light of the measures of normality and the Breusch-Pagan test. Additionally, the models (1), (3) and (4) have issues with normality and should be interpreted with caution.

6 Discussion

This study aimed to answer the research question by testing two major hypotheses. Firstly, it tested if responders in the Ultimatum Game show higher rejection levels when the proposer is a woman than when the proposer is a man. Secondly, it tested if the implicit and explicit associations between gender and agentic or communal attributes are correlated with the rejection levels of responders to female proposers in the Ultimatum Game. The following subsections discuss the results along with their place in the literature, some limitations to the study, and finally suggestions for further research.

6.1 Interpretation of results

6.1.1 Results hypothesis 1

The Mann-Whitney U tests and parametric t-tests did not find evidence to support hypothesis 1, which states that individuals show higher rejection levels to offers from women than to offers from men, for the entire sample or for subsamples based on gender. These findings are not in line with the research of Solnick (2001) or the opposite results found by Eckel and Grossman (2001).

One explored explanation was that some individuals took a long time to answer the question and therefore maybe responded less intuitively. Kubota et al. (2013) used a timer in an experiment using the Ultimatum Game and the Implicit Association Test in a racial setting. However, as can be seen in Table 11 in Appendix C, no subsample based on time shows a significant difference between the rejection levels to female and to male proposers.

Other explanations for the inconclusive finding could be in the design of the online experiment. In the study by Eckel and Grossman, subjects were placed facing each other. A physical presence and non-anonymous setting is likely to trigger different behaviour. In the study by Solnick, subjects did not face each other but were placed in separate rooms. However, the experiment took place in a physical environment and instead of seeing the gender of their counterpart, the subjects saw a name that was easy to identify as male or female.
Firstly, perhaps the online environment in the current paper caused individuals to feel less motivated to try their hardest. The experimenter was unable to keep an eye on the subjects and they may fill it in more casually than they would in another setting.
Secondly, the online survey may give subjects a stronger sense of anonymity. The idea

that there is another person in the next room might create different emotional responses than being matched to a random and nameless person after the experiment.
Thirdly, without the name or physical presence of another person, the subjects may not believe they are actually playing against another real person.
Finally, the current study mentions the gender of the proposer in the Ultimatum Game explicitly while Eckel and Grossman used a face-to-face situation and Solnick wrote down names. It could be that subjects in the current research were more aware of the purpose of the study, and therefore answered differently than how they would behave outside the experiment. This effect can change the answers in either direction, as subjects can either give socially desirable answers or feel the urge to "help" the researcher by showing the effect he is looking for.

The Solnick and Eckel and Grossman studies bring forward that the higher anonymity (not facing each other) and the doubt if there really is another person in the Solnick study might cause the opposite effect from Eckel and Grossman. Following this line of reasoning, one would expect the current study to show even stronger results than Solnick. The lack of this effect can either mean that these are not the main reasons for the different results between the two aforementioned studies, or other factors play a role in the current study.

Evidently, the current study is unable to settle the debate on the ambiguous results obtained in previous research in either direction. Because of the different experimental design and sample characteristics, it also cannot reject either of the previous findings. Therefore, the current results are inconclusive in this debate.

6.1.2 Results hypothesis 2

Based on previous research about similar effects for racial bias (Kubota et al., 2013) and outcomes of negotiations based on perceived negotiation skills (Pardal et al., 2020), the current study aimed to expand the knowledge of the effect of implicit gender associations to the field of negotiations. Similar to Rudman and Glick (2001), the current study finds that in general, women are associated more strongly with communal terms and males with agentic terms. It also finds a correlation between the D-score, measured by the Implicit Association Test (IAT), and the rejection levels in the Ultimatum Game for games with female proposers. This means that responders with stronger implicit associations between females and communal or males and agentic words show higher rejection levels to female proposers.
In line with the current research, Pardal et al. (2020) found significant predictive power

of implicit gender associations in negotiations. However, the current study uses a vastly different approach. It uses the Ultimatum Game with a choice list rather than an elaborate negotiation game, the entire survey is done online and the Implicit Association Test is not done with "good" or "bad" negotiator words but rather with specific agentic and communal words to measure backlash behaviour. Additionally, to isolate the backlash effect, the current study focuses on the response to offers of female proposers rather than on final outcomes, as was done by Pardal et al. Clearly, all these design choices can have an impact on the final results. The current research should be seen as a simplified experimental approach to measure the backlash effect.
If the findings of the current research hold in a broader setting, it seems agentic women do face a backlash in negotiations with biased counterparts. Interestingly, in this setting with the Ultimatum Game, subjects are even willing to suffer financial consequences themselves. The backlash theory would argue this is done to "punish" women for not acting in accordance to their communal stereotype. Alternatively, one could argue that the observed behaviour is not necessarily a backlash exerted to agentic women, but rather a fear of "losing" to women for those subjects who hold more traditional views. Still, the difference between these two ideas is small and the results remains that women get worse results due to their counterparts' biased associations.

There is no evidence to support hypothesis 2b, which states that explicit associations between women and communal words or men and agentic words predict higher rejection levels to offers from women. Particularly relevant for this discussion is the question why the implicit and explicit gender associations have limited correlation, as this was found in many previous studies as well as the current one (Fazio & Olson, 2003; Greenwald et al., 1998; Rudman & Kilianski, 2000; Kawakami & Dovidio, 2001; Monteith et al., 2001; Ottaway et al., 2001). In general, there are two possible explanations for the misalignment between the implicit and explicit associations in the sample.
The first one is that individuals try to hide their real associations when they are asked about this directly in the Gender Stereotype Index. In this case, individuals are aware of their association. However, either because their rational self takes over or because they prefer to give socially desirable answers, they hide their implicit associations. If all individuals hide their implicit beliefs, I would expect the mean GSI to be close to zero. The mean for GSI is actually 1.53 and it has a broad range from -0.33 to 4.17, indicating a clear association between women and communal or men and agentic terms on average in the sample. Still, the limited correlation indicates that it is not necessarily the individuals who have the highest implicit bias that show the highest explicit bias when asked. Looking back at Figure 10, we see several instances of individuals with high D-scores that do not

indicate particularly high explicit associations. The opposite is observed as well, where even individuals who have a D-score below zero show rather high explicit associations. These findings suggest that the hiding behaviour does not hold for the entire sample, yet it could still be that part of the subjects try to hide their (biased) implicit beliefs.
The second explanation for the limited correlation between the implicit and explicit associations is that individuals may be genuinely unaware of their implicit associations. The question arises if the Ultimatum Game displays implicit or explicit associations. One could argue that the Ultimatum Game asks the individuals to explicitly show their preferences. However, due to the time limitation, it is not unlikely that individuals do not get the opportunity to think and they answer intuitively instead, hereby showing their implicit associations. Assuming the Ultimatum Game shows implicit rather than explicit associations, this would explain why the D-scores have explanatory power while the GSI does not in the games with female proposers.

Fazio and Olson (2003) state that motivation to engage and opportunity to deliberate determine the magnitude of the correlation between implicit and explicit associations. Specifically, if either one is relatively low at the time the explicit measure is done, then the implicit and explicit measures should correlate. This is in line with previous psychological and behavioural economic research on systems of two selves. Here, the rational self requires deliberate effort and if there is no motivation or opportunity to exert this deliberate effort, the explicit responses should not deviate from the answers of the automatic, or instinctive responses. The current research has no time limit, but subjects probably want to finish fast and have limited motivation to think long and hard about their answers. Still, the results shows significant differences between the implicit and explicit results, indicating they did put in more effort than their automatic first response when answering the explicit association questions. Additionally, Fazio and Olsen predict that when motivation and/or opportunity to deliberate is low, the behaviour is likely to be predicted by the implicit measure, which indeed holds in the current study.
Besides the limited correlation, I find that the implicit associations hold predictive power over negotiation behaviour in the Ultimatum Game while the explicit associations do not, hereby confirming hypothesis 2c. This could be the case because gender differences are a sensitive social topic (Kawakami & Dovidio, 2001), and the explicit associations subjects express therefore do not reflect their true preferences. Alternatively, due to the perceived time constraint on the Ultimatum Game, it could be that individuals answer too quickly and intuitively for their explicit correction to take place. Therefore, the higher explanatory power of the implicit measure may be due to the potential limited motivation to elaborate, the sensitivity of the topic and potential other factors that influenced the explicit measure.

6.1.3 Implications

There are no general differences between the rejection levels to male and to female responders in the sample. Interestingly however, those individuals that associate women with communal behaviour or men with agentic behaviour show lower rejection levels to female proposers. This implies that if women come across these biased individuals in negotiations, they are facing an uphill battle. Specifically, the counterpart is more likely to reject their offers because of their gender, even at their own personal costs. If the negotiations concern career advancement or wage negotiations, this finding can partially explain the glass ceiling effect and gender wage gap.
This backlash may not only lead to worse outcomes in that particular negotiation, but it can also demotivate women to pursue agentic negotiation tactics in future negotiations, creating a cycle in which agentic women conform to and therefore confirm the gender stereotype of communal behaviour. In order to break this cycle, the focus should not be on training women to be agentic, as this may create backlash. Instead, the focus should be on reducing implicit associations of others to create an environment in which agentic women do not receive backlash. The current research shows that identifying biased individuals may be difficult, since implicit and explicit gender associations do not always align. Institutes that aim to address this problem have to be aware of their inability to identify biased individuals and realize that even individuals who explicitly advocate gender equality may be part of the problem.

6.2 Limitations

All choices that were made for the current experiment in terms of data collection, experimental design and data analysis have their own limitations. I will discuss these three types of potential limitations.

The data collection was done by contacting the network of the author through the the applications Whatsapp and LinkedIn. Ideally, the sample would consist of a somewhat homogeneous group of individuals who actually perform negotiations in their day-to-day life. However, the current study was limited in time and access to such a group. Additionally, participation in the experiment was completely on a voluntary basis, potentially causing a significant self-selection bias. The possibility to earn a reward may have mediated this bias to some extent since it might appeal to individuals who otherwise would have passed

on the possibility to participate. However, most subjects will have participated out of goodwill.

Regarding the design of the experiment, I believe the three major potential limitations are the anonymous and online setting, the choice list design and the potential language barrier.
Most negotiations do not take place in an anonymous setting, let alone an online anonymous setting. Additionally, the Ultimatum Game provides a clean and easy setting to study behaviour, yet this is not very similar to negotiations in the real world either. These design choices served the purpose of this study but they may have influenced behaviour and limit the external validity of the findings. Next to this, subjects may not believe they are actually matched to another participant or that real money is involved in this anonymous setting. If this is the case, it becomes harder to spot any gender effect. Finally, the subjects explicitly read the gender of their counterpart. Because of this, they may have realized they were in a gender-related experiment and answered either socially desirable to appear pro-social or socially undesirable to "help" the author find the effect he is looking for.

The choice list design may have a big influence on the final results as well. In many instances, the Ultimatum Game is played pairwise and a subjects responds to one specific offer presented by another subject. In the current version of the game however, individuals indicated for what price they would no longer accept the offer amount by a random male or female proposer. This was mostly done for practicality and to avoid deception of subjects by providing false offer amounts. However, two potential problems arise in this approach.
Firstly, the choice list design may be hard for subjects to understand. Even though the text provides a thorough explanation, choosing the amount for which they no longer accept an offer is less straightforward than answering "yes" or "no" to a given offer. Additionally, in the online environment the experimenter could not monitor the subjects. If the subjects were in a hurry and read the instructions poorly, the less intuitive choice list design poses a bigger problem.
Secondly, subjects may not answer in a so-called "hot state". A hot state occurs when an emotion influences an individual's mental state. In the classic Ultimatum Game in which individuals are offered a specific amount, they might be annoyed by the low offer or excited about receiving some money right away. In the current study however, the individuals have to think about possible scenarios rather than experience one specific one, leaving them in a cold state.

The final potential design limitation relates to the language of the study. In the current research, I decided to include a Dutch translation because of the many expected Dutch participants, but to keep the Implicit Association Test (IAT) and the Gender Sterotype Index (GSI) questions in English for everyone. The major advantage is that no association with certain terms gets lost in translation between English and Dutch. However, it could be that the somewhat unintuitive English terms were difficult to grasp for a large part of my sample. Previous research found that the IAT is more sensitive for words with high familiarity, but the sensitivity to the IAT is not completely eliminated for lower familiarity terms (Ottaway et al., 2001). The gender association tests may provide somewhat weaker results if the subjects were less familiar with the terms.

Looking at the data analysis, I believe the chosen approach is generally justified. However, there is a critical point of limited explanatory power in all of the linear regression models, as illustrated by the low R-squared, adjusted R-squared and the F statistics. In a different and stronger model, the observed effect might be different. Additionally, the correlations between several independent variables influence the significance of the variables in the regression models. Besides the limited explanatory power, the significance of the relevant interaction variable is a cause for concern as well. In the main Model (2) of Table 7, this is significant at a 5% significance level, but this is less significant in the larger Model (5), which includes the Gender Stereotype Index.

6.3 Suggestions for further research

Using a simplified experimental approach, the current study shows that individuals with higher implicit associations between females and communal or male and agentic show higher rejection levels to female proposers. If this effect holds in other settings, it has major implications on gender in negotiations and on the best approach to achieve equality. One of the main goals of future research should be to test this specific backlash effect in more relevant settings and using more elaborate methods. One suggested approach, focused on the gender pay gap, is to test the effect in a within-company sample of individuals involved in salary negotiations. Firstly, it is important to test if some individuals negotiating on behalf of their firm implicitly associate women with communal words. If this is the case, it must be tested if higher implicit associations predict negotiation behaviour. This can either be done with the Ultimatum Game, as is done in the current study, or using a more elaborate negotiation game like in Pardal et al. (2020). If the observed effect from the current exploratory study also holds for the representative group of

subjects, this would strengthen the evidence for the backlash effect and provide a strong case for the need to reduce implicit associations to reduce the gender wage gap.

Alternatively, some findings in the current study provide other possible avenues of research to pursue. Firstly, the current study fails to find evidence that supports either the research by Solnick (2001) or by Eckel and Grossman (2001). However, there are too many differences between the studies to draw any conclusions about the differences in rejection levels to offers by male and female proposers. Future research could try to replicate results obtained by either study by following a similar approach. If there is more experimental evidence for this effect, future studies can start to build the bridge from lab experiments to real world scenarios, to see if the effect appears there as well. Next to this, a future study could try to find a similar effect as with negotiations in other business domains such as sales or leadership. Finally, future research could focus on the interplay between implicit and explicit gender bias. Fazio and Olson (2003) explain how implicit and explicit associations may align or not based on certain general characteristics. It is likely that there are certain situational cues that cause the implicit and explicit associations to misalign in a gender setting. Studying the different situational cues and their effect on implicit and explicit associations could help us better understand the ambiguous results obtained and increase our understanding of the factors that drive behaviour in gender interaction.

7 Summary and conclusion

The current research aimed to measure the influence of implicit and explicit gender bias on rejection levels in negotiations. The data was collected using an online survey experiment that subjects completed on their computer or laptop in 10-15 minutes. The survey contained general questions, an Ultimatum Game, an Implicit Association Test and explicit association questions. The sample was drawn from the author's personal network and consists of 100 subjects, 94 of whom completed the entire survey, and 90 subjects completed a valid Implicit Association Test. There are more men than women in the sample, the level of education is relatively high and the subject pool is rather young.
The rejection levels in negotiations are based on the rejection levels in the Ultimatum Game. This is deducted from a choice list in which individuals click the amount for which they would no longer accept the offer of the proposer, whose gender is known to the subject. The gender bias in this setting is based on the literature and measured as the implicit or explicit associations of subjects between gender and agentic or communal terms.
The Mann-Whitney U tests indicate that in general in this sample, subjects do not show higher rejection levels to offers from women than to offers from men. Higher implicit associations between men and agentic words or women and communal words seem to be correlated with higher rejection levels in negotiations with female proposers. The models find no correlation between explicit associations and the rejection levels to female proposers.

The most striking finding of this study is the correlation between the implicit associations of subjects and their rejection levels in the the Ultimatum Game when the proposer is female. Responders with a stronger implicit bias in the sample require higher offers from female proposers in order to accept than from male proposers, and they are willing to suffer financial consequences themselves by rejecting the offer.
This is in line with the ideas about the backlash effect discussed throughout this paper, where women receive a backlash or for acting in a stereotype-incongruent manner. The results from this study cannot be directly extrapolated and must be treated with caution. However, these results seem to indicate that agentic women might indeed face a backlash from subjects with a gender bias.

Another interesting finding is the insignificance of the explicit gender associations in negotiations with female proposers and the limited correlation between the implicit and explicit gender associations. Even though there was no evident pressure to give socially

desirable answers, individuals fail to align their implicit and explicit gender bias in the current experiment. This can have two major causes. Either the subjects answered dishonestly, perhaps to seem rational or unbiased, or they are unaware of their own gender associations.

As an answer to the research question, higher implicit gender bias seems to be correlated with higher rejection levels to female proposers. Combined with the information that the average implicit gender bias in the sample is positive, meaning individuals associate men with agentic and women with communal attributes, this is a reason for concern. As was explained in Section 1.2, negotiations are important for salaries and career advancement. If the findings of this study hold within a business, this could be one of the explanations for the glass ceiling effect and the gender wage gap. Additionally, campaigns focussed on "fixing the women" may be ineffective if training women to be agentic in negotiations only leads to more backlash from their counterparts. Instead, campaigns aimed at decreasing implicit gender bias may be a more suitable strategy for institutes aiming to improve gender equality.

All in all, this research on its own does not provide definitive findings on the role of implicit and explicit gender associations in a context of negotiations. As has been described in previous parts, some ambiguous results were found and limitations were addressed in the discussion. The main goal of this paper was to see if implicit and explicit gender associations hold any predictive power over negotiation behaviour with female counterparts. Whilst taking into account the specifics of this research, of the sample and its limitations, we can indeed say that there is reason to believe there is some predictive power in this implicit gender bias and this may be harmful to agentic women in negotiations.

References

Aguiar, F., Brañas-Garza, P., Cobo-Reyes, R., Jimenez, N., & Miller, L. M. (2009). Are women expected to be more generous? *Experimental Economics*, *12*(1), 93–98.

Amanatullah, E. T., & Morris, M. W. (2010). Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of personality and social psychology*, *98*(2), 256.

Amanatullah, E. T., & Tinsley, C. H. (2013). Punishing female negotiators for asserting too much... or not enough: Exploring why advocacy moderates backlash against assertive female negotiators. *Organizational Behavior and Human Decision Processes*, *120*(1), 110–122.

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

Babcock, L., & Laschever, S. (2009). *Women don't ask*. Princeton University Press.

Bakan, D. (1966). The duality of human existence: An essay on psychology and religion.

Bear, J. (2011). "passing the buck": Incongruence between gender role and topic leads to avoidance of negotiation. *Negotiation and Conflict Management Research*, *4*(1), 47–72.

Bear, J. B., & Babcock, L. (2012). Negotiation topic as a moderator of gender differences in negotiation. *Psychological science*, *23*(7), 743.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science*, *9*(3), 226–232.

Bolino, M. C., & Turnley, W. H. (2003). Counternormative impression management, likeability, and performance ratings: The use of intimidation in an organizational setting. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *24*(2), 237–250.

Bowles, H. R. (2013). Psychological perspectives on gender in negotiation. *The Sage handbook of gender and psychology*, 465–483.

Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and human decision Processes*, *103*(1), 84–103.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-role stereotypes: A current appraisal 1. *Journal of Social issues*, *28*(2), 59–78.

Butler, D., & Geis, F. L. (1990). Nonverbal affect responses to male and female leaders: implications for leadership evaluations. *Journal of Personality and social Psychol-*

ogy, *58*(1), 48.

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic perspectives*, *9*(2), 209–219.

Carpenter, T. P., Pogacar, R., Pullig, C., Kouril, M., Aguilar, S., LaBouff, J., ... Chakroff, A. (2019). Survey-software implicit association tests: A methodological and empirical analysis. *Behavior research methods*, *51*(5), 2194–2208.

Cherry, F., & Deaux, K. (1978). Fear of success versus fear of gender-inappropriate behavior. *Sex roles*, *4*(1), 97–101.

Chiu, C.-y., Hong, Y.-y., & Dweck, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of personality and social psychology*, *73*(1), 19.

Conley, T. D., Ziegler, A., & Moors, A. C. (2013). Backlash from the bedroom: Stigma mediates gender differences in acceptance of casual sex offers. *Psychology of Women Quarterly*, *37*(3), 392–407.

Costrich, N., Feinstein, J., Kidder, L., Marecek, J., & Pascale, L. (1975). When stereotypes hurt: Three studies of penalties for sex-role reversals. *Journal of Experimental Social Psychology*, *11*(6), 520–530.

Craver, C. B., & Barnes, D. W. (1998). Gender, risk taking, and negotiation performance. *Mich. J. Gender & L.*, *5*, 299.

Dannals, J. E., Zlatev, J. J., Halevy, N., & Neale, M. A. (2021). The dynamics of gender and alternatives in negotiation. *Journal of Applied Psychology*.

Desmarais, S., & Curtis, J. (1997). Gender and perceived pay entitlement: Testing for effects of experience with income. *Journal of Personality and Social Psychology*, *72*(1), 141.

Dittrich, M., Knabe, A., & Leipold, K. (2014). Gender differences in experimental wage negotiations. *Economic Inquiry*, *52*(2), 862–873.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of personality and social psychology*, *82*(1), 62.

Eagly, A. (1987). *Sex differences in social behavior: A social-role interpretation. hillsdale, nj: Lawrence earlbaum associates.* Inc.

Eagly, A., & Wood, W. (2012). Social role theory. in van lange, p., kruglanski, a., & higgins, et. *Handbook of theories in social psychology*, *2*, 458–476.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological bulletin*, *100*(3), 283.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological bulletin*, *111*(1), 3.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of personality and social psychology*,

46(4), 735.

Eagly, A. H., & Wood, W. (2016). Social role theory of sex differences. *The Wiley Blackwell encyclopedia of gender and sexuality studies*, 1–3.

Eckel, C., De Oliveira, A. C., & Grossman, P. J. (2008). Gender and negotiation in the small: are women (perceived to be) more cooperative than men? *Negotiation Journal*, *24*(4), 429–445.

Eckel, C. C., & Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. *Economic inquiry*, *39*(2), 171–188.

Fabre, E. F., Causse, M., Pesciarelli, F., & Cacciari, C. (2016). The responders' gender stereotypes modulate the strategic decision-making of proposers playing the ultimatum game. *Frontiers in psychology*, *7*, 12.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175–191.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of personality and social psychology*, *69*(6), 1013.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual review of psychology*, *54*(1), 297–327.

García-Gallego, A., Georgantzís, N., & Jaramillo-Gutiérrez, A. (2012). Gender differences in ultimatum games: Despite rather than due to risk attitudes. *Journal of Economic Behavior & Organization*, *83*(1), 42–49.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of" blind" auditions on female musicians. *American economic review*, *90*(4), 715–741.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*(6), 1464.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of personality and social psychology*, *85*(2), 197.

Greig, F. (2008). Propensity to negotiate and career advancement: Evidence from an investment bank that women are on a "slow elevator". *Negotiation Journal*, *24*(4), 495–508.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: reactions to women who succeed at male gender-typed tasks. *Journal of applied psychology*, *89*(3), 416.

Horner, M. S. (1972). Toward an understanding of achievement-related conflicts in women.

Journal of Social issues, *28*(2), 157–175.

Jackson, L. A., Gardner, P. D., & Sullivan, L. A. (1992). Explaining gender differences in self-pay expectations: Social comparison standards and perceptions of fair pay. *Journal of Applied Psychology*, *77*(5), 651.

Janoff-Bulman, R., & Wade, M. B. (1996). The dilemma of self-advocacy for women: Another case of blaming the victim? *Journal of Social and Clinical Psychology*, *15*(2), 143–152.

Jost, J. T. (1997). An experimental replication of the depressed-entitlement effect among women. *Psychology of Women Quarterly*, *21*(3), 387–393.

Kaman, V. S., & Hartel, C. E. (1994). Gender differences in anticipated pay negotiation strategies and outcomes. *Journal of Business and Psychology*, *9*(2), 183–197.

Kawakami, K., & Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin*, *27*(2), 212–225.

Koch, S. C. (2005). Evaluative affect display toward male and female leaders of task-oriented groups. *Small Group Research*, *36*(6), 678–703.

Kolb, D. M. (2009). Too bad for the women or does it have to be? gender and negotiation research over the past twenty-five years. *Negotiation Journal*, *25*(4), 515–531.

Kray, L. J., & Haselhuhn, M. P. (2007). Implicit negotiation beliefs and performance: Experimental and longitudinal evidence. *Journal of personality and social psychology*, *93*(1), 49.

Kray, L. J., & Thompson, L. (2004). Gender stereotypes and negotiation performance: An examination of theory and research. *Research in organizational behavior*, *26*, 103–182.

Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: gender stereotype confirmation and reactance in negotiations. *Journal of personality and social psychology*, *80*(6), 942.

Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The price of racial bias: Intergroup negotiations in the ultimatum game. *Psychological science*, *24*(12), 2498–2504.

Kugler, K. G., Reif, J. A., Kaschner, T., & Brodbeck, F. C. (2018). Gender differences in the initiation of negotiations: A meta-analysis. *Psychological bulletin*, *144*(2), 198.

Kulik, C. T., & Olekalns, M. (2012). Negotiating the gender divide: Lessons from the negotiation and organizational behavior literatures. *Journal of Management*, *38*(4), 1387–1415.

Major, B., McFarlin, D. B., & Gagnon, D. (1984). Overworked and underpaid: on the nature of gender differences in personal entitlement. *Journal of personality and social psychology*, *47*(6), 1399.

Marchetti, A., Castelli, I., Harlé, K. M., & Sanfey, A. G. (2011). Expectations and outcome: The role of proposer features in the ultimatum game. *Journal of economic psychology*, *32*(3), 446–449.

McGee, P., & Constantinides, S. (2013). Repeated play and gender in the ultimatum game. *The Journal of Socio-Economics*, *42*, 121–126.

Meyerson, D. E., & Fletcher, J. K. (2000). A modest manifesto for shattering the glass ceiling. *Harvard business review*, *78*(1), 126–136.

Miles, E. W., & LaSalle, M. M. (2008). Asymmetrical contextual ambiguity, negotiation self-efficacy, and negotiation performance. *International Journal of Conflict Management*.

Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition*, *19*(4), 395–417.

Moss-Racusin, C. A., & Rudman, L. A. (2010). Disruptions in women's self-promotion: The backlash avoidance model. *Psychology of women quarterly*, *34*(2), 186–202.

Nassar-McMillan, S. C., Lambert, R. G., & Hakim-Larson, J. (2011). Discrimination history, backlash fear, and ethnic identity among arab americans: Post-9/11 snapshots. *Journal of Multicultural Counseling and Development*, *39*(1), 38–47.

Oosterbeek, H., Sloof, R., & Van De Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental economics*, *7*(2), 171–188.

Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the implicit association test. *Social Cognition*, *19*(2), 97–144.

Pardal, V., Alger, M., & Latu, I. (2020). Implicit and explicit gender stereotypes at the bargaining table: Male counterparts' stereotypes predict women's lower performance in dyadic face-to-face negotiations. *Sex Roles*, *83*(5), 289–302.

Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly*, *32*(4), 406–413.

Richeson, J. A., & Ambady, N. (2001). Who's in charge? effects of situational roles on automatic gender bias. *Sex Roles*, *44*(9), 493–512.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of personality and social psychology*, *74*(3), 629.

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance. *Journal of personality and social*

psychology, *87*(2), 157.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of personality and social psychology*, *77*(5), 1004.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of social issues*, *57*(4), 743–762.

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and social psychology bulletin*, *26*(11), 1315–1328.

Small, D. A., Gelfand, M., Babcock, L., & Gettman, H. (2007). Who goes to the bargaining table? the influence of gender and framing on the initiation of negotiation. *Journal of personality and social psychology*, *93*(4), 600.

Solnick, S. J. (2001). Gender differences in the ultimatum game. *Economic Inquiry*, *39*(2), 189–200.

Solnick, S. J., & Schweitzer, M. E. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational behavior and human decision processes*, *79*(3), 199–215.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist*, *52*(6), 613.

Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex roles*, *41*(7), 509–528.

Stevens, C. K., Bavetta, A. G., & Gist, M. E. (1993). Gender differences in the acquisition of salary negotiation skills: the role of goals, self-efficacy, and perceived control. *Journal of Applied Psychology*, *78*(5), 723.

Stuhlmacher, A. F., & Walters, A. E. (1999). Gender differences in negotiation outcome: A meta-analysis. *Personnel Psychology*, *52*(3), 653–677.

Tellhed, U., & Björklund, F. (2011). Stereotype threat in salary negotiations is mediated by reservation salary. *Scandinavian Journal of Psychology*, *52*(2), 185–195.

Tinsley, C. H., Cheldelin, S. I., Schneider, A. K., & Amanatullah, E. T. (2009). Women at the bargaining table: Pitfalls and prospects. *Negotiation Journal*, *25*(2), 233–248.

Wade, M. E. (2001). Women and salary negotiation: The costs of self-advocacy. *Psychology of Women Quarterly*, *25*(1), 65–76.

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. In *Advances in experimental social psychology* (Vol. 46, pp. 55–123). Elsevier.

Appendix

A. Survey questions

Table 10: Survey questions

Type	**Text in survey**
Information	The survey software has detected that you are attempting to take this survey from an incompatible device. The survey contains questions that will only function correctly on a computer with a keyboard. Please open this survey from a computer with a keyboard.
Information	Dear participant, before starting the survey, please read the following. The following survey is conducted for research of a master Thesis at Erasmus University Rotterdam. Some more information about the survey is provided at the end. Do note the following: - You can contact me to ask questions about the survey at any time via email: 430613db@student.eur.nl - You can stop taking part in the survey without consequences - Results will only be presented at a group level and cannot be traced back to individuals - Data will be stored anonymously using password protection for a period up to 5 years after publication The data will be handled with care, the results are only used for the purpose of this thesis and you will remain anonymous. Do you consent to the use of the data from the survey for this purpose?
Multiple choice	Would you prefer the survey language to be in Dutch or in English? Please note, 1 part of the survey will be in English for everyone.

Information	Thank you very much for taking the time to fill in this survey! Completing the survey takes approximately 10-15 minutes. Please note that it is crucial for the outcomes to carefully read the instructions for the different parts of the survey. The survey consists of 4 parts: 1. general background questions about you 2. a game with potential prize money 3. an association test that requires speedy answers 4. an association question list The order may differ per participant. It is possible to win a maximum of 10 euros in this survey. Please note, only those individuals that fill in the entire survey and leave their contact details have a chance to win the prize money.
Open (number)	What is your age?
Multiple choice	What is your gender?
Multiple choice	What is your highest completed level of education?
Multiple choice	What is your country of origin?
Multiple choice	What is your marital status?
Multiple choice	How did you come into contact with this survey?

Information	You will now play the Ultimatum Game. You will play against a real participant in this same survey and you can earn money with this game. Please read the instructions carefully. the Ultimatum Game has 2 players, a proposer and a responder. The proposer receives 10 euros, the responder receives nothing. The proposer can decide to offer any part of the 10 euros to the responder, ranging from 0 to all of the 10 euros. The responder can either accept or reject the offer. If the responder rejects the offer, neither player gets any reward. If the responder accepts the offer, both players get the allocated amount.
Multiple choice	Do you understand the general rules of the Ultimatum game?

Information	You will first play in the role of responder, and afterwards in the role of proposer. You are randomly allocated to play against an anonymous male (female) proposer As a responder, you will be matched to a previous male (female) participant in this same experiment. You will not see the name of the participant, but do note that this is an actual person that makes a real offer. You will not see the amount the individual has offered you. Instead, you will be asked at what offer amount you will no longer accept the proposal. "I will click amount X, if I do not want to accept an offer of X or lower from this woman." If the real offer is higher than this price, the offer is accepted. If the real offer is equal or lower than this price, the offer is rejected. Therefore, your potential payoff depends on your answer
Multiple choice	Do you understand your role as a responder?
Information	Due to limited budget of this research, only a part of the subjects will be paid out afterwards. After the experiment closes, 5 couples (proposer and responder) will be randomly selected to win their allocated earnings. In order to receive the amount, you must leave your contact details at the end of the survey. The contact details will only be used in case you win the amount and will not be linked to any other answers to this survey than the Ultimatum Game
Multiple choice	Do you understand the payoff possibility?

information	Please note, the next question is the Ultimatum game. You will randomly be matched to a woman that also completes this survey. Please select the highest amount from the list where, if offered this amount, you no longer want to accept it. The response time is limited to 10 seconds
Multiple choice/choice list	For what amount will you no longer accept the offer from the male (female) participant?
Information	You will now play in the role of proposer. This means that you are endowed with 10 euros. You can choose to offer any part of the 10 euros to a responder, who will fill in this game as well. You will be randomly matched with an individual after the experiment. If the other player accepts your offer, you can both win the allocated amount If the other player rejects your offer, neither of you will get a reward.
Open (number)	Please indicate the amount you would like to offer to the other player (range 0.0-10.0)
Information	The next part of the survey contains several words that may be unfamiliar to you. If you do not understand the meaning of one or several of the words, an explanation and a dutch translation is provided below. If you know all the words, you can skip this explanation.

Information	independent (dutch: onafhankelijk): confident and free to do things without needing help from other people competitive (dutch: competitief): trying very hard to be better than others autonomous (dutch: autonoom): able to do things and make decisions without help from anyone else individual (dutch: individueel): a person considered separately rather than as part of a group hierarchical (dutch: hierarchisch): arranged in a hierarchy self-sufficient (dutch: zelfvoorzienend): able to do or produce everything without the help of other people communal (dutch: gemeenschappelijk): shared by, or for the use of, a number of people attached (dutch: gehecht/verbonden): forming part of something/liking someone or something cooperative (dutch: cooperatief): helpful by doing what you are asked to do together (dutch: samen): with each other kinship (dutch: verwantschap): the fact of being related in a family / feeling of being close to somebody because of similar origins or attitudes commitment (dutch: betrokkenheid): a promise to do something or behave in a certain way / the desire to work hard and give your time and energy to a job or activity

Information	The following task is about associations between words and gender. This is a speed-based task where you will have to click either the 'e' or the 'i' key on your keyboard for multiple associations in a row. More information is found on the next page. We consider some of the previously described words to be 'agentic' words and others to be 'communal' words. The division is as follows: Agentic terms: independent, competitive, autonomous, individual, hierarchical, and self-sufficient Communal terms: communal, attached, cooperative, together, kinship, and commitment
Other	All subjects complete 1 out of 4 randomized orders or the IAT, with the following terms: Gender Male: man, son, father, boy, uncle, grandfather, husband Female: woman, daughter, mother, girl, aunt, grandmother, wife Agentic versus communal Agentic: independent, competitive, autonomous, individual, hierarchical, and self-sufficient Communal: communal, attached, cooperative, together, kinship, and commitment
Information	The following questions are about associations of certain attributes with gender.

Likert scale	Do you associate this attribute more with men or women? (random order) Agentic, Independent, Competitive, Autonomous, Individual, Hierarchical, self-sufficient, Communal, Attached, Cooperative, Together, Kinship, Commitment
Information	Thank you very much for taking the time to fill in this survey! This survey is part of a Master Thesis study on gender associations in a basic negotiation setting. The study considers implicit and explicit associations for this purpose. For any questions about the research, about your results, or for any feedback or complaints, feel free to contact me via email: 430613db@student.eur.nl.
Open (optional)	In order to be able to win the monetary amount, please fill in your email account in the box below. In case you win, I will contact you with this information.

B. Power calculations

The ideal allocation of subjects between the two samples, male proposer and female proposer, depends on the relative cost of the different groups and the expected variance or standard deviation. This can be calculated with the following formula:

$$\frac{n_{maleproposer}}{n_{femaleproposer}} = \frac{\sigma_{maleproposer}}{\sigma_{femaleproposer}} \sqrt{\frac{p_{femaleproposer}}{p_{maleproposer}}}$$

The price is the same, regardless of the gender of the proposer. Because of this, the division of subjects between the two treatment groups should depend solely on the standard deviations in the two groups. In the study by Solnick (2001), the author does not directly mention the standard deviations. However, standard errors of 0.28 on average for both the sample with male and the sample with female proposers were found. The sample sizes for responders to male proposers and female proposers were 48 and 41 respectively. Using these numbers, the standard deviations of the sample means are calculated:

$$\sigma = Standarderror * \sqrt{n}$$
$$\sigma_{maleproposer} = 0.28 * \sqrt{48} = 1.9398969$$
$$\sigma_{femaleproposer} = 0.28 * \sqrt{41} = 1.79287479$$

By inserting the standard deviations calculated before, based on the research by Solnick, ideal allocation of subjects to the two groups is calculated:

$$\frac{n_{maleproposer}}{n_{femaleproposer}} = \frac{1.9398969}{1.79287479} \sqrt{\frac{10}{10}} = 1.08200356$$

Turned into percentage of the total sample, the following formula indicates what percentage of the total sample should be allocated to the male proposer sample:

$$\%n_{maleproposer} = \frac{1.08200356}{1.08200356+1} \approx 52\%$$

C. Additional tables regarding further analysis for hypothesis 1

Table 11: Mann-Whitney U tests on rejection levels to male and female proposers for subsamples on reaction time

	Prob > |z|	**Z**	**N**
subsample			
< 5 seconds	0.655	-0.447	4
< 6 seconds	0.445	0.764	8
< 7 seconds	0.571	0.566	12
< 8 seconds	0.438	0.776	15
< 9 seconds	0.552	0.595	26
< 10 seconds	0.681	0.411	39
< 12 seconds	0.948	0.065	61
< 14 seconds	0.896	0.131	76
< 16 seconds	0.955	-0.056	86
< 18 seconds	0.959	-0.052	91
< 20 seconds	0.729	-0.346	96

Table 12: Count of rejection level choices

	(1)	
	Rejection	
	Number of subjects	**Cumulative percentage**
0	10	10
0.5	6	16
1	8	24
1.5	1	25
2	6	31
2.5	7	38
3	8	46
3.5	8	54
4	14	68
4.5	13	81
5	12	93
5.5	2	95
6	5	100
Total	100	
N	100	

D. Additional tables regarding further analysis for hypothesis 2

Table 13: Alternative regressions on rejection level based on Table 7 Model (2)

	(2A)	**(2B)**	**(2C)**	**(2D)**	**(2E)**
Age	0.046*		0.054**	0.047*	0.030
	(0.027)		(0.026)	(0.027)	(0.026)
Education	0.310	0.484		0.307	0.230
	(0.317)	(0.303)		(0.315)	(0.318)
Femaleresponder	0.013	0.011	0.012	-0.001	-0.053
	(0.388)	(0.393)	(0.388)	(0.386)	(0.392)
Single	0.821*	0.576	0.758*	0.897**	
	(0.459)	(0.440)	(0.455)	(0.436)	
Timing	0.021	0.028	0.021		0.043
	(0.039)	(0.040)	(0.039)		(0.038)
Dscore	-1.021	-0.976	-0.856	-0.982	-0.980
	(0.835)	(0.844)	(0.818)	(0.828)	(0.846)
Femaleproposer	-0.959*	-1.053*	-0.901	-0.920	-0.900
	(0.576)	(0.580)	(0.573)	(0.569)	(0.583)
Femaleproposer xDscore	2.294**	2.512**	2.230*	2.235*	2.356**
	(1.154)	(1.159)	(1.152)	(1.144)	(1.169)
Constant	0.408	1.095	1.171	0.559	1.342
	(1.206)	(1.147)	(0.920)	(1.168)	(1.102)
N	89	89	89	89	89
R^2	0.138	0.108	0.128	0.135	0.104
Adj. R^2	0.052	0.031	0.052	0.060	0.026
F	1.601	1.398	1.694	1.804	1.337

Note. Standard errors in parentheses; * $p \leq 0.10$, ** $p \leq 0.05$.

Table 14: Alternative regressions on rejection level based on Table 7 Model (4)

	(4A)	**(4B)**	**(4C)**	**(4D)**	**(4E)**
Age	0.050*		0.060**	0.051*	0.033
	(0.028)		(0.026)	(0.028)	(0.027)
Education	0.317	0.531*		0.313	0.249
	(0.316)	(0.296)		(0.315)	(0.319)
Femaleresponder	0.013	0.009	0.020	-0.005	-0.026
	(0.389)	(0.394)	(0.389)	(0.385)	(0.394)
Single	0.859*	0.603	0.807*	0.914**	
	(0.455)	(0.437)	(0.452)	(0.438)	
Timing	0.019	0.025	0.018		0.037
	(0.038)	(0.039)	(0.038)		(0.038)
GSI	-0.117	-0.100	-0.126	-0.110	-0.123
	(0.327)	(0.331)	(0.327)	(0.326)	(0.332)
Femaleproposer	-0.506	-0.591	-0.455	-0.475	-0.420
	(0.736)	(0.744)	(0.734)	(0.730)	(0.746)
Femaleproposer x GSI	0.295	0.346	0.297	0.280	0.314
	(0.421)	(0.426)	(0.421)	(0.418)	(0.427)
Constamt	0.119	0.783	0.917	0.280	1.103
	(1.292)	(1.253)	(1.018)	(1.243)	(1.200)
N	94	94	94	94	94
R^2	0.099	0.065	0.088	0.096	0.061
Adj. R^2	0.014	-0.011	0.014	0.023	-0.015
F	1.164	0.854	1.186	1.308	0.797

Note. Standard errors in parentheses; * $p \leq 0.10$, ** $p \leq 0.05$.

Table 15: Alternative regressions on rejection level based on Table 7 Model (5)

	(5A)	**(5B)**	**(5C)**	**(5D)**	**(5E)**
Age	0.046		0.054**	0.047*	0.030
	(0.028)		(0.026)	(0.027)	(0.027)
Education	0.296	0.473		0.293	0.217
	(0.322)	(0.307)		(0.320)	(0.323)
Femaleresponder	0.023	0.025	0.020	0.007	-0.043
	(0.394)	(0.398)	(0.393)	(0.391)	(0.397)
Single	0.823*	0.579	0.763*	0.901**	
	(0.464)	(0.445)	(0.459)	(0.440)	
Timing	0.022	0.029	0.021		0.043
	(0.040)	(0.040)	(0.040)		(0.039)
Dscore	-0.873	-0.836	-0.699	-0.834	-0.835
	(0.874)	(0.883)	(0.852)	(0.867)	(0.885)
GSI	-0.226	-0.217	-0.251	-0.223	-0.222
	(0.343)	(0.346)	(0.341)	(0.341)	(0.347)
Femaleproposer	-1.279	-1.402*	-1.233	-1.224	-1.218
	(0.782)	(0.787)	(0.779)	(0.772)	(0.792)
Femaleproposer x Dscore	2.092*	2.270*	2.041	2.046	2.154*
	(1.243)	(1.252)	(1.240)	(1.235)	(1.259)
Femaleproposer x GSI	0.272	0.297	0.279	0.260	0.271
	(0.446)	(0.451)	(0.446)	(0.444)	(0.453)
Constant	0.704	1.367	1.463	0.859	1.634
	(1.305)	(1.255)	(1.010)	(1.269)	(1.211)
N	89	89	89	89	89
R^2	0.143	0.113	0.134	0.140	0.108
Adj. R^2	0.033	0.012	0.035	0.042	0.007
F	1.302	1.119	1.355	1.425	1.068

Standard errors in parentheses

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

Table 16: Regressions on rejection levels for models (1) and (3) from Table 7 for subsamples on gender

	(1) **Males**	**(1)** **Females**	**(3)** **Males**	**(3)** **Females**
Age	0.104***	-0.003	0.102***	-0.002
	(0.039)	(0.041)	(0.038)	(0.042)
Education	-0.446	0.781	-0.469	1.118**
	(0.470)	(0.472)	(0.437)	(0.514)
Single	1.045*	0.674	0.966*	0.899
	(0.603)	(0.806)	(0.562)	(0.812)
Timing	0.028	-0.071	0.034	-0.078
	(0.044)	(0.104)	(0.040)	(0.109)
Dscore	0.020	1.415		
	(0.789)	(1.041)		
Femaleproposer	0.424	-0.610	0.411	-0.564
	(0.481)	(0.661)	(0.470)	(0.664)
GSI			-0.001	0.534
			(0.272)	(0.339)
Constant	0.541	0.609	0.665	-0.796
	(1.550)	(2.194)	(1.503)	(2.333)
N	57	32	59	35
R^2	0.170	0.249	0.169	0.247
Adj. R^2	0.070	0.068	0.073	0.086
F	1.705	1.379	1.763	1.532

Note. Standard errors in parentheses; * $p \leq 0.10$, ** $p \leq 0.05$.

E. Robustness tests for models of Table 7

Measures of normality

Table 17: Shapiro-Wilk W tests for normality for the models from Table 7

	Observations	W	V	Z	Prob>z
Model (1)	89	0.968	2.408	1.937	0.026
Model (2)	89	0.978	1.624	1.069	0.143
Model (3)	94	0.965	2.729	2.219	0.013
Model (4)	94	0.969	2.425	1.958	0.025
Model (5)	89	0.980	1.518	0.920	0.179

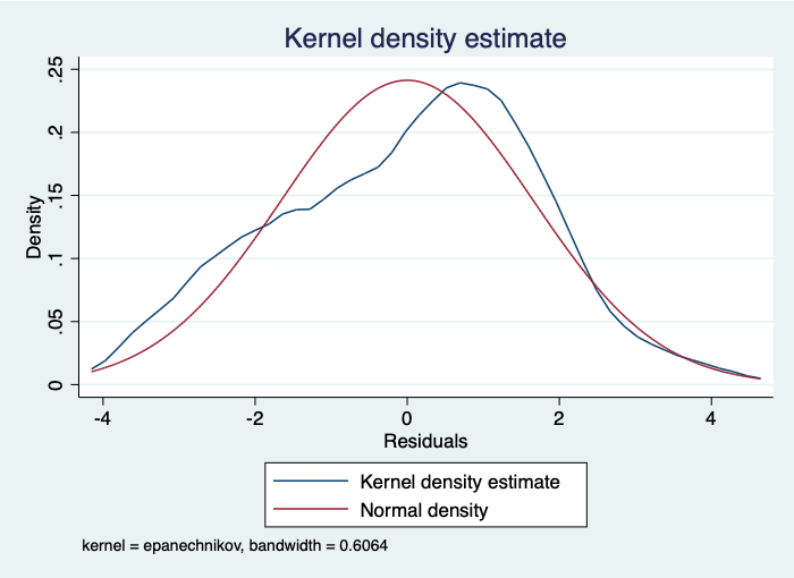

Figure 11: Kernell Density Estimate for Table 7 Model (2)

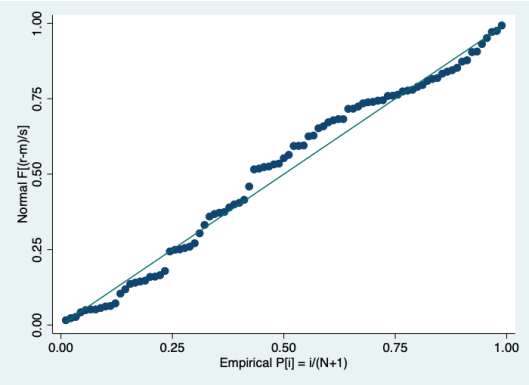

Figure 12: Cumulative density function of the normal distribution of Table 7 Model (2)

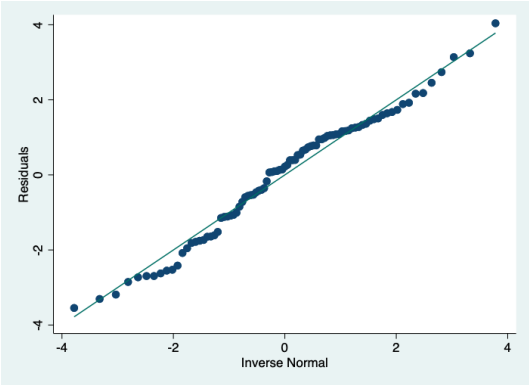

Figure 13: Quantile function of the normal distribution of Table 7 Model (2)

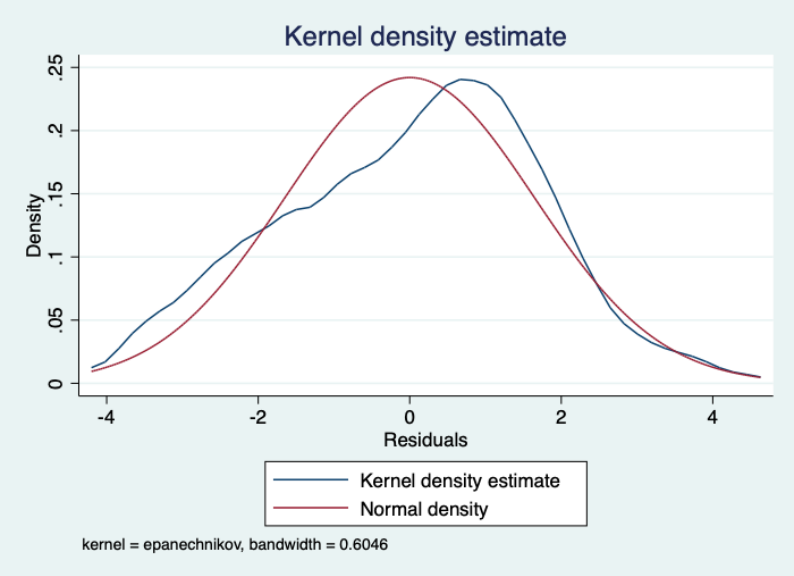

Figure 14: Kernell Density Estimate for Table 7 Model (5)

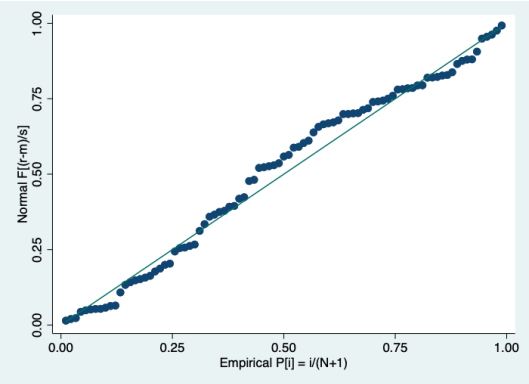

Figure 15: Cumulative density function of the normal distribution of Table 7 Model (5)

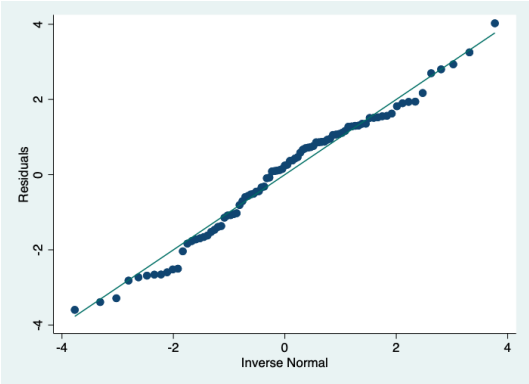

Figure 16: Quantile function of the normal distribution of Table 7 Model (5)

Measures of linearity

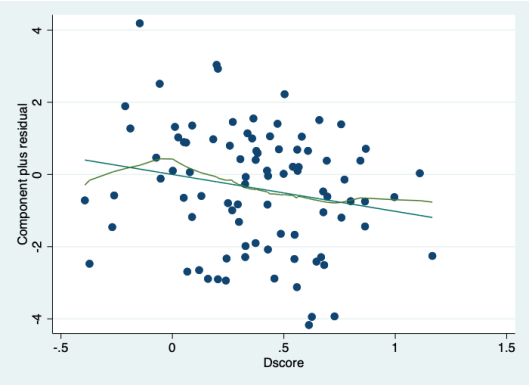

Figure 17: Component-plus-residual plot for Model (2) of Table 7

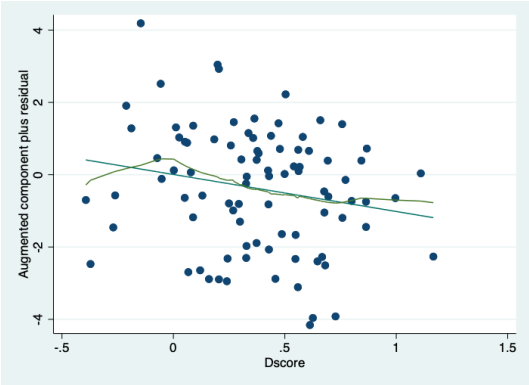

Figure 18: Augmented component-plus-residual plot for Model (2) of Table 7

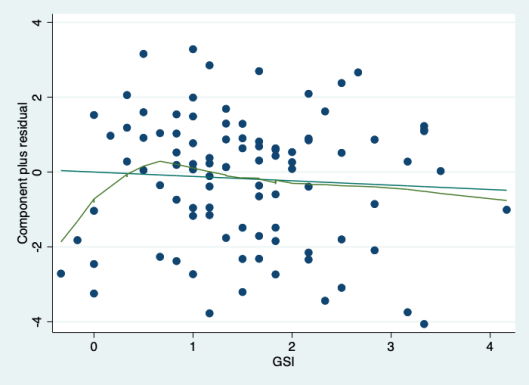

Figure 19: Component-plus-residual plot for Model (4) of Table 7

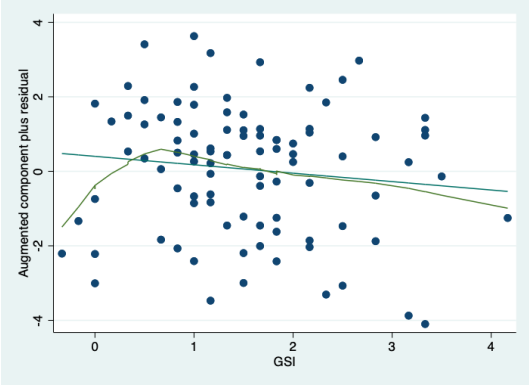

Figure 20: Augmented component-plus-residual plot for Model (4) of Table 7

Measures of homoscedacity of the residuals

Table 18: White/IM and Breusch Pagan test for models of Table 7

		(1)	(2)	(3)	(4)	(5)
Heteroscedacity	χ^2	32.02	38.90	37.45	44.97	53.53
	df	32	38	32	38	54
	p-value	0.466	0.474	0.233	0.203	0.493
Skewness	χ^2	11.49	13.45	11.80	17.57	15.97
	df	7	8	7	8	10
	p-value	0.119	0.097	0.105	0.025	0.101
Kurtosis	χ^2	4.83	2.95	4.97	3.94	2.45
	df	1	1	1	1	1
	p-value	0.028	0.086	0.026	0.047	0.117
BP test	χ^2	1.17	2.09	0.91	0.30	1.40
	p-value	0.279	0.148	0.341	0.582	0.236

Measures of Multicollinearity

Table 19: Variance Inflation Factor tests for models of Table 7

VIF	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Age	1.41	1.43	1.40	1.41	1.43
Education	1.34	1.35	1.28	1.28	1.37
Female responder	1.02	1.03	1.02	1.03	1.03
Single	1.37	1.37	1.31	1.31	1.37
Timing	1.14	1.15	1.09	1.10	1.15
Female proposer	1.06	2.45	1.08	3.95	4.43
Dscore	1.15	2.19			2.35
Female proposer x Dscore		3.76			4.28
GSI			1.05	2.65	2.94
Femaleproposer x GSI				6.21	7.13
Mean VIF	1.21	1.84	1.18	2.37	2.75