University of the Basque Country

Doctoral Thesis

Three Essays in Behavioral and Experimental Economics

Author:
Iñigo Hernández Arenaz

Advisor:
Nagore Iriberri

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

in the

Department of Foundations of Economic Analysis I
Faculty of Economics and Business

2017

Dedicado a
mi familia

Acknowledgments

I am indebted to so many people that I will probably forget to mention some. If you read this and you are not mentioned, it is probably my fault. So, let's begin with the acknowledgments.

First at all, I must acknowledge my advisor, Nagore Iriberri for all the time, patience, and effort she has put into turning me into a serious researcher. At least she has tried hard! Apart from our academic relationship (which is best reflected in chapters two and three of this dissertation but also in my "solo" paper), I have enjoyed her company and shared lots of good (and some not so good) times. It has truly been a pleasure.

But of course, there are many others who also deserve to be mentioned in regard to my academic performance (and mental health).

I must thank all the faculty at *Sarriko* for their support and help (and financial aid). I landed in Bilbao for first time four years ago and met a wonderful group of people who have facilitated my stay here incredibly. Special thanks go out to Elena Iñarra, Jaromír Kováříck, Norma Olaizola, and Marta Escapa. Of course, thanks are also due to the secretary of the department, Esther Vidal, who has suffered from my ineptitude at paperwork but has always shown her best face nevertheless. Nor can I forget all my partners on this 4-year adventure from the *zulo* (to both *I* and *II*).

I must also thank the Faculty of Economics of the *Universidad Pública de Navarra*, where I have spent a lot of time both as an undergraduate and as a graduate student interacting with its staff, who have been fantastic at both the professional as the personal level. Special thanks are due here to José Enrique Galdón for encouraging me to continue studying when I finished my undergraduate studies and was feeling lost, and to Jorge Alcalde Unzu for encouraging me to pursue my PhD in Bilbao at a very difficult time for me personally. Thanks also to my "*other*" partners on this 4-year adventure from the *zulo III* at the *Universidad Pública de Navarra*.

I must also say a few special words for Friederike Mengel and Patrick Nolen who were always available to help and spend time with me during my stay at *Essex University* and have remained so since I left there. Also thanks to Subhasish Chowdhury from the *University of East Anglia* who, together with Friederike Mengel, agreed to write the external report for this thesis.

I also wish to thank the members of my thesis committee (Antonio Cabrales, Javier Gardeazabal, and Pedro Rey-Biel) for agreeing to take part in this event and for always being supportive and kind to me. A special acknowledgment here is due to Pedro Rey-Biel who introduced me for the first time to the world of behavioral and experimental economics during my master's courses at the *Universitat Autònoma de Barcelona*.

But obviously the main reason I have made it this far is my family. They have shaped my character and my critical spirit, made me appreciate the value of hard work and taught me not to give up without a fight. But most importantly, they have always supported me with love and patience. To all of you (wherever you are), thank you so much!! This thesis is also yours.

And last but not least, I have also to thank to my *cuadrilla* (*¡¡La vida es maravillosa!!*) because, even though they tend to distract me from more "productive" activities, they have always been there to cheer me up when needed and have always believed in my ability to pursue this thesis.

¡Muchas gracias a todos!
Mila esker denori!!

Contents

List of Figures

List of Tables

Summary

This dissertation presents three essays which represent my main work during the last 4 (3 and a half to be fair) years. All three address important issues in Economics from a behavioral and experimental perspective. In fact, this dissertation makes use of the three main tools for the study of behavioral economics: Economic modeling, laboratory experiments, and field data. In addition, all three essays are connected in that they deal with important features of the labor market which are related to gender issues and seek to understand the gender wage gap, which is found still to exist and to be of a considerably size.

Chapter 1 explores how stereotypes affect self-selection into different tournaments, an issue which lies at the heart of segregation. I first present a theoretical model in which agents have imperfect self-knowledge about their abilities and must self-select into one of two tournaments: one high-paying and the other low-paying. The model shows that the existence of negative (positive) stereotypes generates underrepresentation (overrepresentation) of stereotyped social groups in the high-paying tournament even when the stereotype is false. Importantly, this model explicitly shows that the main channel driving segregation is that of *self-stereotyping*, a phenomenon that has already been documented in experimental work –mainly in social psychology– but for which a formal formulation was lacking. The second part of this chapter is devoted to experimentally testing the model's predictions by transferring the theoretical setting to the lab. To that end, I ran a lab experiment in which subjects first have to perform a real effort task –which is novel in the literature– and must then choose whether to participate in a high-paying tournament or a low-paying one in which the chances of getting the prize depend on how well they performed in the real effort task. In addition subjects have to self-assess with regard to their performance in the real effort task based on a noisy but informative signal. The key feature of this design is that the real effort task shows a high degree of between-subjects heterogeneity as to its believed gender nature. This enables me to look at the effects of the perceived gender nature of the real effort task on self-assessment and on the tournament in which subjects finally choose to participate. The results are closely in line with the predictions of the model, and show that different perceptions as regard to the gender nature of the real effort task yield different gender self-assessment gaps, which give shape to eventual gender-based segregation between tournaments.

It could be said that watching TV is a waste of time, especially given the general low quality of the shows which are currently being broadcast. Most of the time I agree.[1] However, from time to time watching TV provides an opportunity to gather very interesting data that would otherwise be almost impossible to obtain. The outcome of my spending time in front of the TV, usually after lunch, is presented in Chapter 2. In this chapter I, along with N. Iriberri, study gender differences and gender interaction effects in bargaining in a natural experiment. Data from the TV show "*Negocia como puedas*" offers a unique opportunity to observe both bargaining outcomes and behavior, with sizable stakes that we could hardly implement in the laboratory. We find that the matching between male proposers (strong position) and female responders (weak position) is different from all other matchings in that it is the matching in which the strong party gets the most. No differences are observed in opening offers, but women in the weak position demand less from men than from women. This differential behavior yields bargaining outcomes that are more favorable for men and less favorable for women when male proposers encounter female responders.

In Chapter 3 N. Iriberri and I, inspired by the results in Chapter 2, dig deeper into the issue of gender and gender interaction effects when bargaining. In particular, given that the setting of the TV show in Chapter 2 is characterized by the existence of asymmetries in empowerment, entitlement, and information, and that the results found clearly indicate that gender and strength of bargaining position are interlinked, we wondered whether the existence of different sources of asymmetries might have any effect in the appearance of gender and/or gender interaction effects. Notice that most real-life negotiations are characterized by some kind of asymmetry of the said types. We believe that accounting for the effects of asymmetries therefore deserves attention in its own right. Thus, we carried out a laboratory study to look at gender differences and gender interaction effects in asymmetric bargaining situations. In a symmetric situation, used as benchmark, we find no gender differences or gender interaction effects, and find that 50:50 split of the pie is the norm, as expected. In asymmetric situations, when one bargaining party is made stronger than the other (through empowerment, entitlement or information), we find that gender differences and gender interaction effects are substantial in the first half of the experiment (first 5 periods) but vanish in the second half of the experiment (last 5 periods). Remarkably, we find substantial interaction effects in the informational asymmetry treatment, where only the strong party knows the size of the pie, similar to those found in Chapter 2: the matching between a male holding the strong position and a female holding the weak position is different from all other matchings and is the matching in which the strong party gets the most. Consistently with the results in Chapter 2, we also find that the main driver for this is women in this treatment demanding less from men.

This study could be considered as a starting point for more ambitious projects be-

[1] As a personal disclaimer, notice that I refer here to TV shows and not to broadcasts in general. In particular, there are some good films shown on TV that deserve to be watched. A recent good example is *Loreak* by Goenaga and Garaño (2014).

cause we not only show that gender and gender interaction effects are sensitive to the existence of asymmetries but also find two intriguing results that should be looked at closely. The first is the one mentioned above, which relates to the importance of experience/learning effects. Our results showed that this effect exists but do not explain why. The second point is that, as usual in the literature, we find a substantial deadline effect with many deals being reached close to the deadline. We show that this deadline effect is not free from consequences and in particular that the deals reached just before the deadline have features similar to ultimatum games making them systematically different from those reached further from the deadline. As this seems more of an artifact given by the design of the experiment, ideally we would like to get rid of it.

Resumen Extendido

Esta tesis presenta tres trabajos que abordan temas importantes en Economía desde el punto de vista de la Economía del comportamiento y la Economía experimental. De hecho, esta disertación hace uso de las tres herramientas principales para el estudio de la economía del comportamiento: el modelaje económico, experimentos de laboratorio y el uso de datos de campos. Adicionalmente, los tres ensayos están conectados entre sí en que todos ellos tratan con facetas importantes del mercado de trabajo, relacionados con género y que pueden ser utilizados para entender la brecha salarial que aún a día de hoy está presente y que representa una magnitud importante.

Capítulo 1: Stereotypes and Tournament Self-Selection: A Theoretical and Experimental Approach

El capítulo 1, explora cómo los estereotipos afectan a la auto-selección en distintos torneos perfectamente discriminatorios, un aspecto que se erige como uno de los aspectos claves que dan forma a la segregación laboral y educativa.

Los psicólogos sociales han mostrado que los estereotipos juegan un papel importante a la hora de interpretar los resultados en distintas tareas así como en dar forma a nuestras propias habilidades (véase Lenney (1977); Beyer (1990); Beyer and Bowden (1997); Pomerantz et al. (2002); Ehrlinger and Dunning (2003)). La RAE, en su primera aceptación define "estereotipo" cómo "*Imagen o idea aceptada comúnmente por un grupo o sociedad con carácter inmutable*". El periodista Walter Lippmann –quien primeramente uso el término *estereotipo* en su uso moderno– captura esta idea de estereotipos como puntos de referencia en su influencial libro *Public Opinion* (1922) cunado dice

> "*Las más sutiles y persuasivas de todas las influencias son las que crean y mantienen el repertorio de los estereotipos. Se nos habla del mundo antes de verlo. Imaginamos la mayoría de las cosas antes de experimentarlas*".

En un mundo incierto, evaluar las habilidades es un paso importante en procesos de decisión en cuanto a elecciones educativas y laborales, los cuales representan dos ámbitos importantes en el estudio de la Economía. Desde edades tempranas los niños deben elegir entre itinerarios de ciencias o letras basándose en señales imperfectas sobre sus habilidades, como lo son las notas, en un entorno social que presupone que los niños son mejores en ciencias y las niñas en letras. Esta elección ha mostrado que es un determinante importante en la educación futura así como en los logros en el mercado laboral (Paglin and Rufolo (1990)). Posteriormente, ya en el mercado de trabajo, siguen existiendo ideas preconcebidas sobre diferencias de género incluso entre aquellos individuos que tienen una misma formación educativa (ej. toma de decisiones estratégicas; Atwater et al. (2004)). Estas ideas preconcebidas, o estereotipos, se ha comprobado que están relacionados con cómo hombres y mujeres se ordenan en el mercado de trabajo (Cejka and Eagly (1999); Fernandez and Friedrich (2011); Barbulescu and Bidwell (2013); Leslie et al. (2015)).

A pesar del impacto que los estereotipos tienen en decisiones del tipo mencionado arriba en las que la auto-evaluación es la clave, muy poco trabajo se ha desarrollado en integrar el concepto de estereotipo en el modelaje económico.[2] Esta es la contribución central del capítulo 1: modelar y evaluar los efectos de los estereotipos en situaciones estratégicas que representen situaciones económicamente relevantes.

Para ello, en la primera parte del capítulo 1, presento un modelo teórico. Éste modelo considera una masa de agentes que, simultáneamente, deben auto-seleccionarse en uno de entre dos torneos mutuamente excluyentes. Estos torneos son perfectamente discriminatorios, es decir, los ganadores en un torneo son nombrados únicamente en base de sus habilidades reales en realizar una tarea y sólo pueden optar al premio aquellos que deciden participar en ese torneo en concreto. Estos torneos difieren entre sí en el premio que se otorga a los ganadores: uno paga más que el otro. Sin embargo, los agentes tienen un conocimiento imperfecto de sí mismos tal que no conocen su habilidad real en la tarea a desarrollar sino que observan señales informativas (pero no perfectamente correlacionadas) sobre su habilidad. Por tanto, los agentes realmente encaran dos problemas distintos. Primero, los agentes deben evaluarse a sí mismos basándose en la señal observada. Es decir, deben generar creencias sobre cómo de aptos son en la tarea que les ocupa. Segundo, deben afrontar consideraciones estratégicas pues deben de estimar el comportamiento del resto de agentes para así estimar cómo de competitivo va a ser cada torneo y, por tanto, la habilidad mínima necesaria para ganar un premio en cada uno de los torneos.

Los estereotipos son introducidos en el modelo asumiendo que la población entera puede ser dividida en dos *grupos sociales* diferenciados (hombres/mujeres, negros/no negros, ...). Unos de estos grupos sociales es objeto de un estereotipo negativo (positivo), el cuál dice que, en media, los agentes del grupo social estereotipado son menos

[2]Una notable excepción es la línea iniciada por Akerlof and Kranton (2000). El ensayo desarrollado en este capítulo 1 se enfoca en el análisis de estereotipos *descriptivos*, en contra a los *normativos* los cuales representan el núcleo delo trabajo de Akerlof and Kranton (2000). La literatura en *discriminación estadística* (Phelps (1972); Arrow (1973); Coate and Loury (1993)), cómo en el presente ensayo, también usa la noción de estereotipos descriptivos. Sin embargo, a diferencia del ensayo que nos ocupa, los estereotipos en los modelos de discriminación estadística no afecta la evaluación propia sino la evaluación realizada por terceras personas. En este capítulo de la tesis muestro que la existencia de estereotipos es suficiente para generar segregación incluso si no hay discriminación de ningún tipo.

(más) hábiles que los no estereotipados. Esta creencia sobre la existencia de una diferencia *ex-ante*, toma la forma de un desplazamiento en la distribución de habilidades moviendo su valor medio. Importante, el estereotipo es conocimiento común. Este supuesto, aunque fuerte, está dado por la propia definición del concepto de estereotipo y respaldado por hallazgos empíricos en la psicología social (Frome and Eccles (1998); Furnham and Gasson (1998); Furnham et al. (2002); Petrides et al. (2004); Nosek et al. (2009))

La existencia de estereotipos en este entorno tiene dos efectos. Primero, afecta al razonamiento estratégico *de todos los agentes* al incorporar sus mensajes en sus creencias (*priors*) sobre como la habilidad está distribuida en el conjunto de la población. Segundo, y más importante, la existencia de estereotipos genera diferencias en cómo se encara el problema de conocimiento imperfecto lo que resulta, a través de un proceso de inferencia Bayesiana, en el surgimiento de *auto-estereotipación*: una gente estereotipado y uno no estereotipado observando *la misma* señal sobre su habilidad, diferirán en la estimación sobre sus habilidades reales en el sentido impuesto por el estereotipo. Este efecto es altamente consistente con los hallazgos documentados principalmente en psicología social (Lenney (1977); Chatard et al. (2007)) pero también en economía (Coffman (2014)).

El modelo concluye que este efecto de auto-estereotipación tiene importantes efectos en cómo los agentes se comportan cuando tienen que elegir en qué torneo competir y las posibilidades de obtener un premio están determinadas *únicamente* por su ranking dentro de cada torneo. En esta línea Lazear and Rosen (1981) ya avanzaban que, "*en el mundo real, donde hay población heterogénea, los participantes en el mercado se ordenan en distintos tipos de torneos. Hay jugadores que se saben de mayor calidad a priori y participarán en torneos con mayores premios en juego*" (n. 5).

Cuando el juego Bayesiano descrito arriba es resuelto, las diferencias en auto-evaluación se trasladan en diferencias en las estrategias de equilibrio seguidas por agentes estereotipados y no estereotipados. En particular, habrá un conjunto de señales para las cuales los agentes estereotipados y no estereotipados se comportarán de forma distinta: si el estereotipo es negativo (positivo), para ese conjunto de señales los agentes estereotipados decidirán participar en el torneo de pagos bajos (altos) mientras que los no estereotipados lo harán en el de pagos altos (bajos). Si el estereotipo resulta ser falso –tal que las diferencias en comportamiento no pueden ser justificadas– el modelo predice que existirá un segregación entre torneos que no será óptima, tal que los agentes que ex-ante son percibidos como mejores estarán sobrerrepresentados en el torneo de pagos altos. Es más, este problema de representatividad se dará tanto a nivel de participación (quién participa en qué torneo), como a nivel de ganadores de los premios (en la distribución de premios entre grupos sociales). Una implicación directa de todo esto es que, en el mundo real, donde muchas decisiones educativas y laborales implican elegir entre distintos entornos competitivos con distintos pagos potenciales, distintos grupos sociales se ordenarán en el mercado de forma diferente simplemente por la existencia de estereotipos. De particular interés en este aspecto resulta la segregación laboral por sexos, puesto que numerosos estudios muestran que hombres y mujeres se ordenan de forma muy diferente en el mercado laboral (Fernandez-Mateo and Fernandez (2016); Card et al. (2016)) y que esta diferencia en el comportamiento explica una parte substancial de la brecha salarial (Macpherson and Hirsch (1995); Blau and Kahn (1997); Bertrand and Hallock (2001); Bayard et al. (2003); O'Neill (2003); Ludsteck (2014)).

La segunda parte de este primer capítulo, está dedicada a testar las implicaciones del modelo teórico en el laboratorio y, en particular, la existencia de la auto-estereotipación y de sus efectos sobre el comportamiento. Para ello, hago uso de una tarea de esfuerzo real que muestra ser neutral en cuanto al rendimiento de hombres y mujeres, pero altamente heterogénea en cuanto a las creencias idiosincráticas de los participantes en el experimento con respecto a si en la tarea lo hacen, en media, mejor los hombres o las mujeres. La estrategia de identificación pasa primordialmente por explotar la variación en estas creencias sobre qué género es mejor en la tarea de esfuerzo real para ver como esas percepciones afectan a la auto-evaluación en la habilidad y en la decisión de participar en un torneo con pagos altos o en otro con pagos bajos.

El grueso del experimento consiste en lo siguiente. Primero los sujetos realizan una tarea de esfuerzo real 14 veces sin obtener, en ningún momento, ninguna información sobre su rendimiento. Segundo, se le presenta a cada sujeto su puntuación en una de las 14 repeticiones de la tarea así como un gráfico que muestra cómo se ha distribuido la puntuación media durante las 14 repeticiones en esa sesión. Basándose en esa información, los sujetos deben estimar cuál es su puntuación media durante las 14 repeticiones (su habilidad). Tercero, los agentes deben elegir entre participar en el torneo A, que paga mucho a los ganadores, o en el torneo B, que paga menos. Ambos torneos tienen un número limitado de premios. Los premios se otorgan en base al ranking de habilidades (puntuación media en la tarea de esfuerzo real) dentro de cada torneo. Cuarto, se elicitan creencias individuales sobre la naturaleza de género de la tarea: qué género lo ha hecho mejor en media durante la tarea de esfuerzo real y cuánto mejor.

Los resultados experimentales están en línea con las predicciones del modelo teórico. Las percepciones que cada sujeto tiene sobre la naturaleza de género de la tarea inicial afecta directamente a la auto-evaluación: Para aquellos que afirman que la tarea es neutral en género no hay diferencias en auto-evaluaciones, mientras que entre aquellos que afirman que la tarea es femenina (masculina) las mujeres (hombres), para una misma señal observada, estiman mayores niveles de habilidad que los hombres (mujeres). Si miramos al rol jugado por las percepciones en la elección entre torneos, no se encuentran diferencias de género en la propensión a elegir el torneo de pagos altos (A) entre aquellos que dicen que la tarea es neutral en género, pero desviaciones de esta percepción de neutralidad lleva a la aparición de diferencias de género en la elección del torneo a participar. Más importante aún, y apoyando el principal canal de la auto-estereotipación propuesto por la teoría, estas diferencias en la elección de torneos pueden ser explicadas por las diferencias en auto-evaluación. Por último, los resultados también muestran que estas diferencias en auto-evaluación tienen efectos directos en la probabilidad de ganar el premio alto.

CAPÍTULO 2: WOMEN ASK FOR LESS (ONLY FROM MEN): EVIDENCE FROM ALTERNATING-OFFER BARGAINING IN THE FIELD

Aunque la brecha salarial entre hombres y mujeres ha sido un importante objeto de estudio en economía, las aproximaciones clásicas al problema no han sido capaces de explicarla satisfactoriamente (Blau and Kahn (2000)).

Las diferencias de género en negociación han sido planteadas como una alternativa. Los salarios iniciales son, en muchos casos, el producto de negociaciones bilaterales. El influencial libro de Linda Babcock y Sara Laschever "Women don't Ask: Negotiation and the gender divide" revela importantes diferencias de género en la propensión a negociar. Un estudio mencionado en este libro indica que entre los graduados de la Carnegie Mellon University, el 57% de los hombres negociaron sus salarios iniciales mientras que entre las mujeres esta cifra caía al 8%. Además, los salarios también se ven afectados por negociaciones que se van produciendo a lo largo de la vida laboral, ej. incrementos salariales. Si las mujeres son menos propensas a negociar sus salarios, esto iría claramente en el sentido de explicar la brecha salarial (Azmat and Petrongolo (2014); Card et al. (2016)).

Un concurso de televisión nos proporciona una oportunidad única de observar los resultados de negociaciones en una situación real con importantes pagos en juego. En el concurso (*Negocia como puedas*, emitido en Cuatro durante el verano del 2013), se le hace una pregunta al concursante (*Proponente*) y se le da una cantidad de dinero que puede variar entre 100 euros y 1600 euros. Sin embargo, éste no puede dar una respuesta a la pregunta directamente, sino que tiene tres minutos para encontrar a alguien que de la respuesta por él (*Respondedor*) y negociar un precio para esta respuesta a través de una negociación de ofertas alternativas. En una ronda de negociación típica, el proponente hace una oferta inicial, la cuál puede ser aceptada o rechazada por el respondedor. Si el respondedor rechaza la oferta, entonces puede hacer una demanda la cual puede ser aceptada o rechazada por el proponente. La negociación puede alargarse cualquier número de rondas dentro del límite de los tres minutos. Si la respuesta proporcionada por el respondedor es correcta, entonces la división del dinero acordada en la negociación se lleva a cabo.

En este capítulo estudiamos tanto diferencias de género tanto efectos de interacción de género en la negociación tanto en el resultado final obtenido, cómo en el comportamiento durante la negociación.

Con respecto a los resultados obtenidos de la negociación, encontramos que las negociaciones entre un hombre proponente y una mujer respondedora es el único emparejamiento que se diferencia del resto de posibles emparejamientos: son el emparejamiento en el que el proponente obtiene mayores ganancias y en el que el respondedor obtiene menores ganancias. El resto de emparejamientos posibles (hombre con hombre, mujer con mujer y mujer proponente con hombre respondedor), son indistinguibles entre sí. Al explicar los detalles de nuestro entorno de negociación, argumentamos que los proponentes tienen una posición más fuerte que los respondedores, pues pueden romper la negociación unilateralmente (empoderamiento), tienen más derechos morales sobre el dinero a repartir (legitimidad), y sólo ellos conocen la cantidad total de dinero disponible (información). En ese sentido, es precisamente cuando los hombres ocupan la parte fuerte y las mujeres la débil que los hombres terminan con mayores resultados de negociación.

Para entender estos efectos de interacción de género en los resultados de la negociación, miramos a los efectos de interacción de género en el comportamiento durante la negociación. Analizamos ofertas de los proponentes, demandas de los respondedores, y las probabilidades de aceptar ofertas y demandas. No encontramos diferencias en las ofertas iniciales ni en las ofertas subsecuentes realizadas por hombres y mujeres proponentes. Tampoco encontramos que hombres y mujeres respondedores difieran en las ofertas recibidas. Más interesante, encontramos que son las mujeres las que discriminan entre hombres y mujeres proponentes, lo cual explica por qué el emparejamiento entre un hombre proponente y una mujer respondedora da lugar a las mayores ganancias para los proponentes. No es el caso de que los hombres ofrezcan menos a las mujeres, sino que las mujeres demandan menos a los hombres cuando el hombre ocupa la posición fuerte de la negociación.

Nuestros datos ofrecen múltiples ventajas sobre los experimentos de laboratorio típicos en dónde también se han estudiado los efectos del género en la negociación (por ejemplo, Eckel and Grossman (2001); Solnick (2001); Rigdon (2012); Andersen et al. (2013); Dittrich et al. (2014); Exley et al. (2016)). Primero, el total de dinero sobre el que se negocia es, de media, 345 euros lo que es muy superior a lo que típicamente está en juego en un experimento de laboratorio. Segundo, excepto por el límite de los tres minutos, la negociación no está estructurada y es libre, lo que acerca el marco a una situación real. Nuestros datos también aportan ventajas sobre los experimentos de campo típicos (por ejemplo Ayres (1991); Ayres and Siegelman (1995); Castillo et al. (2013); Leibbrandt and List (2014)). Primero, todo el proceso queda grabado, por lo que podemos estudiar el comportamiento de los sujetos y no limitarnos exclusivamente a analizar los resultados finales. Segundo, tenemos una alta variación en los roles de proponentes y respondedores lo que nos permite mirar no sólo efectos de género, sino también efectos de interacción de género.

Sin embargo, la naturaleza de nuestros datos también presenta ciertos problemas. En primer lugar, los individuos estudiados podrían no ser del todo representativos en cuanto están dispuestos a participar en un programa de televisión. Además, existe la posibilidad de encontrar un *efecto audiencia* por el cuál el comportamiento podría verse afectado. Estas dos limitaciones sin embargo, son comunes a todos aquellos estudios que utilizan datos provenientes de programas de televisión (List (2006); Post et al. (2008); van Dolder et al. (2015)).

A grandes rasgos, nuestros resultados están en línea con la literatura que encuentra que las mujeres demanda menos y se les ofrece menos (a través del efecto de las demandas sobre siguientes ofertas). Además, el trabajo añade tres contribuciones importantes. Primero corrobora que las interacciones de género son cruciales para explica las diferencias de género. Segundo, el trabajo ofrece evidencias basadas en datos de negociaciones no estructuradas recogidos en el campo con pagos importantes en juego, complementando los resultados hallados anteriormente en el laboratorio. Por último, mostramos que el elemento del comportamiento clave en las diferencias de resultados en las negociaciones no radica en la propuestas iniciales de los proponentes, sino en las demandas de los respondedores. En otras palabras, no encontramos signos de un comportamiento discriminatorio en ofertas. Son las mujeres quienes demandas menos a los proponentes hombres.

CAPÍTULO 3: THE ROLE OF GENDER AND ASYMMETRIES IN ALTERNATING-OFFER BARGAINS

En el tercer capítulo, inspirados por los resultados en el capítulo 2, profundizamos en el tema de diferencias de género y efectos de interacción de género en las negociaciones. En particular, dado que el entorno en el que se desarrolla el programa de televisión presentado en el capítulo 2 está caracterizado por fuertes asimetrías entre el rol de proponente y el de respondedor en empoderamiento, legitimidad, e información y que los resultados hallados claramente muestran que los efectos de género están interconectados con los roles asumidos, en este capítulo nos preguntamos si la existencia de distintas fuentes de asimetría tienen impacto sobre la aparición de efectos de género y/o interacciones de género. Nótese que la mayoría de las situaciones reales de negociación están caracterizadas por la existencia de algún tipo de asimetría de las mentadas anteriormente. Por tanto, creemos que el efecto de las asimetrías merecen atención per se.

Para ello, llevamos a cabo un experimento de laboratorio. En él, usamos primero como control un entorno de negociación simétrico en dónde los roles tienen la misma fuerza tal que un reparto del dinero 50:50 es esperable, y en dónde no esperamos encontrar ni efectos de género ni interacciones. A continuación introducimos asimetrías, haciendo un rol (*proponente*) más fuerte que el otro (*respondedor*). Hombres y mujeres son, ex-ante, igualmente probables de ser asignados el rol de proponente o el de respondedor. Es de esperar que la introducción de asimetrías rompa con el reparto igualitario del 50:50, otorgando mayores beneficios al proponente. Estudiamos, de forma separada, tres tipos de asimetrías: empoderamiento (sólo el proponente tiene un dinero garantizado si la negociación fracasa), legitimidad (el proponente tiene un mayor derecho moral hacia la cantidad a repartir), e información (sólo el proponente conoce el total de dinero a repartir). En estas cuatro situaciones estudiamos tanto el resultado final de la negociación (la probabilidad de llegar aun acuerdo, el tiempo que toma llegar a un acuerdo, y la división del dinero) como el comportamiento durante la negociación (ofertas y demandas).

El experimento consiste de tres partes principales. Primero los sujetos deben realizar una tarea de esfuerzo real, a través de la cuál obtienen una productividad. En la segunda parte, los sujetos son emparejados aleatoriamente de dos en dos y tienen tres minutos para llegar a un acuerdo en cómo repartir una cantidad de dinero que está determinada por sus productividades en la parte anterior. Esta parte consiste en 10 períodos en los que se empareja, cada vez, con un participante distinto. En la última parte del experimento, elicitamos creencias individuales con respecto a sus autoevaluaciones en la tarea de la primera parte y en sus habilidades negociadoras, así como sus preferencias sociales y de riesgo.

En el tratamiento simétrico, como esperábamos, el reparto del 50:50 es la norma. Asimismo, no encontramos diferencias de género ni efectos de interacción. Sin embargo, cuando se introducen asimetrías encontramos tanto efectos de género como efectos de interacción de género. Cuando el proponente está empoderado, en el rol de respondedor los hombres obtienen menos que las mujeres (6 puntos porcentuales menos), lo que es explicado por que los hombres respondedores reciben ofertas iniciales menores. Esta es la única excepción, pues en el resto de tratamientos asimétricos, las mujeres obtienen en media peores resultados. Cuando el proponente tiene una mayor legitimidad, los hombres obtienen mejores resultados en ambos roles que las mujeres (4 y 6 puntos porcentuales para proponentes y respondedores respectivamente). Esto se explica por el hecho de que los hombres tienden a ofrecer menos y a demandar más que las mujeres. Cuando el proponente tiene más información que el respondedor, en línea con los resultados del capítulo 2, los proponentes hombres negociando con mujeres respondedoras obtienen mejores resultados que en cualquier otro tipo de emparejamiento (alrededor de 6 puntos porcentuales más), mostrando que las interacciones de género en este caso son importantes. Este hecho, y de nuevo en línea con los hallazgos en el capítulo 2, es explicado por el lado de las demandas, puesto que son las mujeres respondedoras las que demandan menos a los hombres que a otras mujeres. Muy interesante es el resultado de que todos estos efectos de género e interacciones de género están presentes de una forma importante sólo en la primera parte del experimento (primeros 5 períodos) pero desaparecen en su segunda parte (5 últimos períodos), de forma que cuando se consideran todos los períodos conjuntamente los efectos encontrados son mucho más débiles.

También es interesante, dados los 3 minutos de límite que tienen los sujetos para cerrar un acuerdo, que encontramos un "*deadline effect*" muy importante, por el cuál casi el 25% de las negociaciones exitosas se cierran en los últimos 10 segundos. Esto es consistente con los resultados típicamente reportados en este tipo de experimentos (Roth et al. (1988) y Gneezy et al. (2003)). Además de mostrar su existencia, también mostramos que aquellos acuerdos cerrados en los últimos 10 segundos (*ultimatums*) son sistemáticamente diferentes de los demás acuerdos (*acuerdos amigables*).

Este trabajo por tanto ofrece tres contribuciones principales. Primero, es el primer estudio que aborda los distintos tipos de asimetrías y los conecta con la existencia de diferencias de género. Esto es importante pues la mayoría de los estudios realizados incluyen uno o más tipos de asimetrías y, en muchas ocasiones, se confunden entre ellos. Este trabajo pues, aísla cada una de las tres fuentes de asimetría para evaluar, de forma separada, diferencias de género y efectos de interacciones en cada uno de los casos. Los resultados señalan que la existencia y naturaleza de asimetrías en negociaciones no es irrelevante en este aspecto. Segundo, con este experimento podemos estudiar en un entorno controlado, no sólo el resultado final de las negociaciones, sino también el comportamiento durante la negociación. Tercero, los estudios que pueden examinar efectos de interacción además de diferencias de género son escasos. Nuestro estudio utiliza avatares del género del participante con el objetivo de poder separar diferencias de género de los efectos de interacción.

Chapter 1

Stereotypes and Tournament Self-Selection: A Theoretical and Experimental Approach

1.1 Introduction

Social psychologists have shown that stereotypes play an important anchoring role in interpreting private performance feedback and shaping beliefs about own abilities (see Lenney (1977); Beyer (1990); Beyer and Bowden (1997); Pomerantz et al. (2002); Ehrlinger and Dunning (2003)). The Oxford dictionary defines "stereotype" as "*a widely held but fixed and oversimplified image or idea of a particular type of person or thing*". The journalist Walter Lippmann –who first coined the term *"stereotype"* in its modern usage– captures the idea of stereotypes as reference points in his influential book *Public Opinion* (1922) when he states that

> "*The subtlest and most pervasive of all influences are those which create and maintain the repertory of stereotypes. We are told about the world before we see it. We imagine most things before we experience them*".

In an uncertain world, assessing one's own ability is a key step in decisions processes regarding educational and labor market choices, which are important settings for study in Economics. From early ages, children choose between arts and science itineraries based on imperfect signals, such as grades, under a social environment that presupposes that boys are better at math and girls at arts. This choice has been shown to be an important determinant in their later educational and labor market attainments (Paglin and Rufolo (1990)). Later, in the labor market, preconceived gender differences in skills still exist even among individuals who choose the same degree or field (e.g. strategic decision-making; Atwater et al. (2004)). These have been shown to be related to observed gender differences in sorting in the labor market (Cejka and Eagly (1999); Fernandez and Friedrich (2011); Barbulescu and Bidwell (2013); Leslie et al. (2015)). Despite the impact that stereotypes have on decisions of this type in which self-assessment is the key, little work has been done on integrating the concept of stereotypes into the economic modeling.[1] This is the central research question addressed in this paper.

We study how stereotypes affect self-selection into a high paying or a low paying tournament. First, a game theoretical model is presented in which different effects of stereotypes can be identified. Second, we move the setting of the theoretical model into the laboratory in order to test its validity.

The theoretical model considers a mass of agents who have to simultaneously self-select into one of two perfectly discriminatory tournaments, such that winners are named based on their true ability at performing a particular task. These tournaments differ in the prizes that they award: One pays more than (and is therefore preferred to) the other.[2] Agents suffer from *imperfect self-knowledge* as they do not know their own real abilities but observe informative signals about them. Agents face two different problems. First, agents have to cope with imperfect self-knowledge by assessing their (subjective) probability distribution of abilities based on their privately observed signals. Second, they face a strategic situation in which they must forecast others' behavior to maximize the expected return on their participation decision.

Stereotypes are introduced in the model by assuming that the whole population is split into two different social groups. One social group is targeted by a stereotype (positive or negative), which states that, on average, the stereotyped group is better/worse than the non stereotyped one. This *ex-ante* difference takes the form of a displacement of the ability distribution by shifting its mean. Importantly, the stereotype is common knowledge. This assumption, although strong, is given by definition and supported by empirical findings in social psychology.[3]

[1] A notable exception is the work begun by Akerlof and Kranton (2000). Our paper focuses on the analysis of *descriptive* stereotypes as opposed to *normative* ones which are at the core of *identity economics*. This distinction is crucial. *Descriptive* stereotypes simply describe how different a social group *is*, but normative stereotypes dictate how a particular social group *should behave*. Statistical discrimination, as in our paper, also uses descriptive stereotypes. However, in contrast to our paper, stereotypes in statistical discrimination do not affect own ability assessments but assessment of other individuals' abilities. Our approach shows that the existence of stereotypes suffices to generate segregation even in the absent of discrimination of any kind.

[2] The setting presented here is based on Morgan et al. (in press). Important differences must be noted. In the present model, in contrast to Morgan et al. (in press), agents are ignorant of their real abilities but instead observe a signal. Also, we only consider the case in which the total number of vacancies is enough to allocate the entire population. More importantly, guided by our research question, we consider *de facto* two ex-ante different groups.

[3] For example, psychologists have found that both fathers and mothers consistently attribute higher IQs to their sons than to their daughters (Furnham and Gasson (1998); Furnham et al. (2002)). It has also been shown that people attribute higher IQs to their fathers than to their mothers (Petrides et al. (2004)). In other areas, such as mathematics, there is also evidence that stereotypes are widely held (Frome and Eccles (1998); Nosek et al. (2009)). Finally, it should also be noted that stereotypes are part of our cultural inheritance, so it is not that hard to assume that everyone at least has some knowledge of their existence.

The existence of stereotypes has two effects. First, affects the strategic reasoning of *all agents* when incorporating their messages into their beliefs as to how abilities are distributed across the whole population. Second, and more importantly, the existence of stereotypes generates differences in how agents cope with imperfect self-knowledge which will result, through a Bayesian updating process, in the appearance of *self-stereotyping*: a stereotyped and a non stereotyped agent observing the *same* ability signal will differ in the estimation of their abilities and that difference will be in accordance with the stereotype. Consistent with this self-stereotyping mechanism, social psychologists have found important interactions between gender differences in self-assessment and stereotypical perceptions of tasks (see Lenney (1977) for an early review). Chatard et al. (2007) show that when asked to recall past grades in mathematics and arts girls and boys overestimate or underestimate them in a way consistent with existing stereotypes. Interestingly, the size of this effect is moderated by the degree of endorsement of the stereotype. In economics, experimental studies have also shown the importance of self-assessment of own abilities in decision-making. For example, Coffman (2014) shows that when individuals are asked general knowledge questions they display different degrees of confidence in their answers depending on how gender-congruent the topic is, leading to inefficiencies in cooperative settings. Gender differences in self-assessment have also been shown to be important in explaining gender differences when choosing between competitive and alternative payment schemes (Niederle and Vesterlund (2007); Dohmen and Falk (2011); Kamas and Preston (2012)). Related to this, stereotypical perceptions behind tasks and cultural differences have also proved to be important factors (Gneezy et al. (2009); Booth and Nolen (2012); Shurchkov (2012); Dreber et al. (2014); Grosse et al. (2014)). The theoretical model proposed here provides an explanation for the findings in social psychology and economics by explicitly modeling the interaction between existing stereotypes and self-assessment.

After agents have estimated their own ability in light of their signal and the existing stereotype they choose between the high and low paying tournaments, where the probability of succeeding in a given tournament is determined not only by own ability but also by the abilities of other competitors. This leads us to consider choices in the context of rank-order tournaments. As noted by Lazear and Rosen (1981), "*in the real world, where there is population heterogeneity, market participants are sorted into different contests. There players (and horses, for that matter) who are known to be of higher quality ex ante may play in games with higher stakes*" (n. 5).

When the game is solved the differences in self-assessment will generate differences in the signal-based equilibrium strategies followed by stereotyped and non stereotyped agents. In particular, there will be a set of signals for which stereotyped and non stereotyped agents behave differently. If the stereotype is false –such that no behavioral differences are justified– the model predicts that a non optimal group segregation between tournaments will arise, such that the group assumed to be ex-ante better will be overrepresented in the high paying tournament. Furthermore, this participation gap will be transferred to the set of winners. One straightforward implication of these predictions is that in the real world, where many educational and labor choices involve choosing between different competitive environments, social groups will sort differently just because of the existence of stereotypes. Of special interest on this regard is gender-based labor segregation, as several researches point out that men and women sort into the market in very different ways. For example, Fernandez-Mateo and Fernandez (2016) find that supply forces are crucial factors in explaining the low proportion of women in top managerial jobs, providing evidences for "*women self-steering away from the active pursuit of top management jobs*". Related to this, Card et al. (2016) find that women are less likely to work in firms offering higher wages to either gender than comparable men. Similar results linking gender-based segregation with labor outcome can be found in Macpherson and Hirsch (1995); Blau and Kahn (1997); Bertrand and Hallock (2001); Bayard et al. (2003); O'Neill (2003); Ludsteck (2014). Understanding the mechanisms through which stereotypes affect decisions about *where* to compete could shed light on the problem of educational and labor segregation, especially the gender-based segregation which is at the heart of the gender wage gap.[4]

In order to empirically test the implications of the theory, a lab experiment was run in which the theoretical setting was transferred to the lab. Note that the model predicts that there are two mechanisms through which stereotypes affect behavior: self-stereotyping and beliefs about the distribution of abilities over the whole population. However self-stereotyping is the crucial mechanism, so the experiment focuses exclusively on it. In the experiment, we make use of a novel task that proves to be gender neutral in performance but highly heterogeneous regarding the idiosyncratic beliefs about its gender nature. We exploit this heterogeneity in beliefs to see how those affect estimates of abilities and subsequent choices in self-selection into tournaments. Note however that individually (and not commonly) held perceptions are the only important aspect of stereotypes for identifying the self-stereotyping mechanism. In the experiment we therefore refer to perceptions and stereotypes interchangeably as they play the same role.

The bulk of the experiment consists of the following. First, subjects perform a real effort task 14 times consisting of the following: subjects see a picture of a glass with certain quantity of water in it for three seconds. After that time the picture disappears and a picture of an empty glass appears in which they have to indicate the level of water shown in the first picture. This task, which is novel in the literature, proves to be simple to understand, gender neutral in performance, and without no clear nature as to the ability required to perform it well, favoring the existence of heterogeneity as to its gender nature. Second, subjects observe their score in one randomly chosen repetition, which is their only feedback about their performance, and the distribution of

[4]In addition to a pure normative motivation concerned with fairness, recent studies have shown that gender-based segregation has important effects at the aggregate level. In particular, a recent study by Hsieh et al. (2013) estimates that between 16% and 20% of the growth that took place in the US between 1960 and 2008 can be attributed to a change in the labor market sorting of blacks and women.

average scores in that session. Based on this information, they are asked to estimate their average score over the 14 repetitions. This represents the self-assessment stage of the theoretical model. Third, subjects face a decision matrix in which they have to choose between two options (tournaments) –A and B– under different situations. Those situations vary in regard to how many prizes each option offers and to the relative sizes of the prizes. In all situations, option A offers the highest prize and the sum of prizes offered by option A and option B is equal to the number of participants in that session. Fourth, after the main tournament decision part, beliefs about which gender performs better at the real effort task are elicited. Designing the experiment in this way enables us to link the perceived gender nature of the task with differences in self-assessment and in tournament choices.

The experimental results are in line with the predictions of the theoretical model. The perceptions held by subjects about the gender nature of the task directly affect self-assessment: for those who state that the task is gender neutral no gender differences on self-assessment are detected, while for those who state that the task favors one gender a self-assessment gap appears which is consistent with the stated perceptions. When looking at the role played by perceptions as to the gender nature of the real effort task when choosing between tournaments, no gender differences in the likelihood of choosing the high paying tournament (A) are found among those who state that the task is gender neutral, but deviations from gender neutrality in perceptions lead to gender gaps in tournament choices. More importantly, supporting the main self-assessment mechanism proposed by the theory, these differences in tournament choices can be explained by differences in self-assessment. Finally, we show that these differences in self-assessment and tournament choices are passed on to the set of winners.

Two related strands of literature must be mentioned. First, Bordalo et al. (in press) proposes a rationalization of the existence of stereotypes and their persistence over time. By contrast, we take stereotypes as given and inherent to the social environment, and study their effect on individual decision-making. Second, Niederle and Vesterlund (2007) study gender difference in choosing between a tournament and a piece-rate scheme. Our paper, by contrast, studies how stereotypical societal traits affect agents' decisions about *where* to compete. In addition, the setting used by Niederle and Vesterlund (2007) has no strategic component, but our setting does.[5]

The rest of the paper is organized as follows. Section 2 develops the theoretical model: Subsection 2.1 characterizes the equilibrium and subsection 2.2 analyzes the equilibrium prediction in order to identify the effects of stereotypes on segregation, and establishes our main hypotheses for experimental testing. Section 3 is devoted to the experiment: Subsections 3.1 and 3.2 detail the design of the experiment and

[5] Another important difference with respect to the experimental setting used by Niederle and Vesterlund (2007) is that agents have to perform the task after choosing the payment scheme, i.e. there is competition on performance, while in our setting competition consists of submitting the completed real-effort task to the high or low paying tournament.

introduce basic descriptive analyses. Subsection 3.3 analyzes the experimental data in depth. Section 4 concludes.

1.2 A Model of Self-Selection with Stereotypes

Consider a continuum of risk neutral agents of mass 1 (a cohort).[6] Each agent is endowed with ability at a task denoted by a.

The cohort is split into two different social groups $\{S, N\}$. The cohort consists of a mass λ of S-agents and a mass $1-\lambda$ of N-agents. Assume that there is a stereotype about social group S which holds that the distribution of abilities of social group S is equal to that of social group N except for a shift in the mean. If the stereotype is positive (negative), the shift is positive (negative). In particular we assume that, with $M^N > 0$ being the average ability for social group N, according to the stereotype the mean ability for social group S is $M^S = (1+\alpha)M^N$, where the parameter α captures the strength and direction of the stereotype. We restrict our attention to the case in which $\alpha \neq 0$, as having $\alpha = 0$ means *de facto* that there are no stereotypes. Therefore, a negative stereotype is modeled through $\alpha \in [-1, 0)$, while a positive one is consistent with $\alpha \in (0, \infty)$. The stereotype is common knowledge.

Let the distribution for social group N be distributed as $(\widetilde{a}|N) \sim \mathcal{N}(M^N, \sigma_a^2)$ with associated CDF and PDF, $F^N(.)$ and $f^N(.)$, respectively. Thus, according to the stereotype, abilities in social group S are supposedly distributed as $(\widetilde{a}|S) \sim \mathcal{N}(M^S, \sigma_a^2)$ whose CDF and PDF are represented by the functions $F^S(.)$ and $f^S(.)$, respectively. It is immediately evident that the existence of the stereotype translates into a first order stochastic dominance for the distribution of abilities such that if the stereotype is negative then $F^S(a) > F^N(a)$, $\forall a$, and the other way around if the stereotype is positive.

There are two perfectly discriminating and mutually exclusive tournaments, $t \in \{A, B\}$.[7] Tournament t awards a prize $W_t \geq 0$ to, at most, a mass $\delta_t < 1$ of those participants with the greatest abilities. Note that the assumption about tournaments naming their winners based on abilities saves us from any effort-related consideration. Two final assumptions are considered. First $\delta_A + \delta_B \geq 1$, which means that there is no scarcity of prizes, i.e. if coordination occurs everyone could get a prize. Second, without loss of generality, tournament A is more attractive than B, i.e. $W_A > W_B$.

[6] The assumption of risk neutrality plays no role in the model's conclusions but it is made for the sake of clarity. In particular all the results in subsection 1.2.2 remain qualitatively the same under (homogeneous) non-neutral risk preferences.

[7] We could alternatively assume that tournaments are not perfectly discriminating but there is noise in measuring the real ability of agents. Given the characterization of the equilibria (proposition 1.2.1) this would not change the results. This change would just imply a common, higher degree of uncertainty. In particular, it would imply that the self-assessed probability of winning has greater variance which would not bring any additional insight under the risk neutrality assumption.

In this setting, agents have to self-select into one of the two tournaments. However, agents make this choice under *imperfect self-knowledge* about their abilities. In particular, agents are ignorant of their real abilities, but instead they observe a noisy signal about them, r_i, which is given by the following expression:

$$r_i = a_i + \mu_i$$

where a_i is the real ability with which the agent is endowed and μ_i is a realization of a random shock distributed as a zero-mean normal with variance σ_μ^2 which is i.i.d. across agents and abilities. Let $\Phi(.)$ and $\phi(.)$ denote the CDF and PDF of this random shock. When interpreting received signals, agents therefore have two thoughts: How good am I? and how (un)lucky have I been?.[8]

Note that once the agent observes r_i, there is an infinite sequence of pairs of abilities and shocks compatible with that signal, given by the set $\mathcal{L}_{r_i} = \{(a,\mu) : r_i = a + \mu\}$. Given the distributional assumptions, an agent from social group $g \in \{S, N\}$ receiving a signal of his/her ability of r_i can set a probability distribution for his/her abilities.[9] Therefore, agents will compute the conditional distribution of a given r_i given by

$$(a|r_i, F^g) \sim \mathcal{N}\left((1-\gamma)M^g + \gamma r_i, (1-\gamma)\sigma_a^2\right) \tag{1.1}$$

with $\gamma = \sigma_a^2/(\sigma_a^2 + \sigma_\mu^2)$. This expression simply represents a Bayesian updating process which actually depends on the agent's social group through the distributional assumptions (the prior) that are assumed to apply, F^g. The extent to which the updated distribution of beliefs will differ from the prior distribution depends on the quality of the signal, which is captured by the parameter γ: if the signal is very noisy (i.e. is not informative, large σ_μ^2) the updating will have a very low impact and the updated mean and the updated variance will be very close to the one of the prior.[10]

Denote by $\hat{F}(a|r_i, F^g)$ and $\hat{f}(a|r_i, F^g)$ the CDF and PDF of this updated distribution of beliefs over abilities. Notice that $\hat{F}(a|r_i, F^g)$ accounts for the subjective probability that an agent from social group $g \in \{S, N\}$ observing signal r_i attaches to the event of having a real ability lower than or equal to some level a. It is easy to see that if the

[8] For example, consider a student observing his/her grade in Mathematics. One observation –or any finite sequence of observations– cannot account perfectly for real ability in Mathematics, although it surely conveys valuable information about it. It may be the case that the student was lucky that day (and was asked what he/she knew most about) or had an especially bright day. Thus, the student, who is aware that there are factors outside his/her control, cannot say with certainty whether that A$^+$ in Mathematics is representative of his/her *real ability* or is just the product of a positive shock.

[9] The exceptions are given by the degenerate cases in which $\sigma_a^2 \to 0$ or $\sigma_\mu^2 \to 0$, in which agents know their real abilities with probability one.

[10] This estimation process yields that agents observing $r \gtreqless M^g$ will estimate $E(a|r, F^g) \lesseqgtr r$. This is implied by the fact that agents observing high (low) signals undervalue (overvalue) their abilities by attaching the signal to having had good (bad) luck. This is consistent with the *impostor effect* (see Clance and Imes (1978)) and the *Dunning–Kruger effect* (see Kruger and Dunning (1999)).

stereotype is negative ($\alpha < 0$), $\hat{F}(a|r_i, F^S) > \hat{F}(a|r_i, F^N)\ \forall a, r_i$, i.e. an S-agent observing signal r_i will estimate a higher likelihood of having a real ability below a given level than an N-agent observing the same signal. This Bayesian updating process thus ends up with S-agents *self-stereotyping* themselves by transferring the prevailing stereotype from their social group to their self-assessment.[11] Obviously, the opposite occurs if the stereotype is positive.

1.2.1 Characterization of the Equilibrium

The setting described represents a Bayesian game. Formal definitions including the derivation of generic expected utilities are given in Appendix A.1.1. The equilibrium is characterized by a strategy profile that maps types (signal×social group) into actions, such that no agent has an individual incentive to deviate from the equilibrium strategy given the common prior (beliefs) and type.[12]

Proposition 1.2.1. *Given the competitive environment defined by the tuple* $(W_A, W_B, \delta_A, \delta_B, F^N(.), \Phi(.), \lambda, \alpha)$ *the equilibrium of the game is represented by a tuple* $(a^\alpha, r_N^\alpha, r_S^\alpha)$*, such that all agents from social group* $g \in \{S, N\}$ *who observes a signal higher (lower) than a threshold signal* r_g^α *choose to participate in tournament* A *(*B*). The tuple* $(a^\alpha, r_N^\alpha, r_S^\alpha)$ *must hold simultaneously*

1. $\int_{a^\alpha}^{\infty} \lambda[1 - \Phi(r_S^\alpha - y)]f^S(y) + (1-\lambda)[1 - \Phi(r_N^\alpha - y)]f^N(y)dy = \delta_A$
2. $1 - \hat{F}(a^\alpha|r_g^\alpha, F^g) = \frac{W_B}{W_A}, \forall g \in \{S, N\}$

The above proposition formally characterizes the equilibria (a formal derivation can be found in Appendix A.1.2). The first condition represents a "prize clearing condition" for tournament A that can be interpreted as follows: Given the behavior rule summarized by the pair (r_N^α, r_S^α), the mass of prizes of tournament A *should* be cleared, given the beliefs, by those who self-select into tournament A who have real abilities higher than a^α. Importantly, notice that this real ability threshold is estimated to be the same for both social groups, leading them to coincide in the perceived competitiveness of tournament A. Interestingly, the more negative (positive) the stereotype is, the lower (higher) the required believed standard a^α is, i.e. $\partial a^\alpha / \partial \alpha > 0$. In other words, the existence of negative (positive) stereotypes has an encouragement (deterrent) effect that applies to all agents by making them believe that the overall degree of competition is lower (higher).

It is also worth commenting that the first condition implies that tournament A is going to be overcrowded while tournament B shows vacancies, leading to an inefficient outcome. This is because, given imperfect self-knowledge, the probability of an

[11] This seems consistent with evidence of *stereotype threat*. Although this concept has been mostly studied as affecting performance directly, there is some research which shows that it can also affect expectations and ultimately choices (see for example Davies et al. (2002) or Davies et al. (2005)).

[12] It is important to note that the common knowledge assumption of the stereotype is crucial for the equilibrium solution but not for the self-stereotyping in the self-assessment process.

agent from social group $g \in \{S, N\}$ with real ability a getting a signal higher than the required ability threshold r_g^α, and thus participating in tournament A, is

$$Prob(r > r_g^\alpha | a) = Prob(a + \mu > r_g^\alpha) = 1 - \Phi(r_g^\alpha - a)$$

On the one hand, this means that not everyone with abilities higher than the believed real ability threshold, a^α, is going to self-select into tournament A, although the probability increases with the real ability of each agent. On the other hand, this also implies that some low-ability agents self-select into the high-paying tournament, mistakenly believing that they are high-skilled. These two effects –crowding out of high-skilled agents and crowding in of low ones– result in over-entry in tournament A.[13] Implicitly from the over-entry in tournament A, agents expect a mass strictly lower than δ_B entering tournament B with the expectation of getting W_B with probability one.[14]

The second condition of the proposition defines marginal types: Given the *believed* real ability threshold needed to win the prize in tournament A (a^α), there is a signal level for each social group that makes agents indifferent between the two tournaments. Notice that within each social group the expected returns from tournament A are strictly increasing on the signal, while the expected returns from tournament B are constant. The signal level r_g^α thus works as a cutoff point, where agents from social group $g \in \{S, N\}$ observing $r_i < r_g^\alpha$ self-select into tournament B, and agents observing $r_i > r_g^\alpha$ choose tournament A.[15]

Importantly, proposition 1.2.1 implicitly establishes the existence of a mapping from real ability thresholds to signal thresholds and vice versa, enabling us to check that a unique equilibrium always exists (see Appendix A.1.3 for the proof).

Proposition 1.2.2. *There is a real ability threshold signal level, a^α, such that its uniquely induced behavioral rules, $r_S(a^\alpha), r_N(a^\alpha)$, induce simultaneously a^α as a real ability threshold, i.e.*

$$a(r_N(a^\alpha), r_S(a^\alpha)) = a^\alpha$$

Moreover, given the properties of the mappings $a : (r_N, r_S) \to a$ and $r_g : a \to r$ stated in lemmas A.1.1 and A.1.2, this element a^α exists and is uniquely determined given the setting $(W_A, W_B, \delta_A, \delta_B, F^N(.), \Phi(.), \lambda, \alpha)$ in which the game takes place.

Having characterized the equilibrium and shown its existence, we can now analyze the resulting equilibrium strategy profiles for addressing the particular effects of

[13] This result is closely related to the experimental findings in Camerer and Lovallo (1999). The model predicts that agents self-select into the risky activity even when they expect that there will not be prizes for everyone.

[14] This implies that the equilibrium displayed is robust for some environments in which $\delta_A + \delta_B < 1$. In particular the lower bound for δ_B in order to maintain the structure of the equilibria described above is $\int_{-\infty}^{\infty} \lambda\Phi(r_S^\alpha - y) f^S(y) + (1-\lambda)\Phi(r_N^\alpha - y) f^N(y) dy$, which is clearly lower than $1 - \delta_A$. This extreme case corresponds to the case in which all slots in tournament B are filled at equilibrium.

[15] Agents observing exactly this threshold signal are indifferent between the two tournaments but they represent a mass of measure 0.

stereotypes on behavior. From the characterization of the equilibrium in proposition 1.2.1 it should be clear that the cutoff strategy is different for each social group. Given that the real ability threshold for winning the prize in tournament A is common to both types of agents –as both types perceive the same overall distribution of abilities– and given that imperfect self-knowledge generates self-stereotyping, the agent defined as marginal for each social group will necessarily observe different signals. In other words, the minimum signal required for S-agents differs from the one required by N-agents, generating *behavioral differences* between social groups based on signals.

The characterization of the equilibria in proposition 1.2.1 further implies that $\hat{F}(a^\alpha|r_S^\alpha, F^S) = \hat{F}(a^\alpha|r_N^\alpha, F^N)$. Additionally, the signal is perceived to be equally informative for both groups (same γ). These two features mean that

$$\hat{f}(a|r_S^\alpha, F^S) = \hat{f}(a|r_N^\alpha, F^N), \forall a$$

so the threshold signals for the two social groups are belief-equivalent. In words, the marginal S-agent entering tournament A and the marginal N-agent entering tournament A hold exactly the same beliefs about the distribution of their abilities.

Corollary 1.2.3. *For any environment $(W_A, W_B, \delta_A, \delta_B, F^N(.), \Phi(.), \lambda, \alpha)$ the equilibrium signal thresholds are related such that*

$$r_S^\alpha = r_N^\alpha - \alpha M^N \left[\frac{(1-\gamma)}{\gamma}\right]$$

This implies that under the existence of a negative (positive) stereotype, the social group target of the stereotype requires that higher (lower) signals be observed in order to participate in tournament A.

This corollary summarizes the effect of *self-stereotyping* on determining signal-based behavioral differences between social groups. In order for an S-agent to hold the same beliefs about his/her ability as an N-agent observing a signal r, he/she needs to observe a signal $r - \alpha M^N[(1-\gamma)/\gamma]$, which means in essence that S-agents are self-stereotyping. Therefore, since this also applies to marginal agents, there is a set of signals for which the behavior of N-agents and S-agents will differ, even though all agents agree on the minimum ability required to win the prize in tournament A.[16]

Corollary 1.2.3 sheds light on the factors that influence asymmetry into the behavior of the two social groups. First, if the stereotype is very strong (large $|\alpha|$) the

[16] This is very important when looking at empirical data. An external observer who fails to take into account the effect of self-stereotyping could mistakenly conclude that the two social groups differ in their self-confidence. In particular, under the existence of a negative (positive) stereotype, the external observer could conclude that S-agents are underconfident (overconfident). However, this difference in confidence may not be a universal trait but rather task-specific and related to the existence of stereotypes on each particular task. Controlling for self-assessment of ability will determine whether the explanation lies in differences in confidence or differences in self-assessment.

asymmetry intuitively increases. Second, and more interestingly, the size of the self-stereotyping effect is highly dependent on the quality of the signal, γ. If the signal is very noisy about the real ability (not informative, low γ) the asymmetry in behavior between the two social groups increases, as the effect of self-stereotyping becomes very prominent. By contrast, if the signal is very precise about the real ability agents face a setting which approximates *perfect self-knowledge* (γ close to 1), where the effect of self-stereotyping becomes negligible. This makes it clear that the stereotype is just a piece of information that agents take into account only when they do not have other sources of information. This result concurs with previous experimental research on attitude toward competition which indicates that providing feedback – which in our model could be interpreted as increasing γ– is a good policy for dealing with gender differences in competition entry (see Ertac and Szentes (2011), Ewers (2012), Wozniak et al. (2014)). Third, related to the previous point, according to the Bayesian process through which agents update their beliefs about their abilities, the more signals an agent receives, the less he/she is likely to rely on distributional assumptions and the more on the set of signals received. Therefore, the effects of stereotypes are likely to vanish as agents become more experienced in the task itself. Note, however, that acquiring more experience might not be trivial. If an agent persists in an activity in which he/she is believed to be at disadvantage, he/she could realize that he/she is really not, but given that he/she believes from the very beginning that he/she is at a disadvantage, his/her incentives to persist in that activity might be low and he/she may well drop out before collecting enough observations. In short, stereotypes could be a vicious circle. This has clear policy implications, as any intervention providing more and more precise information on agents' abilities would lower the impact of stereotypes. Previous research shows that such interventions can be successful.[17]

1.2.2 Representativity and Comparative Statics

The ultimate aim of this model is to identify potential channels through which the existence of stereotypes may affect the decision of whether to compete in the high or low-paying tournament. Up to now, we have concluded that the existence of a stereotype has two main effects; first it encourages/deters participation in the best tournament for the whole population by reducing/increasing the perception of the abilities of the

[17] In the lab, Brandts et al. (2014) find that the gender gap on entry into a real effort tournament decreases significantly when subjects have a better informed advisor. In particular, this decrease is driven by high-ability women being more likely to enter and by low-performing men being less likely. Interestingly, the effects of advice are found to have no effect on the gender gap among those with intermediate performance levels. Interpreting these results in light of our model, the effect of advice could be seen as a reduction of the magnitude of the $\alpha\,[(1-\gamma)/\gamma]$ term from corollary 1.2.3 generated by an improvement in the signal through the advice received, so that the set of signals for which stereotyped and non stereotyped agents' behavior differs is decreased. Also, Joensen and Nielsen (2015) find that reducing the opportunity cost of taking advanced Math courses at high school increases the enrollment rate on such courses. Note that taking these extra courses can be seen in light of our model as receiving extra signals about mathematical/quantitative skills. This increases the college enrollment rate of women in male-dominated fields. It also increases women's probability of building a professional career in more competitive itineraries and of climbing towards the top of the earnings distribution.

rest of the population (condition one from 1.2.1). This affects both stereotyped and non stereotyped agents equally, so it cannot be seen as a source for explaining behavioral differences between the two social groups. Second, the existence of a stereotype impairs self-assessment between non stereotyped and stereotyped agents in what is referred to as self-stereotyping.

The combination of these two effects enables us to predict that N-agents will enter tournament A under a different set of signals from S-agents. We now address the question of how this sorting affects the *representativity* of each social group in each tournament. All proofs of results in this section are provided in Appendix A.1.4. We start by defining what we mean by representativity.

Definition *We say that a social group is underrepresented (overrepresented) in tournament A if the proportion of agents from that social group in tournament A is lower (higher) than would be expected under perfect information.*

In order to manage the concept of overrepresentation and underrepresentation presented in the above definition we focus solely on the S group. Furthermore, from now on we restrict the analysis to the case in which the stereotype is false.[18] If the stereotype is false, the distribution of abilities for both social groups is the same and equal to $\widetilde{a} \sim F^N(.)$. To avoid confusion, denote the CDF for the common distribution when the stereotype is false as $F(a)$ with associated PDF $f(a)$. To account for effects of over/underrepresentation we construct the following index:

$$\frac{\lambda_A^\alpha}{\lambda} = \frac{\int_{-\infty}^{\infty}[1-\Phi(r_S^\alpha - y)]f(y)dy}{\int_{-\infty}^{\infty}\lambda[1-\Phi(r_S^\alpha - y)]f(y) + (1-\lambda)[1-\Phi(r_N^\alpha - y)]f(y)dy} \tag{1.2}$$

where λ_A^α is the proportion of agents in tournament A who come from social group S as predicted by the equilibrium of the model and λ is the proportion of agents from social group S in society. Notice that under the assumptions that the stereotype is false and that agents have perfect information about their real abilities the proportion of agents in tournament A from social group S should be exactly λ.[19] Therefore, social group S will be underrepresented when $\lambda_A^\alpha/\lambda < 1$ and overrepresented when $\lambda_A^\alpha/\lambda > 1$. Thus when the stereotype is false any deviation from $\lambda_A^\alpha/\lambda = 1$ is undesirable as it indicates the existence of representativity problems. More interestingly, the reading of the expression $|1 - \lambda_A^\alpha/\lambda|$ provides us with a measure of the *representativity gap*.

[18] The case in which the stereotype is true is hard to manage, as whether it causes over/underrepresentation depends not only on whether the stereotype is positive or negative but also on the joint values of the magnitude of the stereotype, the spread of prizes and the actual variance of abilities. For this reason no clear conclusion can be drawn, so we stick to the case in which the stereotype is false.

[19] Under perfect information (when real abilities are perfectly known to the agents) the decision rule for both social groups would be the same (to choose A if ability exceeds a certain threshold) independently of whether a stereotype exists or not. Therefore the social composition in each tournament would replicate that of the whole population. The role of stereotypes in this setting would be just to decrease/increase the believed required minimum ability for obtaining W_A.

Proposition 1.2.4 (**Segregation**). *If the stereotype is false, agents from social group S are underrepresented (overrepresented) when the stereotype is negative (positive).*

This proposition implies that when the stereotype is false there is always a representativity gap in the pool of participants in tournament A independently of the particular environment in which the choice takes place. However, given that there is an excess of entries in tournament A, not all agents will obtain a prize, so this imbalance in participation may not translate into an imbalance in the set of winners. This is of great concern, especially taking into account corollary 1.2.3, which implies that if the stereotype is false then S-agents participating in tournament A will have a higher average rate of success when the stereotype is negative as on average they will have higher ability levels than those from social group N, and vice versa when the stereotype is positive. However a look at how the mass δ_A of prizes is split between the two social groups reveals that these differences in winning rates do not cancel out the imbalance in participation.

Proposition 1.2.5 (**Wage gap**). *If the stereotype is false, agents from social group S are underrepresented (overrepresented) among the set of winners from tournament A when the stereotype is negative (positive).*

This is because lower signal thresholds increase the chances of high-ability agents who receive low signals correctly choosing tournament A. Thus, when the stereotype is negative N-agents have a lower signal threshold than S-agents –sustained by the belief that the stereotype is true–, so high-ability S-agents are less likely to enter tournament A than equally skilled N-agents. If the stereotype is positive the opposite reasoning applies. Notice that this implies the existence of a wage gap at social group level. Among those who win any prize (get a job, for example), S-agents are more likely to get the position in which less is paid than equally skilled N-agents and vice versa if the stereotype is positive. It is important to remark that this result is found in the absence of any kind of discrimination arising naturally from the existence of a stereotype in which all agents believe.

These two results show that representativity issues in the high-paying tournament are present both in participation in the high-paying tournament and among the winners. Moreover, the model enable us to conduct a comparative statics exercise to further understand the effects by addressing the issue of under which competitive environments the effects are more severe.

Proposition 1.2.6. *If the stereotype is false, the representativity problem measured through $|1 - \lambda_A^\alpha/\lambda|$ is*

1. *higher the higher the ratio W_B/W_A is*
2. *higher the lower δ_A is*
3. *higher the lower λ is*
4. *higher the higher γ is*

Several interesting issues are worth mentioning. First, in regard to the spread of prizes (W_B/W_A) comparative statics show that representativity problems increase when tournaments are balanced in terms of the amount of prizes. In other words, competitive structures in which one tournament is clearly superior in terms of potential outcomes tend to make differences in behavior disappear. Second, regarding the mass of prizes offered by tournament A (δ_A), participation gaps become greater as the number of prizes offered by the high-paying tournament become smaller, i.e. the more elitist tournament A is. These two results are related in the sense that both lower W_B/W_A and higher δ_A encourage agents observing low signals to participate in tournament A, thus lowering the representativity gap. This is implied because, in general, the higher the signal threshold that needs to be observed to participate in tournament A is (the higher r_S^α and r_N^α are), the bigger the representativity gap becomes.[20] Third, the result regarding the proportion of the whole population accounted for by the stereotyped social group (λ) is very interesting as it posits that stereotype-related participation gaps are negatively related to that proportion. This point is noteworthy because it implies that representativity issues will be greater when stereotyped social groups account for a minority. Finally, the representativity gap becomes greater as the signal becomes less informative (lower γ). This last result reinforces the idea mentioned above that stereotypes have behavioral consequences because there is imperfect self-knowledge.

1.3 Experimental Test

A laboratory experiment was run to test the relevance for the theoretical model on explaining self-selection into tournaments under the existence of stereotypes. The experiment explicitly tests for the self-stereotyping mechanism described in the theoretical model as it represents the main source for differences in self-selection.[21]

The experiment took place at the Bilbao Laboratory of Experimental Analysis (Bilbao Labean) at the University of the Basque Country on a computer based form using z-Tree experimental software (Fischbacher (2007)). Subjects were recruited through ORSEE (Greiner (2015)), which resulted in 140 participants –69 (49.29%) men and 71 (50.81%) women– split into four different sessions that took place on two consecutive days. Recruiting was carried out such that the gender balance in each session could be assured but in such a way that subjects were unaware of this at the time of recruiting.[22]

At the beginning of each session, subjects were provided with written general in-

[20] See equation (A.7) in Appendix A.1.4.

[21] As will be made clear in the presentation of the design, the encouragement/deterrent effect which affects both social groups is canceled out. This is further reflected in the data that we collected.

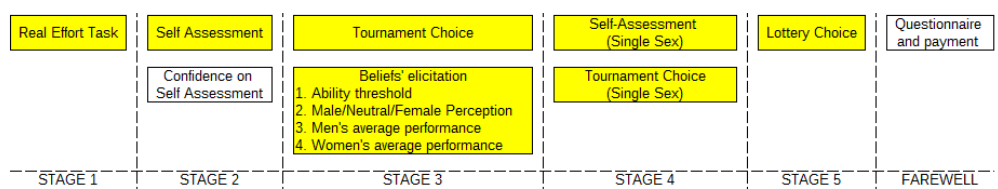
[22] The number of participants in each session and the percentage of women were 36 (47.2%), 37 (54%), 33 (57.6%) and 34 (44.1%) respectively. All sessions were statistically gender balanced as intended. Since subjects could see each other at the beginning and during the experiment, it is assumed that they were aware of this gender balance.

structions which were read aloud to guarantee that they were public knowledge. In particular, these general instructions informed them that the experiment consisted of 5 different stages whose instructions would be displayed on their computer screen immediately before the start of each stage. A translation of all the instructions can be found in Appendix A.2. All subjects took all decisions simultaneously, so all agents entered all stages at the same time. Each session lasted for around one and a half hours, including payment. Average earnings were 16.03 euro (s.d. 3.51) including a show-up fee of 3 euro, and total earnings ranged from 7.3 euro to 24.3 euro.

1.3.1 Experimental Design

The experiment was designed to follow closely the theoretical framework presented in Section 2. In particular, in stages 1, 2 and 3 the setting from the theoretical model is transferred to the lab.[23] At the end of stage 3, subjects' beliefs about the gender nature of the task and beliefs about other participants' behavior are collected. Stage 4 repeats stage 2 and 3 but in a single sex setting. Stage 5 consists of a lottery choice intended to elicit risk preferences (Eckel and Grossman (2002)). The experiment ended with a non-incentivized questionnaire in which subjects were asked about demographics, their taste for competition, social risk preferences (Weber et al. (2002)), and the big five personality traits (Gosling et al. (2003)). For a graphical summary of the experiment see Figure 1.1, which is explained in detail below.

Figure 1.1– Roadmap of the experimental design



Notes: All incentivized tasks are shown in yellow.

The first stage of the experiment consisted of a real effort task. In this part, subjects saw a glass with some water in it for 3 seconds (Figure 1.2, picture on the left). The picture then disappeared and an empty glass appeared with a blue slider at the bottom (Figure 1.2, picture on the right). Subjects were asked to drag the slider and locate it as close as possible to the level of water shown in the previous picture. In order to prevent cheating, the first picture was displayed in the top half of their screens while the picture of the empty glass was in the bottom half. Subjects had to repeat this task fourteen times for different initial pictures. The only thing that changed between one picture to another was the level of water shown. All subjects saw the same pictures in the same order. Before starting the task, subjects had a practice trial to get used to

[23]To avoid framing effects, instead of referring to choices as "tournaments" the instructions during the experiment referred to them simply as "options".

how the task worked and to settle any potential doubts. No feedback was given to the subjects on how well they did on this trial and it was made clear that this had no effect on the rest of the experiment.

FIGURE 1.2– PICTURES SHOWN TO SUBJECTS DURING THE REAL EFFORT TASK.

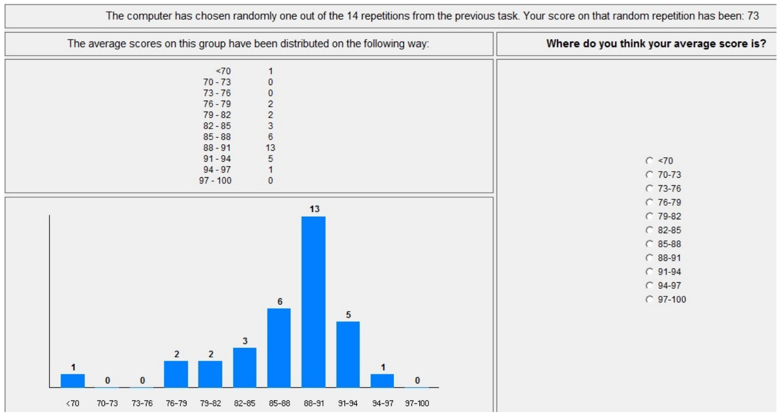
The score awarded to each subject in each repetition was inversely proportional to the distance between the actual level of water in the first picture and their guess. Subjects were told that for payment only one of the fourteen repetitions would be randomly selected.[24]

This particular task, which is novel in the literature, was selected for several reasons. First, participants were very unlikely to have any prior experience with this or any related task. This means that they had no expectations about their own performance or that of others prior to participating in the experiment. Second, the nature of the actual ability required to perform well is not clear. This helps to ensure heterogeneity in the beliefs of subjects as to what ability is important to perform well and therefore as to the gender nature of the task. To further check this point, a pre-experimental survey was run on a similar subject pool. This confirmed that without performance the task is perceived on average as gender neutral but that there is a significant degree of heterogeneity in beliefs as to its gender nature.[25] Third, it is very complicated to figure out how well one has performed on the task without feedback. Fourth, the task is extremely simple to understand and no gender differences in performance were expected

[24]Specifically, the score in each repetition was computed as $100 - 5 \times Distance|Real\_Water\_Level, Slider\_Location|$ such that the maximum score attainable was 100. Notice that the minimum score varies from picture to picture. Negative scores were extremely unusual and only represent 8 out of the 1960 times that the task was performed in total (0.41%). In terms of signals only 1 out of the 140 subjects received a negative one. The rule for payment was the following: $Payment_{stage\_1} = max(5 \times score\_selected/100, 0)$.

[25]The survey was conducted at the Public University of Navarre to gather information about beliefs held by undergraduate students about the task. Once the task was described in the same way as during the experiment, participants were asked which gender they thought would perform better at it. Immediately afterwards they were asked to explain their choice in an open form question. Most subjects reasoned their choice by linking the task to a particular ability. These abilities were extremely heterogeneous in nature, which indicates that unanimity in this regard is far from being the case. The required abilities reported included intuition, precision, logic, visual memory, spatial vision, and others.

FIGURE 1.3– SELF-ASSESSMENT TASK



which was indeed corroborated in the data (t-statistic=-0.069, two-sided p-value=0.74).

In stage 2, participants were asked to estimate their average score over the fourteen repetitions at stage 1 task. Instructions emphasized that this average score represented their *ability* on that particular task. Participants were asked to select where in a ranking of 11 possible levels they thought their ability lay. These ranks ranged from "<70" to "97-100" in intervals of 3. Subjects were informed that if their estimation was correct they would be paid an additional 1.5 euro. In this stage, we used two different instruction sets.

In the first two sessions before entering the estimation subjects were shown their score in one randomly chosen repetition out of the fourteen times that they performed the real effort task in stage 1, i.e. they were presented with an informative but not perfectly correlated *signal* of their ability levels.[26] They were also presented with the distribution of the abilities of all participants in the session. This information was presented as a chart and as a histogram (see Figure 1.3).[27]

In the last two experimental sessions, before subjects were shown their signal a brief description of the task was provided in which the task was framed as being related to spatial vision and manipulation of visual information, and it was stated that such skills were essential for engineering. Note that the worst case scenario for this experiment is the case in which the general perception is gender neutrality or the absence of gender-based perceptions as the model would predict no behavioral differences. The

[26]The actual correlation between signals and abilities for the full sample was 0.43 (p-value<0.001).

[27]This feature of the experimental design makes it impossible to test the mechanism addressed in the model by which stereotypes have an encouraging/deterrent effect. When the real distribution of abilities is provided there is no way in which stereotypes can affect perceptions about the skillfulness of the *whole* population. We decided to provide this information in order to gain power in the estimation stage, as failing to provide the distribution of abilities during the session would probably result in a high degree of randomness in estimations which would greatly reduce the power of the econometric analysis.

purpose of this additional information was to increase the proportion of subjects who reported that the task was male. This manipulation did indeed increase the perception of the task as being male by 20% (p-value<0.05) but it had no other noticeable effect, so we pooled the data in all sessions.[28] The rest of the instructions were exactly the same as in the two previous sessions.

After entering their estimations, subjects were asked to enter their level of confidence concerning that estimation through a non-incentivized question. Confidence was measured through a 7 level Likert scale with 1 standing for "*not sure at all*" and 7 for "*completely sure*".

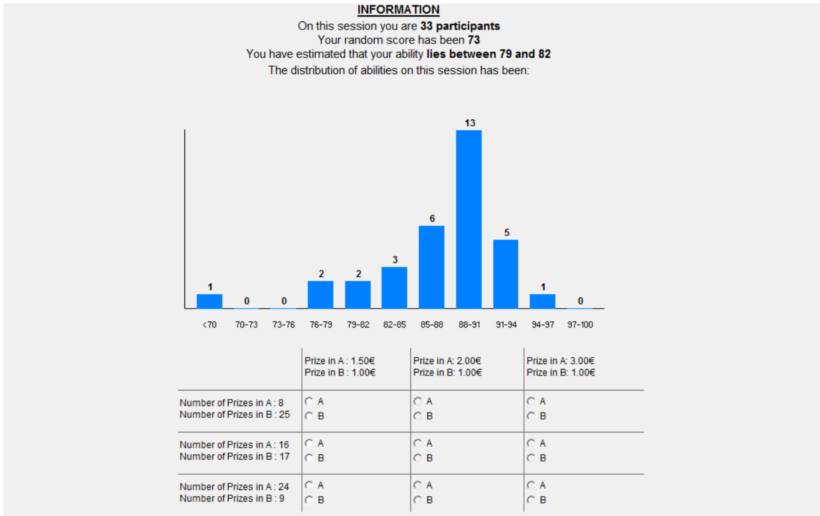
Stage 3 involved three different parts. In the first part subjects were presented with a decision matrix that showed 9 different situations. In each situation they were asked to choose between option A and option B, each of which offered different prizes. Subjects were informed that if they chose an option in which the number of participants choosing that same option was lower than the number of prizes available under that option, they would get the option's prize for sure. When the number of participants opting for that option exceeded the number of prizes available for that option in that situation, only those with the highest abilities in the stage 1 task would get prizes. Thus, these "options" resemble the tournament choice presented in the theoretical framework shown in Section 2.

The 9 situations differed in the number of prizes offered by option A and option B and in the value of the prize in option A. Importantly for this part of stage 3, in all situations the sum of the number of slots in option A and option B was always equal to the number of participants in that session of the experiment. This means that the only way in which everyone could get a prize was by perfectly coordinating on choices. Option B always offered prizes worth one euro while option A's prizes could be 1.5 euro, 2 euro or 3 euro. Similarly, the number of prizes offered in option A (B) were 1/4 (3/4), 1/2 (1/2) or 3/4 (1/4) the number of subjects participating in each session. The combination of these 3 spreads of prizes and the 3 possible allocations of prizes gives the 9 situations observed by subjects (see Figure 1.4).[29] Subjects were told that at the end of the experiment one of the 9 situations would be taken randomly and the amount resulting from their choice and that of the other participants in that situation would be paid. While choosing between option A and option B subjects could see at all times the number of participants in the session, their signals obtained from their performance in stage 1, their estimated ranking of their ability levels from stage 2, and

[28] Results go in the same direction if all the analyses are conducted on either the manipulated or the non manipulated group. See tables A.1 and A.3 in Appendix A.3. Moreover, all the analyses performed in this paper include session fixed effects and thus control for this manipulation. Excluding session fixed effects and including instead a dummy variable with a value of 1 if the session was manipulated and 0 otherwise and the interaction effect of this variable with the *Female* dummy show that the manipulation itself has no noteworthy effect on the variables studied apart from modifying perceptions.

[29] After the reading aloud of the instructions for this stage, subjects had to answer a test with control questions to guarantee that they had understood correctly how winners were chosen.

FIGURE 1.4– TOURNAMENT SELF-SELECTION



the distribution of abilities in the session.

For the sake of completeness, it is worth saying that the second part of stage 3 was identical to the first except that this time in all 9 situations option B offered the same number of prizes as there were subjects in the session. Decisions in this setting however are not really about *where to compete* but about *whether to compete or not*. Given that we use gender as social groups this distinction seems to be important.[30] Consequently, results from this setting are analyzed in a companion paper.

Stage 3 concluded with 5 questions –which were displayed one after another on different screens– related to the session. At the end of the experiment the computer selected one of them and subjects received 1.5 euro if the answer provided was correct. A translation of the exact wording can be found in Appendix A.2. Questions 1 and 2 were intended to measure potential heterogeneity in forecasting others' behaviors by asking for subjects' forecasts about the minimum ability required in order to obtain a prize in tournament A at the first and the second part of stage 3 respectively. In terms of the theoretical model, we were asking subjects to forecast the *real ability threshold*. Subjects' answers to these questions showed us that the vast majority of them correctly expected that tournament A was going to be overcrowded in both parts of stage 3.[31]

[30] In particular, under this setting we find a large, significant gender gap in entry into tournament A that is absent in the setting analyzed in this paper. However, for this setting perceptions do not appear to play a significant role in explaining behavior, although coefficients always go in the direction predicted by the theory and even become significant at the 5% level when looking at the subsample given by marginal agents (see subsection 1.3.3) and when extra controls are included. Results are available upon request. The consistency of coefficients with the theory suggests that the aforementioned gender difference in behavior may be hindering our analyses when we look at perception-related behavior.

[31] In the instructions for this question, subjects were informed that if they believed tournament A was going to be uncrowded –so that there was no minimum ability–, they should choose ability "<70". Only

Questions 3, 4 and 5 were intended to measure subjects' perception of the gender nature of the task performed in stage 1. Question 3 was introduced to directly elicit beliefs about stereotypes by asking subjects to guess who had performed better on average in the real effort task: men/no differences/women. Questions 4 and 5 were asked to gauge subjects' perceptions as to the average performance of men and women respectively by using an eleven-option item ranging from "<70" to "97-100" in intervals of 3.

Stage 4 resembles stages 2 and 3 except that *subjects only interacted with those of their same sex*. We refer to this stage as the *single-sex* setting, and to the previous one as *mixed-sex*. In this *single-sex* setting, subjects again had to estimate their ability levels. The only difference with respect to stage 2 was that this time we showed them the distribution of abilities within their own gender instead of the distribution of the whole session. They were told that if their estimation was correct, they would be paid an additional 1 euro. Next they again had to choose between tournament A and B. The rules were the same as in stage 3, but this time subjects only interacted with their own gender. This meant that at all levels the session was split into two gender-homogeneous groups and subjects played stages 2 and 3 within their corresponding group.

Stage 5 consisted of a lottery choice to elicit risk attitudes. The elicitation method used resembles the one in Eckel and Grossman (2002), enabling 8 different degrees of risk attitude to be differentiated.[32]

Finally, subjects also completed a non-incentivized questionnaire which collected data on some sociodemographic variables and self-reported measures of taste for competition, social risk attitude (Weber et al. (2002)), and the big five personality traits (Gosling et al. (2003)). Given the sample size and its low impact on the main results, social risk, personality traits scores and some socio-demographic variables were not included on the main analyses of the paper. A robustness check of the main results including these set of control variables can be found in Appendix A.3.

Hypotheses and Overview of the Results

The experimental design described above enable us to directly test the following claims established by the theoretical model presented in section 1.2.

C_1: *Perceptions held about the gender nature of the real effort task bias subjects' ability estimations in a consistent way.*

This claim captures the *self-stereotyping* mechanism. Results with respect to this claim (section 1.3.3, Table 1.4) support the theoretical predictions.

6 subjects (4.2%) made this choice.

[32] Lotteries were chosen in order to distinguish risk neutrality, 4 different levels of risk aversion, and 3 of risk loving according to a constant relative risk aversion utility function of the form $\mu(x) = x^{1-r}$.

C_2: *Perceptions held about the gender nature of the real effort task affect self-selection into tournament A.*

C_3: *Perceptions held about the gender nature of the real effort task affect the ex-ante probability of obtaining the prize in tournament A.*

C_2 and C_3 describe the consequences that perceptions have on behavior and outcomes according to the theoretical model. In particular, C_2 refers to corollary 1.2.3 in the theory and this is supported by experimental evidence (section 1.3.3, Table 1.5). C_3 contains the theoretical prediction regarding the impact of perceptions on subjects' outcomes which is also supported by the experimental data (Table 1.6).

C_4: *The effects of perceptions in C_2 and C_3 can be explained by C_1.*

According to the theoretical model, two agents with the same ability estimation should behave identically. Thus, claim C_4 states that controlling for differences in self-assessment should cancel out the direct effect of perceptions on both self-selection (C_2) and winning rates (C_3). The data shows that this is indeed the case as controlling for self-assessment leads to perceptions no longer affecting behavior or winning rates (Tables 1.5 and 1.6).

As a robustness check to enable us to underpin the previous results, we also add the following claim:[33]

C_5: *In the single-sex environment perceptions held about the gender nature of the real effort task play no role in self-assessment or self-selection into tournament A.*

This last issue is addressed in section 1.3.3, Table 1.8. This claim is supported by our data, reinforcing the premise that results in C_1, C_2, and C_3 come from the effects of subjects' perceptions and not from unobserved factors and/or self-selection issues.

1.3.2 PERCEPTIONS AND DESCRIPTIVE STATISTICS

We gather information about subjects' perception as to the gender nature of the task using the answers provided at the end of stage 3 (questions 3, 4, and 5). In particular, question 3 provides us with a discrete variable on whether subjects perceived the task in stage 1 as male, female or gender neutral. We refer to this variable as *Perception* of the task. Meanwhile, questions 4 and 5 ask about the perceived average performance for men and women, respectively.

Based on subjects' answers to these two questions we construct two variables: *Maleness* and *Relative Maleness*. The former is just the difference between the reported averages in abilities for men and for women. *Relative Maleness* is computed as the difference between the reported *relative* averages in abilities for men and for women, so

[33] Although it is not explicitly addressed in the theoretical model, this prediction follows immediately because by design only the self-stereotyping mechanism is in play in the experiment. So during stage 4 of the experiment, in which we provide subjects with the distribution of abilities within their own gender, we should not observe any effect related to perceptions as there is no way to suffer from *self-stereotyping*.

it captures not only the absolute but also the relative difference between the believed average of men's and women's performance by taking into account distributional information.[34]

Notice that both *Maleness* and *Relative Maleness* provide continuum measures of the perceived maleness of the task, while the variable *Perception* offers a discrete measure. However all three should be consistent one with another such that if a subject claims that men have a higher ability level than women (*Maleness*, *Relative Maleness*>0) his/her associated *Perception* of the task should be male. In the subsequent analysis we therefore consider only those subjects who are consistent in these measures. This leaves us with 120 out of the total sample of 140 subjects (85% of the original sample). See Table 1.1.

Although all three variables (*Perceptions*, *Maleness*, and *Relative Maleness*) return similar results, in the rest of the paper we show only the analyses for *Relative Maleness* as it proves to be much more powerful in terms of both answering our research question and statistical power. Supplementary analyses making use of the variables *Perceptions* and *Maleness* are available upon request.

TABLE 1.1– CONSISTENCY BETWEEN RELATIVE MALENESS AND PERCEPTION

		Perception			
		Male	Neutral	Female	Total
Relative Maleness	<-60%	0	0	3	3
	[-60%,-40%)	0	0	4	4
	[-40%,-20%)	0	5	18	23
	[-20%,0%)	2	7	6	15
	0%	10	31	6	47
	(0%,20%]	10	1	0	11
	(20%,40%]	25	4	0	29
	(40%,60%]	6	1	0	7
	>60%	1	0	0	1
	Total	54	49	37	**140**

		Perception			
		Male	Neutral	Female	Total
Relative Maleness	<-60%	0	0	3	3
	[-60%,-40%)	0	0	4	4
	[-40%,-20%)	0	0	18	18
	[-20%,0%)	0	0	6	6
	0%	10	31	6	47
	(0%,20%]	10	0	0	10
	(20%,40%]	25	0	0	25
	(40%,60%]	6	0	0	6
	>60%	1	0	0	1
	Total	52	31	37	**120**

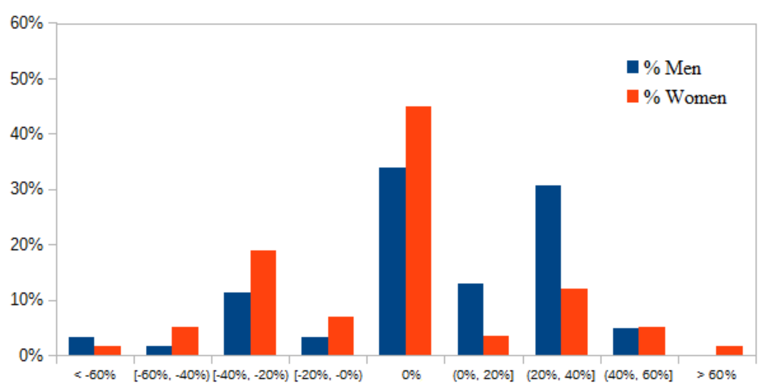
Note: Relationship between *Relative Maleness* (rows) and *Perception* (columns) of the task for the original sample (left) and for the final sample (right). Inconsistencies are displayed in red.

[34] In particular, *Relative Maleness* is computed as follows: suppose that in a session with N subjects one of them claims that men's average ability lies in the interval X, for which the observed frequency is N_X. Then we count the number of subjects whose ability is strictly below that interval, and denote this by N_X^-. Thus we can argue that this agent believes that men, on average, are better than $[N_X^- + (N_X/2)]/N$ percent of the subjects participating in that session. By computing this percentage for the relative position of women and computing its difference relative to men, we get the *Relative Maleness* variable. For example, using the distribution of abilities displayed in Figure 1.4, a subject claiming that men's average is in the interval 91-94 and women's in the interval 85-88 would be translated into relative positions of 89% ([27+(5/2)]/33) for men and 33% ([8+(6/2)]/33) for women, thus returning a *Relative Maleness* of 56%. If instead these claims were 85-88 for men and 79-82 for women, then the *Relative Maleness* would be 21%. Notice that in both cases though, the variable *Maleness* would be the same at 2. These examples illustrate that accounting for the underlying distribution conveys important information that should be taken into account in the analyses.

Thus, this *Relative Maleness* is taken as our variable for assessing the impact of the gender-based perception of the task, so it deserves a closer look. In particular, although this variable is constructed based on incentivized elicitation, it is of crucial importance to rule out potential endogeneity and self-selection issues. Given that beliefs about the gender nature of the task are not randomly allocated, there may be important idiosyncratic factors that affect subjects' perceptions. However, our main concern for the purpose of econometric analysis is to address whether there is any systematic bias in the way in which these beliefs are set. We now assess the potential existence of biases in this *Relative Maleness* variable with respect to other key control variables.

We first compare the distribution of beliefs across genders as regards *Relative Maleness* (see Figure 1.5). We find an own-gender bias in these beliefs such that both men and women are more likely, on average, to state that their gender performs better (although it is only significant for men, p-value<0.1). Moreover, this own-gender bias translates into a small gender difference in beliefs (p-value<0.1). This seems to be a pure gender effect and should not be problematic when it comes to analyzing the data.

FIGURE 1.5– DISTRIBUTION OF THE RELATIVE MALENESS VARIABLE WITHIN GENDERS



Second and more importantly, we should check whether subjects used perceptions to self-justify their performances. It is plausible that men (women) observing high (low) signals could justify their performance by reporting the task as male and vice versa if the signals observed are low (high). That is, the reported *Relative Maleness* of the task might be correlated with the observed signal.[35] In that case, our analysis would suffer from a strong identification problem as we would not be able to properly identify the effect of perceptions about the gender nature of the task on the outcome variables of interest. The regression analysis shown in Table 1.2 checks for and rules out this possibility.

In columns (1) and (2) of Table 1.2, analyses are run separately for men and women.

[35] According to cognitive dissonance, the signal observed could affect the belief formation process such that, within genders, beliefs and signals could be correlated.

Regressions show no significant effects in either gender, as can be checked through the statistical insignificance of the coefficient for the variable *Signal*. Analysis on column (3) checks for whether the size of the effect of the signal in determining *Relative Maleness* differs for men and women. The statistically insignificant coefficients for the interaction effect *Signal∗Female* further confirms that the size of the impact of signals does not differ for men and women. These analyses enable us to reject the hypothesis that a self-selection problem driven by observed signals could be affecting the reported *Relative Maleness* of the real effort task.

TABLE 1.2– OLS REGRESSION FOR RELATIVE MALENESS

Dep. Variable: *Sample:*	*Rel. Maleness* *Men* (1)	*Rel. Maleness* *Women* (2)	*Rel. Maleness* *All* (3)
Female			-0.0313
			(0.263)
Signal	0.0223	-0.0379	0.0223
	(0.199)	(0.215)	(0.200)
Signal∗Female			-0.0602
			(0.294)
Constant	0.0398	0.00849	0.0398
	(0.183)	(0.190)	(0.183)
Observations	62	58	120
R-squared	0.000	0.000	0.024

Notes: Female takes value 1 when the subject is a woman and 0 otherwise. *Signal* is a continuous measure accounting for the performance signal observed by the subject during the experiment. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Panel A of Table 1.3 looks at potential self-selection patterns based on other potentially relevant control variables by showing the correlations within each gender with the *Relative Maleness* variable. As was expected, within each gender the only variables significantly correlated with *Relative Maleness* are those related to beliefs themselves: believed relative average ability for men (*Bel.Ab.Men*) and the equivalent for women (*Bel.Ab.Women*). Thus, Panel A rules out the existence of self-selection issues in observables, validating the use of *Relative Maleness* as our treatment variable.

Panel B of Table 1.3 presents an overview of the results by looking at the raw data of the main variables os interest: *Relative Assessment*, *Prob(A)* and, *Prob(W_A)*.[36]

[36]Notice that we talk about the *Relative Assessment* instead of referring to the assessment measure collected directly during the experiment. This is because the original assessment variable collected during the experiment conveys only cardinal information, but this *Relative Assessment* variable captures the relative assessment within the pool of subjects, which is actually the relevant assessment when choosing between tournaments. This *Relative Assessment* variable is computed as follows: suppose that in a session with N subjects one of them claims that his/her ability lies in the interval X for which the observed frequency is N_X. Then we count the number of subjects whose ability is strictly below that interval and denote this by N_X^-. Thus we can argue that this agent believes that he/she is better than $[N_X^- + (N_X/2)]/N$ percent of the subjects participating in that session. For example, using the distribution of abilities dis-

TABLE 1.3– CORRELATIONS FOR MAIN CONTROL AND OUTPUT VARIABLES WITH RELATIVE MALENESS

PANEL A: CONTROL VARIABLES

Obs.	*Men* 62 (1)	*Women* 58 (2)	Obs.	*Men* 62 (1)	*Women* 58 (2)
Signal	0.0151	-0.0187	Risk Pref.	0.1726	-0.1472
	(0.907)	(0.889)		(0.180)	(0.270)
Ability	0.1333	-0.0181	Age	0.0872	0.1547
	(0.301)	(0.893)		(0.500)	(0.246)
Conf.Est.	0.0871	-0.1592	Difficulty	-0.0007	-0.1546
	(0.501)	(0.233)		(0.996)	(0.246)
Min.Ab.Win	0.1498	-0.0046	Taste Comp.	0.1413	-0.1941
	(0.245)	(0.973)		(0.273)	(0.144)
Bel.Ab.Men	0.7736***	0.7192***	Bel.Ab.Women	-0.7230***	-0.6284***
	(0.000)	(0.000)		(0.000)	(0.000)

PANEL B: OUTCOME VARIABLES

	Mixed-Sex			**Single-Sex**	
Obs.	*Men* 62 (1)	*Women* 58 (2)	Obs.	*Men* 62 (1)	*Women* 58 (2)
Rel. Assessment	0.3322***	-0.1961	Rel. Assessment$_{ss}$	0.1859	0.0345
	(0.008)	(0.140)		(0.148)	(0.797)
Prob(A)	0.1599***	-0.0646	Prob(A)$_{ss}$	0.0647	-0.0473
	(0.000)	(0.141)		(0.127)	(0.281)
Prob(W_A)	0.3088**	-0.0622	Prob(W_A)$_{ss}$	-0.0490	-0.0388
	(0.015)	(0.643)		(0.7053)	(0.772)

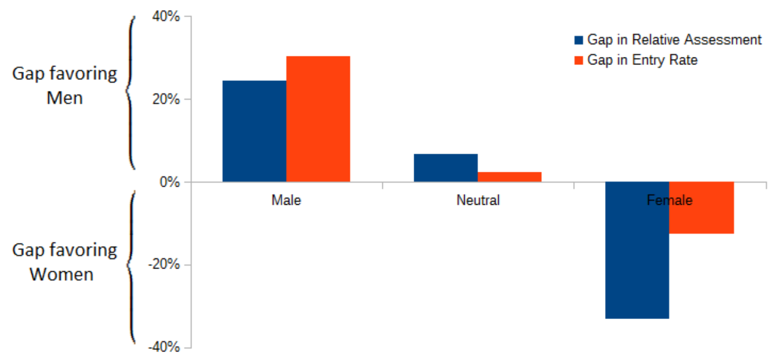
Notes: Signal is the random draw observed by the subject in the real effort task. *Ability* is the average score on the real effort task. *Conf.Est.* is a discrete variable between 1 and 7 denoting the confidence with which subjects make their estimation of ability. *Min.Ab.Win* is an 11-level discrete variables accounting for beliefs about the minimum ability required to win the prize in tournament *A*. *Risk Pref* is an 8-level variable measuring subjects' risk attitude. *Age* is subjects' age in years. *Difficulty* is a 5-level item measuring subjects' perceptions on task difficulty. *Taste Comp.* is a 5-level item accounting for self-reported taste for competition. *Bel.Ab.Men* and *Bel.Ab.Women* are the relative assessment for men's and women's average abilities. *Rel. Assessment* is subjects' relative self-assessment. *Prob(A)* is a situation specific dummy variable that takes the value 1 if a subject enters tournament *A*. *Prob(W_A)* is a situation specific dummy variable that takes the value 1 if a subject gets the prize at tournament *A*. *Rel. Assessment$_{ss}$*, *Prob(A)$_{ss}$* and *Prob(W_A)$_{ss}$* are the corresponding variables for the *single-sex* setting. Significance levels in parentheses.

First, it shows an important link between subjects' perceived maleness of the task and subjects' self-assessment (*Relative Assessment*) which is consistent with the self-stereotyping hypothesis: the more male the task is perceived as being, the higher men's self-assessment is and the lower women's is. Second, this pattern seems to be replicated in the entry rate into tournament A (*Prob(A)*), with men entering more often as the task is perceived to be more male. As regards gender gaps, Figure 1.6 suggests that the behaviors of the gender gap in both variables across perceptions are related to each

played in Figure 1.4, a subject claiming that his/her ability is in the interval 91-94 would be assigned a self-assessed relative position of 89% ([27+(5/2)]/33). All results are robust to the analysis of the original self-assessment provided by subjects. Results available are upon request.

other. Finally, Panel B of Table 1.3 also shows that the perceived maleness of the real effort task has an impact on men's and women's probability of obtaining the top prize W_A (*Prob*(W_A)), suggesting that the aforementioned behavioral differences impact the final outcomes of men and women.

FIGURE 1.6– GENDER GAP IN RELATIVE SELF-ASSESSMENT AND IN ENTRY RATE FOR TOURNAMENT A



Notes: Average gap in relative self-assessment (blue bars) and in entry rate for tournament A (red bars) for the *mixed-sex* environment relative to women's average values. A positive (negative) score means that there is a gender gap "favoring" men (women).

Panel B of Table 1.3 also shows the correlation between the perceived relative maleness of the task and the main outcome variables in the *single-sex* setting. Notice that in this setting the self-stereotyping mechanism is, by design, canceled out. Raw data suggest this is actually the case as the variable *Relative Maleness* is uncorrelated with self-assessments, behaviors and outcomes. This suggests that the correlations found in the *mixed-sex* setting are the result of perceptions and not of unobserved characteristics which correlate with perceptions.

Overall, Panel B of Table 1.3 provides some important insights from the raw data that fully support the theoretical model. In the rest of this section, regression analyses are performed to test these insights formally and bring to light the mechanisms that lie behind them.

1.3.3 REGRESSION ANALYSIS

In this section we carry out formal regression analyses to establish the relationship between perceptions as to the gender nature of the task and the three main outcome variables: self-assessment of ability (*Relative Assessment*), entry rate into the high-paying tournament (*Prob(A)*) and winning the high-tournament prize (*Prob*(W_A)). In particular, we take y_i as the outcome observed for subject i of any dependent variable of interest y and we perform the following analysis:

$$y_i = \beta_0 + \beta_1 Female_i + \beta_2 Rel.Maleness_i + \beta_3 Rel.Maleness_i * Female_i + \boldsymbol{\theta}_1^T \mathbf{x}_i + \boldsymbol{\theta}_2^T \mathbf{x}_y + \varepsilon_{iy}$$

where $\mathbf{x}_i$ is a vector of individual-specific control variables for subject i and $\mathbf{x}_y$ is a vector of variable-specific controls for the outcome variable y. The main independent variables are *Female*, *Relative Maleness*, and more importantly the interaction effect *Rel.Maleness∗Female*. This interaction effect captures how the gender gap in different dependent variables changes when the perceived maleness of the task increases. This specification enables us to separate pure gender effects (summarized by the parameter β_1) from gender differences arising from perceptions (coefficient β_3). It also enables us to compare men and women with the same perception *and* perception strength, and to check whether the impact of perceptions comes from men's reactions, women's reactions or a combination of the two.

SELF-ASSESSMENT

The estimation of their abilities that subjects draw up from the signal observed, and in particular their relative self-assessment, is a key variable for explaining behavior in later stages of the experiment. Here we are primarily interested in whether subjects' beliefs about the maleness of the task have any impact on the ability estimation process and, more importantly, whether they generate *self-assessment gaps*. In short, this subsection tests the self-stereotyping hypothesis posed by the theoretical model.

Table 1.4 shows the effect of perceptions concerning relative self-assessment through the coefficients of the *Rel.Maleness* and *Rel.Maleness∗Female* variables.

TABLE 1.4– OLS FOR RELATIVE (SELF) ASSESSMENT

Sample:	*All* (1)	*All* (2)	*Men* (3)	*Men* (4)	*Women* (5)	*Women* (6)
Signal	1.059***	0.958***	0.916***	0.785***	1.280***	1.299***
	(0.212)	(0.216)	(0.210)	(0.240)	(0.333)	(0.330)
Female	0.00543	0.0337				
	(0.0411)	(0.0421)				
Rel.Maleness	0.341***	0.303***	0.343***	0.337***	-0.192*	-0.230**
	(0.120)	(0.113)	(0.118)	(0.111)	(0.103)	(0.110)
Rel.Maleness∗Female	-0.559***	-0.535***				
	(0.158)	(0.151)				
Constant	-0.297	-0.497	-0.199	-0.295	-0.509*	-1.090**
	(0.201)	(0.346)	(0.190)	(0.504)	(0.298)	(0.462)
Other Controls	NO	YES	NO	YES	NO	YES
Session FE	YES	YES	YES	YES	YES	YES
Observations	120	120	62	62	58	58
R-squared	0.429	0.489	0.451	0.563	0.408	0.519

Notes: OLS for *Relative Assessment*. Other controls include taste for competition, risk aversion, age and difficulty of the task. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

In column (1), without controls, and in column (2), with controls, a look at the coefficient of the interaction effect *Rel.Maleness∗Female* shows that how male the task is perceived to be affects estimations by men and women differently, and that this actually determines the size and direction of the gender gap in relative self-assessment (p-value<0.01). In particular, this negative coefficient, taken together with the insignificant coefficient for *Female*, which tells us that under neutral perceptions (*Rel.Maleness*=0) men and women make comparable estimations, illustrates that when the task was perceived as male (*Rel.Maleness*>0) men relative self-assessment was significantly higher than that of comparable women, but when the task was perceived as female (*Rel.Maleness*<0) the opposite happened. Moreover, it also reveals that the size of these gender gaps is moderated by the believed size of the gender differences. This result is robust to the inclusion of other covariates and the use of different samples (see Table A.1 in Appendix A.3).

In an attempt to bring to light the reason for this behavior in the self-assessment gap, separate regressions for men (columns (3) and (4)) and women (columns (5) and (6)) are displayed in Table 1.4. This reveals that the average effect of *Rel.Maleness* is positive on men but negative on women, suggesting that actual perceptions simultaneously trigger the effects of positive and negative stereotypes, each targeting a different gender.[37]

SELF-SELECTION AND SET OF WINNERS

Subjects had to self-select into one of two tournaments labeled A and B under 9 different situations. In all them the prize offered to the winner in tournament A was strictly higher than the one offered in tournament B. In this section we look at whether beliefs about which gender is perceived to be ex-ante better at the real effort task can explain gender segregation across tournaments. In other words, we analyze whether the perceptions held by subjects affect their signal-contingent behavior and whether this effect of perceptions on behavior is different for men and women. The main dependent variables are the probability of entering tournament A (*Prob(A)*) and the probability of obtaining a prize in it (*Prob*(W_A)).

In the absence of the effects of stereotypes and perceptions, if subjects had followed an optimal decision rule during the experiment the rate at which men and women won the prize from tournament A should be the same (see Table A.4 in Appendix A.3). This immediately means that the optimal signal-based behavior should be the same for men and women and unrelated to their perceptions regarding the maleness of the task. In other words, neither *Relative Maleness* nor *Rel.Maleness∗Female* reflect differences in the characteristics of the subjects' pool that could explain any perception-contingent be-

[37] A natural follow up is to see what happens to the confidence level with what ability estimations are reported, as this could be crucial for decisions involving a ranking of abilities. An analysis of the self-reported confidence level (see Table A.2 in Appendix A.3) shows that confidence in estimation is not affected by how gendered the task is perceived but by the actual estimation. This is consistent with the existence of a *self-serving attribution bias* as those with higher estimations are more confident in their accuracy. That said, Table A.2 in Appendix A.3 illustrates the existence of a marginal pure gender effect on the level of confidence with which estimations are made.

havior.[38]

TABLE 1.5– PROBIT MODEL FOR THE PROBABILITY OF ENTERING TOURNAMENT A

Sample:	*All* (1)	*Men* (2)	*Women* (3)	*All* (4)	*Men* (5)	*Women* (6)
Relative Assessment				0.355***	0.518***	0.231**
				(0.0827)	(0.112)	(0.105)
Signal	0.471***	0.391**	0.582***	0.103	-0.0355	0.263
	(0.174)	(0.193)	(0.205)	(0.139)	(0.132)	(0.187)
Female	-0.0204			-0.0261		
	(0.0405)			(0.0354)		
Rel.Maleness	0.197*	0.222**	-0.139*	0.0859	0.0337	-0.0809
	(0.115)	(0.111)	(0.0796)	(0.0973)	(0.0968)	(0.0773)
Rel.Maleness∗Female	-0.309**			-0.112		
	(0.142)			(0.129)		
Tournament Controls	YES	YES	YES	YES	YES	YES
Session FE	YES	YES	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES	YES	YES
Number of Clusters	120	62	58	120	62	58
Observations	1,080	558	522	1,080	558	522

Notes: Marginal effects are reported. Other controls include risk preferences and believed minimum ability for winning at tournament A. Tournament controls include prize awarded by tournament A and allocation of prizes. Clustered standard errors at subject level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

In Table 1.5 the actual probability of entering tournament A is regressed as a function of the observed signal, the gender of the subject and his/her stated perception as to the maleness of the task. The analysis also includes the main control variables (risk aversion and belief as to minimum ability require to win the prize in tournament A) and tournament fixed effects.[39] As suggested by the descriptive statistics in Table 1.3 and Figure 1.6, there is a clear pattern in the gender gap on entry rate that seems to be driven by the actual perception of the task. Column (1) shows that there are no pure gender differences and that men and women who hold a neutral gender perception regarding the task (*Relative Maleness*=0) behave identically based on signal observed. This means that in the absence of any biased preconception as to which sex perform better on average, men and women enter tournament A at the same rate. However, the negative and statistically significant coefficient for the interaction effect between gender and perceived maleness (*Rel.Maleness∗Female*) indicates behavioral differences when perception departs from gender neutrality. In particular, it confirms that men (women) enter more often than women (men) if the task is perceived as male (female)

[38] When this same analysis is performed within the sample of marginal subjects (see Section 1.3.3), the variable *Relative Maleness* is marginally significant (p-value<0.1) but, importantly, the interaction effect *Rel.Maleness∗Female* is still far from being so. This implies that although *Relative Maleness* correlates to some extent with real ability in this subsample, its correlation is the same for men and women such no gender differences in the signal-contingent strategy should be expected.

[39] Results throughout this section are robust to the inclusion of other covariates and to the use of other statistical techniques and samples (see Table A.3 in Appendix A.3).

and that the size of this gap depends on the actual size of the gender bias perceived by the subjects.

Columns (2) and (3) of Table 1.5, look at the effect that the perceived maleness of the task has for each gender, separately. These two regressions suggest that *Relative Maleness* affects the entry rate of men positively and that of women negatively.

All this together reinforces the idea that the perceived gender nature of the task affects the threshold signal level required by subjects to participate in tournament A in the way suggested by the theoretical model: When the task is perceived as male (female), women (men) need to observe higher signals in order to self-select into the high-paying, more competitive tournament so they end up entering less often.

The patterns found at columns (1), (2), and (3) in Table 1.5 suggest that the effect of *Relative Maleness* on the behavior of agents is related to the behavior of the self-assessment gap shown in Table 1.4. In columns (4), (5), and (6) we perform the same analysis as in columns (1), (2), and (3) but including the relative estimation of abilities as an extra control. Importantly, in these cases the effect of maleness on determining the gender participation gap disappears. This strongly suggests that men and women who estimate the same ability are behaving similarly. Moreover this supports the claim that perceptions as to the maleness of the task have very little impact on participation decisions on their own, but most of the effect of perceptions on behavior is due rather to their effects on self-assessment (see Table 1.4).

Explaining participation gaps is important in understanding segregation, but the most relevant question for welfare analysis is probably whether the behavioral differences addressed above translate into gender representativity gaps in the set of winners. Notice that gender participation gaps should not necessarily translate into gender representativity gaps in the set of winners: there could be a participation gap favoring men, i.e. men enter more often, but one which is driven by men with very low ability entering tournament A too often. In that case, we would expect no gender representativity gaps in the set of winners as the gender differences in behavior would have no implications for the final outcomes.

Table 1.6 shows the effect of perceptions on the probability of finally obtaining the big prize. In column (1) it can be seen that when we only control for *Signal* the interaction effect *Rel.Maleness∗Female* is negative and significant, suggesting that the more male (female) the task is perceived as being, the more likely men (women) are than women (men) to end up winning the top prize. However, although based on signals both genders should win the high prize at the same rate independently of the perceptions held (see Table A.4 in Appendix A.3), winners are named based on real abilities, not on signals. So in column (2), we control for the real ranking of abilities within each session (*Ability Rank*), with ranking 1 being the worst performer, in addition to signals in order to test for any unbalancedness between genders in this regard that may

TABLE 1.6– PROBIT MODEL FOR THE PROBABILITY OF WINNING THE PRIZE AT TOURNAMENT A

	(1)	(2)	(3)	(4)
Signal	1.336***	0.561***	0.854***	0.284**
	(0.190)	(0.119)	(0.233)	(0.118)
Ability Rank		0.0197***		0.0192***
		(0.00119)		(0.00112)
Relative Assessment			0.308***	0.190***
			(0.102)	(0.0581)
Female	0.00394	-0.00534	-0.00298	-0.00798
	(0.0477)	(0.0306)	(0.0461)	(0.0296)
Rel.Maleness	0.223**	0.180**	0.129	0.122*
	(0.112)	(0.0789)	(0.118)	(0.0738)
Rel.Maleness∗Female	-0.328**	-0.202*	-0.178	-0.108
	(0.159)	(0.116)	(0.171)	(0.113)
Tournament Controls	YES	YES	YES	YES
Session FE	YES	YES	YES	YES
Number of Clusters	120	120	120	120
Observations	1,080	1,080	1,080	1,080

Notes: Marginal effects are reported. Tournament controls include prize awarded in tournament A and allocation of prizes. Clustered standard errors at subject level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

be correlated with the interaction effect *Rel.Maleness∗Female*. Our results show that although including this control reduces their effect, perceptions still have a significant and sizable impact on generating gender gaps. It is also worth mentioning that, as in all previous analyses, the dummy *Female* is not significant, which shows that when the task is perceived as gender neutral both genders win the top prize at the same rate –as would be expected– but once gender neutrality is broken, asymmetries arise in regard to the winning rate.

Finally, the role played by the self-stereotyping mechanism remains to be considered. Not surprisingly in view of the previous results, columns (3) and (4) show that once individual self-assessment is accounted for, perceptions held by subjects do not contribute to generating gender differences in the winning rate, independently of whether we control for the real ability rank or not. Notice that the only effect of *Relative Assessment* on the probability of winning the prize in tournament A is in shaping participation decisions. Thus, the results in Table 1.6 suggest that the interaction effect *Rel.Maleness∗Female* is found to be significant in columns (1) and (2) because it is actually evidencing the effect of perceptions in self-assessment (Table 1.4) and its impact on behaviors when choosing between tournaments (Table 1.5).[40]

[40] This is related, though in a completely different context, to the findings on Reuben et al. (2012). They find that women are less likely to be appointed as group leaders when the task to be performed relates to mathematical skills, but that this underrepresentation can be almost explained by differences in self-assessment. This enables the authors to conclude that in their experiment discrimination against women plays no part, but that lower self-assessment by women relative to men does.

MARGINAL SUBJECTS

The above results are fully in line with the model. However, the model makes a stronger prediction: although self-stereotyping affects self-assessment for the full sample, its behavioral consequences should show up only in a given subset of subjects. In other words, according to the self-stereotyping mechanism displayed by the theoretical model, the effects from Tables 1.5 and 1.6 should be driven by those subjects who are borderlines as regards being winners in tournament A. This excludes those with very high/low signals, who are almost sure they will/will not obtain a prize in tournament A, so their decision to participate/not participate should not be affected by their particular perception about the gender affinity of the task. To address this issue, we construct a dummy variable called *Marginal* which takes a value of 1 when the subject in any particular situation can be considered to be reasonably uncertain of his/her chances of winning in tournament A and 0 otherwise.[41]

TABLE 1.7– PROBIT FOR THE PROBABILITY OF ENTERING TOURNAMENT A FOR MARGINAL AGENTS

Sample:	*Marginal* (1)	*Nonmarginal* (2)	*High Signal* (3)	*Low Signal* (4)
Signal	1.259	0.380**	-2.135**	0.205
	(0.782)	(0.163)	(0.952)	(0.178)
Female	0.00193	-0.0386	-0.0606	-0.0451
	(0.0570)	(0.0435)	(0.0500)	(0.0667)
Rel.Maleness	0.343*	0.113	0.0269	0.274
	(0.197)	(0.120)	(0.115)	(0.191)
Rel.Maleness∗Female	-0.598**	-0.166	-0.131	-0.266
	(0.242)	(0.156)	(0.182)	(0.242)
Tournament Controls	YES	YES	YES	YES
Session FE	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES
Number of Clusters	97	120	71	72
Observations	342	738	354	384

Notes: Marginal effects are reported. Other controls include risk preferences and believed minimum ability for winning at tournament A. Tournament controls include prize awarded by tournament A and allocation of prizes. Clustered standard errors at subject's level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

By performing this classification and replicating the analysis in column (1) of Table 1.5 separately for subjects considered as marginal and for those who are not, we find that the effect of maleness in explaining the gender participation gap is only significant for the marginal individuals (Table 1.7, columns (1) and (2)). It can be also checked that the coefficient in this subsample of marginal subjects is greater (β: -0.60

[41] As a rule of thumb, we consider 30% of the sample as marginal. This subsample was identified as follows: We sorted all subjects within each session by their observed signals. Then, for a situation in which tournament A offers a fraction of prizes δ_A, we considered that those who observed a signal above (below) the percentile $[1-\delta_A]+15$ ($[1-\delta_A]-15$) of the signal distribution would behave in the same way independently of the perception held. The results remain quite similar when different interval amplitudes are used. Lastly, notice that there are agents who are considered as marginal under several δ_A's.

vs -0.33) and more significant (*p*-value: 0.01 vs 0.03) than for the whole sample. As a final test, notice that among those who are not considered marginal a distinction can be drawn between those with high and low observed signals. Separate analyses for these two groups (columns (3) and (4) of Table 1.7), reveal that *Relative Maleness* has no significant effect either. The results shown in Table 1.7 show that, as predicted by the model, stereotypes affect the whole population but their behavioral consequences in this particular setting show up only locally, in some particular parts of the sample. Similar results hold when the probability of winning the prize in tournament A is considered. Results are available upon request.[42]

Robustness: Tournament Self-Selection in a Single-Sex Environment

In the fourth stage of the experiment, subjects go through the exact same choices (self-assessment and choice of tournament) but in a single-sex environment. The difference with respect to the previous situation is that subjects now interact with and see the information for their own gender only. Importantly, except for the fact that subjects now observe the distribution of abilities of those of their own gender, the information that they hold in terms of their own and others' abilities is the same. In particular, they observe the same signal provided in stage 2 of the experiment and no information is provided about their payoff at stage 3.

According to the theory, in this environment perceptions should not play any role in affecting subjects' decisions: Subjects observe the ability distribution *only for their own gender*, so whether a subject believes the task to be male/neutral/female is completely irrelevant as in all cases he/she sees him/herself in an *ex-ante homogeneous* group in which perceptions of this type cannot play any role. Consequently, under this *single-sex* environment there is no scope for self-stereotyping.

The contribution of this treatment to the general picture of the experiment is that it enables us to further reject other potential selection biases regarding perceptions or, on the contrary, to admit that there are important unobserved characteristics which play a role in agents' decisions. For example, men who claim that the task is male and men who claim that it is female could differ in key aspects such as general optimism or self-esteem. If so, we should still observe an effect of perceptions on the new estimates of ability and on signal-based behavior. On the other hand, the loss of the effect of perceptions in this *single-sex* environment would be further evidence in support of the initial premise that the channel driving the results in subsections 1.3.3 and 1.3.3 is actually self-stereotyping rather than unobserved differences in subjects' characteristics across perceptions.

[42] Dohmen and Falk (2011) also apply this idea of marginal subjects. In their paper subjects have to self-select into either a fixed or a variable payment scheme. They find, consistently with our findings, that personal characteristics are stronger predictors of behavior for marginal subjects than for nonmarginal ones.

TABLE 1.8– RESULTS FOR THE SINGLE-SEX ENVIRONMENT

Dep. Variable:	*Rel. Assessment$_{ss}$*	*Prob(A)$_{ss}$*	*Prob(A)$_{ss}$*	*Prob(W_A)$_{ss}$*	*Prob(W_A)$_{ss}$*
Sample:	*All*	*All*	*Marginal*	*All*	*Marginal*
	(1)	(2)	(3)	(4)	(5)
Signal	0.993***	0.537***	1.505*	1.144***	1.687*
	(0.210)	(0.153)	(0.788)	(0.278)	(0.951)
Female	0.0482	-0.00878	-0.0712	0.0449	0.0675
	(0.0382)	(0.0420)	(0.0680)	(0.0522)	(0.0839)
Rel.Maleness	0.150*	0.0590	-0.0645	0.158	-0.0855
	(0.0876)	(0.126)	(0.206)	(0.158)	(0.249)
Rel.Maleness∗Female	-0.0849	-0.136	-0.128	-0.149	-0.174
	(0.124)	(0.146)	(0.248)	(0.194)	(0.310)
Tournament Controls	–	YES	YES	YES	YES
Session FE	YES	YES	YES	YES	YES
Other Controls	YES	YES	YES	NO	NO
Number of Clusters	–	120	97	120	97
Observations	120	1,080	342	1,080	342
R-squared	0.555				

Notes: Column (1) shows the OLS estimates for the subject's relative self-assessment. Columns (2)–(3) and (4)–(5) show the marginal effects of the probit model for the probability of choosing tournament A and the probability of getting W_A for the full sample and the subsample of agents classified as marginal, respectively. Other controls include taste for competition, risk aversion, age and difficulty of the task for column (1) and risk preferences together with believed minimum ability for winning in tournament A for columns (2) and (3). Tournament controls include prize awarded by tournament A and allocation of prizes. Robust standard errors for column (1) and clustered standard errors at subject level for columns (2)–(5) displayed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.8 shows the main analyses performed in Tables 1.4, 1.5, 1.6 and 1.7 for the data from this *single-sex* environment. The first noteworthy finding is that having information about the distribution of abilities of their own sex cancels out the effects of subjects' perceptions on self-assessment (column (1)). Notice that the coefficient for the interaction *Rel.Maleness∗Female* switches from being strongly significant (p-value<0.01) to statistically insignificant (p-value=0.49).

Since in this setting no differences of any kind are observed in self-assessment, the natural expected result would be to find no differences in behavior or outcomes. Estimation results confirm that this is the case. In column (2) it can be confirmed that for the same observed signal men and women have the same probability of entering tournament A in the *single-sex* environment. Furthermore, column (3) shows that this is also the case for those agents classified as marginal. Columns (4) and (5) show that perceptions do not affect subjects' probabilities of winning the top prize either for the full sample or for the sample of marginal subjects.

These findings reinforce the initial premise that differences in behavior and outcomes are driven largely by differences in self-assessment and that perceptions affect self-assessment significantly, but only when the information can be accommodated to preset perceptions. In other words, data from this stage supports the initial hypothesis that perceptions –and by extension stereotypes– are critical when comparing social groups, with confidence not being affected directly but by means of social comparisons.

However, despite the great consistency shown with the hypothesis put forward, several caveats regarding the design of the experiment lead us to take data from this stage with a grain of salt. Firstly, subjects may be influenced by past decisions and may thus show some degree of inertia in behavior which pushes them to replicate past decisions at this stage. Notice that this point is contrary to our findings as it implies that perception-related behavior must be still present through this inertia. The second caveat is that the aim of the experiment has been partially disclosed to subjects, given that in stage 3 they were asked about perceptions. They could then react to that knowledge by changing their behavior. The last concern is that subjects may learn more accurately what their real ability levels are as we provide them with new information. This may make the signal more informative, so the role of perceptions will diminish. However, this learning argument seems to be only weakly supported by data. First, for all experimental sessions a Kolmogorov-Smirnov test rejects the hypothesis that the distributions of abilities of men and women are different. Therefore the picture of the distribution of abilities in the first part of the experimental session and the picture observed in this *single-sex* stage should be similar. Consequently, switching from the *mixed-sex* stage to the *single-sex* one should not greatly improve the quality of the signal received. In addition, subjects' guess rates for *Rel.Assessment* and $Rel.Assessment_{ss}$ are 32% and 38%, figures that are not statistically different (p-value>0.1). The drawback is that since the *mixed-sex* and *single-sex* distributions of abilities shown to subjects are very similar, subjects who hold non-neutral perceptions may learn that there are no gender differences and thus behave in the same way as agents who hold neutral perceptions. Although we cannot rule out this last possibility, this alternative interpretation of the results of this section still confirms that the findings in the *mixed-sex* environment are driven by perceptions and not by unobserved characteristics.[43]

1.4 Concluding Remarks

In this paper we present a simple yet natural way of introducing stereotypes into economic modeling. Given the large amount of research, especially in social psychology,

[43]We have two major competing explanations for the results on this section. On the one hand, according to the theory in Section 2, subjects who observe the distribution of abilities of an ex-ante homogeneous group will not self-stereotype independently of their beliefs about the gender nature of the task. On the other hand, it may be the case that all agents learn that there are no gender differences in performance so they update their priors regarding the gender nature of the task and conclude that it is gender neutral. In both cases, the result for the *single-sex* environment is the same: *Maleness∗Female* is not significant. In the first case, this would be because perceptions do not play any role in this particular setting. In the second case it is because at this stage of the experiment all agents perceive the task as neutral. In short, both explanations conclude that, one way or another, perceptions are no longer relevant in this *single-sex* setting. Thus, the *single-sex* setting is identical to the *mixed-sex* one but free of perception effects, so any variable other than perception that lies behind the results in the *mixed-sex* setting should still be in play.

which suggests that stereotypes have important effects on self-assessment, we believe there is a need to explore more deeply when and how stereotypes matter in relevant economic situations.

Accordingly, this paper contributes to the relevant literature by applying the above model dealing with stereotypes and self-assessment in a strategic situation which represents economically relevant settings. In particular, we look at how stereotypes affect self-selection into different tournaments offering different prizes. The theoretical model identifies the channels through which stereotypes may affect agents' decisions. In short, the model predicts that the existence of a stereotype will generate *self-stereotyping* which ultimately affects agents' decisions as to *where* to compete in a self-fulfilled way. These predictions are strongly supported by the experimental data collected.

The setting studied in this paper concerning the way in which stereotypes affect decisions on where to compete may be of great interest in explaining the gender-based segregation observed in education and labor choices.[44] In particular, the model presents a novel approach explaining gender-based segregation through the impact of stereotypes on the supply side of the labor market. This idea is supported by the observation that, in many cases, the segregation observed in the data is consistent with existing gender stereotypes (Cejka and Eagly (1999); Barbulescu and Bidwell (2013); Haveman and Beresford (2012)).

Understanding whether gender-based segregation in the labor market can be explained through existing stereotypes is an extremely important issue because women are the targets of negative stereotypes in at least three areas strongly related to wage levels: quantitative skills (Frome and Eccles (1998); Nosek et al. (2009)), leadership (Schein (2001); Atwater et al. (2004)), and general IQ (Furnham and Gasson (1998); Furnham et al. (2002); Petrides et al. (2004); Bian et al. (2017)). According to the model proposed in this paper those stereotypes hurt women's self-assessment and therefore undermine their professional goals, leading them to self-select into lower paying itineraries than men.

In addition, as discussed on Section 2, the model also provides interesting insights into how stereotypes affect choices, and suggests directions for policies dealing with them (see corollary 1.2.3 and the discussion therein). In particular, a promising result is that increasing the quantity and quality of the feedback received by those who are the target of a stereotype may alleviate its effects.

[44] Although the model developed in this paper presents a setting that deals with *horizontal segregation*, it can be extended straightforwardly to a *vertical segregation* setting. Notice that in a setting where agents have to self-select into competing in tournaments that involve different abilities (e.g. science and arts) and in which one social group is stereotyped in the skill relevant to the high paying tournament (say quantitative skills at science), the results found here will also apply.

Overall, this paper shows strong evidence that stereotypes are very likely to be playing an important role in shaping segregation –particularly gender-based segregation– at the labor market through self-selection. Needless to say, we do not posit that all the gender segregation observed in the market is generated through this mechanisms. Obviously there are other factors that are very likely to play a role in shaping it on both the demand and supply sides. For example, continuing with stereotypes, segregation could be due to gender roles. According to the theory put forward by Akerlof and Kranton (2000), it may well be the case that women choose different jobs than men because they are supposed to do so. In this setting stereotypes play the role of social norms. Another important component of segregation can be discrimination. On the demand side discrimination –both taste-based and statistical– can explain the segregation observed in the labor market by restricting the access of women to high-paying positions. On the supply side, expectations about discrimination may also deter women from applying for some positions. Moreover, differences in risk aversion could also generate sorting differences as less competitive itineraries usually involve lower degrees of uncertainty. Lastly, it may be also the case that men and women simply differ in their choices because of differences in preferences (Croson and Gneezy (2009)).

Considering the above, the extent to which this self-stereotyping mechanism is playing a role is thus open to argument. However, in light of the experimental results presented in the paper, it cannot be denied that the self-stereotyping mechanism has potential for explaining gender differences in the real world. Notice that the experiment involves a setting which is free from discrimination and in which tastes and social norms should not play any role. The case of risk aversion deserves further discussion. In our sample, as usual, women are more risk averse than men. Thus, it could be argued that this explains why women choose tournament A less often than men. However, this same argument would lead us to expect women to also choose tournament A in the *single-sex* environment less often than men, which is not the case.[45] Therefore, the results obtained in the experiment cannot be attributed to any of the above alternative explanations. Thus, the experimental evidence presented in this paper lead us to believe that the interaction between stereotypes and self-assessment plays an important role in shaping the labor decisions of agents both previously and in the labor market, and thus contributes to our understanding of education and labor segregation.

Finally, we would like to point out a methodological caveat. The data collected in our experiment shows that perceptions about real effort tasks can be critical in explaining subsequent behavior related to that task. Notice further that the consequences of

[45] In addition there is no an intuitive way in which risk aversion should matter in determining subjects' self-assessment, which we show explains differences in entry rates, as subjects must enter a point estimate, which means that this thus is independent of the degree of risk aversion. This is confirmed by the observation that risk aversion is found not to be significant in determining subjects' self assessment (see Table A.1 in Appendix A.3).

those perceptions manifest themselves in the data *before* we collect them, so it is not a priming issue but something that subjects have in mind during the experiment (see also Grosse et al. (2014); Bian et al. (2017) or Iriberri and Rey-Biel (in press) for similar findings). We therefore think that in studies dealing with real effort tasks in which uncertainty of the kind covered on this paper –i.e. self-assessment– plays a role, it may be important to elicit perceptions as to the gender nature of the task to obtain a potentially important control for subsequent econometric analysis.

Chapter 2

Women ask for less (only from men): Evidence from alternating-offer bargaining in the field

2.1 INTRODUCTION

The gender wage gap has long been an important object of study in economics. Although it has shown a decreasing trend over time, its persistence in developed countries challenges the classical explanations based on differences in human capital, preferences or statistical discrimination (Blau and Kahn (2000)).

Gender differences in negotiation have been put forward as an alternative explanation for the gender gap. Starting wages are often the result of bilateral negotiation. The influential book by Linda Babcock and Sara Laschever "Women don't Ask: Negotiation and the gender divide" reveals important gender differences in the likelihood of negotiating. A study mentioned in the book shows that among graduates of Carnegie Mellon University 57% of men negotiated the starting salary offered, while only 8% of women did so. Moreover, wages are also affected by negotiations that come later during one's career, e.g. for pay increases. If women are less likely to negotiate starting salaries, and/or if women are less likely to ask for a pay increase, this will clearly go some way towards explaining the gender wage gap (Azmat and Petrongolo (2014); Card et al. (2016)).

Alternating-offer bargaining at a TV show offers a unique opportunity to observe bargaining outcomes and behavior in a real-life situation with sizable stakes. In the show a contestant who plays the role of the proposer in the bargaining, is asked a question. The contestant cannot provide the answer herself/himself but has three minutes to find someone, who plays the role of responder, from whom the answer can be bought in an alternating-offer setting. In a typical bargaining round, the proposer makes an initial offer which may be accepted or rejected by the responder. If the responder rejects it then he/she can post a demand, which may be accepted or rejected by the proposer. Bargaining can extend over any number of rounds within the three minute limit. If the answer is correct, the pie is divided up as agreed in the bargaining.

In the paper we study two main research questions. First, are male and female contestants equally likely to choose male and female responders to bargain with? We study gender differences when choosing the gender of the bargaining partner. Second, is bargaining between male-male, male-female, female-male and female-female matchings different? We study both gender differences and gender interaction effects in bargaining.

We find that although both male and female proposers are more likely to choose male responders, male proposers show a stronger preference for male bargaining partners. This is consistent with taste-based discrimination but also with gender differences in beliefs.

With respect to bargaining outcomes, we find that negotiations between male proposers and female responders stand out from negotiations between all other gender combinations: they are the most favorable for men and the least favorable for women in terms of earnings. In explaining the details of our bargaining setting, we argue that proposers hold a stronger position than responders. In that sense, it is precisely when the proposers are male and the responders are female that men are found to end up with higher bargaining outcomes.

To understand the gender interaction effects in bargaining outcomes, we then look at gender interaction effects in bargaining behavior. We analyze offers, demands and probabilities of accepting by responders and proposers. We find no differences in opening offers between male and female proposers, or in opening offers to male and female responders. More interestingly, we find that it is women who discriminate between male and female proposers, demanding less from men than from women, which explains why the matching between male proposers and female responders end up with the highest earnings for the proposer. It is not the case that men offer less to women but it is women who demand less to men, when men hold the role of the proposer and women hold the role of the responders.

Male proposers who choose female responders are not of a particular type and female responders who are chosen by male proposers are not of a particular type. Based on important observable characteristics of both proposers and responders, we show that male proposers who choose male responders do not significantly differ from male proposers who choose female responders. Similarly, female responders who are chosen by male proposers do not significantly differ from female responders who are chosen by female proposers. In addition, using the probability score matching technique, we find that our results are not driven by selection problems. On the downside, we cannot rule out selection based on characteristics that are unobservable, such as beliefs.

Gender differences in bargaining have been studied by economists. For example, male proposers' behavior has been analyzed in studying discrimination by carrying out field experiments in which the gender of potential scripted buyers is varied (Ayres (1991); Ayres and Siegelman (1995); Castillo et al. (2013)). To study gender differences in wage negotiation, Säve-Söderbergh (2007) uses wage bids and wage offers of recent graduates and finds that women post lower wage bids, and receive lower offers. More recently, Leibbrandt and List (2014) find, using a field experiment, that women are less likely to negotiate their salary when wages are not described explicitly as negotiable, but that the difference disappears when they are described as negotiable. Closer to our setting, using also data from a TV-show, van Dolder et al. (2015) find gender to be an insignificant determinant of contestants' initial claims, their hardball announcements and concessions, and the shares they end up with. There are however, important differences between our setting and theirs. First, bargaining is between three parties in their setting, while it is between two parties in ours. Second, bargaining in their setting consists of an allocation problem, where the pie is split into three unequal shares and contestants have to decide who gets which, while in ours it consists of an alternating-offer bargaining. Most importantly, they study only gender differences and not gender interaction effects. Economists have also studied gender differences in controlled settings such as the laboratory, mostly using the ultimatum game, which represents a reduced-form bargaining setting, as it allows for a single offer (or demand) and the response to it. Rigdon (2012) finds that women demand less than men in a demand-ultimatum-game in the laboratory, and Andersen et al. (2013) find that gender differences in bargaining depend on culture. More recently, Exley et al. (2016) study gender differences in the choice to negotiate, and in their baseline treatment, where subjects are *forced* to negotiate, they find that men and women achieve similar returns.

Gender interaction effects in bargaining have received less attention. Given that bargaining requires interaction between two agents, gender differences in one role may crucially depend on the gender of the interlocutor. Existing studies based on field data or field experiments do not study gender interaction effects, either because the gender of the person in one role is not known (e.g. Leibbrandt and List (2014)), or because there is not enough variation (e.g. Castillo et al., 2013). Economists are thus limited to the use of laboratory experiments. Using mostly face-to-face ultimatum games, Eckel and Grossman (2001); Solnick (2001) show that offers to women are lower than offers to men, and that women are more likely to accept offers.[1] Sutter et al. (2009) find more competition and retaliation between same gender matchings than mixed gender matchings using the power-to-take game. Eriksson and Sandberg (2012) find that women are less likely to initiate a negotiation if they are matched with a female partner. More closely related to our setting, Dittrich et al. (2014), using a laboratory face-to-face alternating-offer wage-bargaining game, find that starting salaries offered by men to

[1] To be precise, Eckel and Grossman (2001) find that women are more likely to accept offers, while Solnick (2001) finds that women are more likely to accept offers from male proposers than from female proposers.

women are lower than those offered by women to men, resulting in significant gender interaction effects on wage-bargaining outcomes. This line of research shows that observed gender differences depend crucially on the gender of the interacting individual, so such interactions deserve equal attention.

Our setting offers multiple advantages over laboratory experiments. First, the pie to be divided is worth an average of 345 Euro, so the stakes are sizable and significantly bigger than in a typical laboratory experiment. Please, see van Dolder et al. (2015) for similar arguments on the importance of stakes. Second, except for the three minute limitation the bargaining is not structured, so the observed negotiations in the tv-show are closer to the type of bargaining that happens in real-life than the structured negotiations in the laboratory. The setting also offers some advantages over standard field data given that the bargaining process is recorded. First, the setting allows not only bargaining outcomes, such as whether the negotiation was successful, proposers' and responders' bargaining outcomes or the duration of the negotiation but also bargaining strategies and bargaining behavior itself to be observed, such as round by round offers, demands and probabilities of accepting. Second, there is gender variation in the roles of both proposers and responders, so it is possible to study not only gender differences but most importantly, also gender interaction effects within different bargaining matchings.

Our setting also presents limitations that can affect external validity. One limitation that should be kept in mind is the extent to which the individuals studied are representative, as they are willing to, and indeed do, participate in a TV show. In that respect, the participation bias might be lower than in other shows as the show in question is recorded in the main streets of major cities in Spain and not in a studio. The recruitment process is therefore somewhat non-standard, as the contestants are picked on the spot. Second, audience effect must be taken into account, as the observed behavior might be influenced by it. These two limitations are common to all studies that use behavior at a TV show (List (2006); Post et al. (2008); van Dolder et al. (2015)). Finally, the interactions we observe are one-shot, while reputation built in repeated-interactions is important when bargaining in the labor market.

Overall, our findings are consistent with the literature that finds that women demand less and are offered less. In addition, the paper makes three important contributions. First, in line with the few papers that look at gender interactions (Eckel and Grossman (2001); Solnick (2001); Dittrich et al. (2014)) it confirms that gender interactions are crucial in understanding gender differences. When looking at both gender differences and gender interaction effects, we show that looking at only gender differences can show a misleading interpretation of results. As an example, the findings in this paper are consistent with findings by Säve-Söderbergh (2007) in the sense that women demand less. However, we show that this is only the case when the interaction is with men, i.e. women demand less *only from men*. Second, the paper offers evidence based on unstructured bargaining behavior observed in the field with sizable stakes, supplementary to the bargaining behavior observed in the laboratory. Finally, and in sharp contrast with Dittrich et al. (2014), we find that in our setting the determinant behavior that results in gender interaction effects does not reside in proposers' initial offers but in responders' demands.[2] In other words, no evidence of such discriminatory behavior is found in initial offers. It is female responders who demand less from male proposers, both in initial demands and subsequent demands.

The rest of the paper is organized as follows. Section 2.2 describes the TV show, the data and the identification strategy. Section 2.3 contains the analysis and results. We first analyze gender differences in the choice of the gender of the bargaining partner (Section 2.3.1). Then we study gender differences and gender interaction effects in bargaining outcomes (Section 2.3.2). Thirdly, we analyze bargaining behavior studying actual offers, demands and probabilities of accepting (Section 2.3.3). Section 2.4 shows the results of three robustness tests that address both the selection and unbalancedness problems in our data. Section 2.5 concludes.

2.2 THE DATA

2.2.1 THE SETTING: ALTERNATING-OFFER BARGAINING IN A TV SHOW

We use the bargaining behavior from the Spanish TV show *Negocia como puedas*.[3] This was a quiz-show that took place in the streets of major cities in Spain in the summer of 2013.

In a typical episode, the presenters pick a contestant, who is endowed with 100 Euro. He/she is then asked a question whose answer he/she cannot provide him/herself, independently of whether he/she knows the answer, and has three minutes to find someone on the street and negotiate a prize for the right answer via an alternating-offer bargaining. The contestant searches for a potential responder. The contestant can approach as many potential responders as he/she wants, and only when the potential responder provides an answer that is considered satisfactory to the contestant, he/she starts bargaining. In the bargaining, if an agreement is reached and the answer is correct then the 100 Euro is divided up as agreed. If the contestant does not reach an agreement within the three minutes, the game ends and he/she wins nothing. Negotiations may end with no agreement for two reasons: The time limit may be reached with no agreement, though this occurs very rarely (in 4% of the breakdowns), or (in the remaining 96% of the breakdowns) the contestant may decide to break off the negotiation and look for someone else to start a new negotiation. This is allowed as long as it occurs within the three minute time limit. The game is repeated up to 4 different times

[2] Although the setting in Dittrich et al. (2014) is closest to ours in that the negotiation occurs as in an alternating-offer game, there are also important differences. The most important one is that, contrary to our setting, proposers do not choose to break the negotiation, but the matching and breaking of negotiation is done randomly and exogenously.

[3] The show's name translates as "Bargain How You Can". It was shown on national TV channel Cuatro.

with different bargaining partners, as long as the questions are answered correctly. The answers to the questions are quite easy and trivial, and correct answers were given on 87% of occasions.[4] We refer here to the four potential repetitions as stages. In the second stage the contestant is endowed with a further 200 Euro in addition to the money kept from the first stage, so the maximum size of the pie to be divided up in the second stage is 300 Euro. In the third stage there is an extra endowment of 300 Euro, so the maximum size of the pie is 600 Euro. The fourth and final stage is optional, but if the contestant decides to continue the extra endowment is 1000 Euro, such that the maximum size of the pie to be divided up is 1600 Euro. The amount of the pie is known only to the contestant. Note that the stakes are real and sizable.

We refer to the contestant as the proposer and to the person selected on the street as the responder. In a typical round of bargaining, the proposer starts with an offer, which the responder either accepts or rejects. If the responder rejects, then he/she can post a demand, which may be accepted or rejected by the proposer. We refer to the combination of proposer's offer, responder's response, responder's demand and proposer's response to the demand as a *round*. Note that a round does not need to be complete as one of the bargaining partners can remain silent in a specific round. However, in each round at least one of the bargainers must make an active move by posting an offer or a demand, and responding to an offer or to a demand. The bargaining process can take any number of rounds within the three minute limit. In the data used here the shortest bargaining process lasted for one round while the longest lasted for 15.

Given our setting, we argue that the proposer's position is strong while the responder's position is weak. First, it is the proposer who starts the negotiation with an opening offer (95% of the time). Research into bargaining has shown that the starting offer is an important determinant of outcomes (see for example Van Poucke and Buelens (2002)). Second, proposers know with certainty the actual size of the pie to be divided, while responders do not. Finally, and more importantly, while the proposer can break off the negotiation and look for another responder at any time, the responder has no such option. These three characteristics make the positions of the proposer and responder asymmetric, giving the former a strong role and the latter a weak one.

[4] If the contestant reaches an agreement but the answer is incorrect, the amount negotiated is deducted. He/she is then allowed to use the *wildcard*, which consists of a phone call to a friend who must provide the correct answer, while the contestant is allowed to help but not to use any word from a list of forbidden words. The wildcard therefore, enables contestants to continue in the game even though the answer to the question is incorrect. There is only one wildcard. Most importantly, to use the wildcard an agreement must have been reached, such that the bargaining behavior is equally valid regardless of whether the contestant starts bargaining because he/she thinks the answer is correct or because he/she is running out of time and is doing so only to be able to use the wildcard.

2.2.2 The Database

We have created a panel database that records the bargaining behavior in the TV show.[5] For the analysis, we use bargaining behavior from 428 matchings between 134 proposers and 428 different responders.[6] There are 73 female proposers (54%) and 61 male proposers (46%), and 157 female responders (37%) and 271 male responders (63%). The sample of proposers is balanced in terms of gender composition, but the sample of responders contains more men than women.[7]

The proposers introduce themselves briefly, providing their names, ages and occupations. In terms of occupation we identify those who are students, retired, and unemployed. We also classify proposers as having a low-level or high-level occupation.[8] The responders only reveal their names, so we have no information on their ages and occupations. However, given that we have footage from the TV show, we have elicited subjective perceptions of the ages of both proposers and responders (a scale from 1 (below 30) to 6 (above 70)), socio-economic status (0 for medium-low, and 1 for medium-high) and attractiveness (using a scale between 0 (very unattractive) and 8 (very attractive)). We showed snapshots of all participants - both proposers and responders - to 10 different raters and averaged them.[9] The instructions given to these raters are available upon request.

We also recorded the pie to be shared, *Pie*, the stage at which the contestant is, *Stage*, and the deviation from the mean pie at each stage, *(Pie-Mean) by stage*.[10] *Bargain-*

[5] We have saved all the original TV shows so the database can be replicated and recoded.

[6] There is a total of 437 matchings. Information from 9 bargaining matchings was dropped because the presenter made comments about the proposer's behavior, for example accusing the proposer of being stingy, which influenced both the bargaining outcome and behavior.

[7] A Chi-Square Goodness of Fit test does not reject the null hypothesis that the sample of proposers is representative in terms of gender (p-value 0.30), but it does reject it for the sample of responders (p-value 0.00).

[8] We follow the 2 digit classification used by the Spanish National Institute of Statistics (INS) and consider as low-level those occupations whose digits are strictly above 49, along with occupations in the armed forces. The low-level occupation variable includes occupations such as waiters and hairdressers. Accordingly, we consider as high-level those occupations whose digits are below 49. The high-level occupation variable includes occupations such as engineers and clerical jobs.

[9] Each rater evaluated 125 participants in about an hour and was rewarded with a fixed amount of 15 Euro. For proposers, five men and five women were recruited. For responders, if the responder was chosen by a male proposer then all 10 raters were male, while if the responder was chosen by a female proposer then all 10 raters were female. Since we have the proposers' real ages, we have computed the correlation between the real age and the perceived one. The result is 0.94, which confirms the validity of these ratings. In addition, given that we are using the average of the ratings, we have calculated the interim reliability scale (Cronbach's alpha), which gives a score of 0.99 for perceived age, 0.94 for socio-economic status and and 0.85 for attractiveness, confirming the alignment of raters on each of the variables.

[10] As is clear from the description of the TV show, the size of the pie increases proportionally from stage to stage, so these two variables are highly correlated (over 0.9). As we can not include both, pie and stage, we decided to control for the stage, which ranges from 1 to 4, and also for the deviation from the mean pie at each stage.

ing Time refers to the time in seconds left for bargaining, where the maximum is three minutes. Finally, we classify each bargaining matching according to whether the question asked is perceived to be male, neutral or female (*Male Question*, *Neutral Question*, and *Female Question*). Remember that the contestant is asked a question whose answer he/she needs to buy in an alternating-offer bargaining. So the choice of the bargaining partner, and in particular the gender of the chosen bargaining partner, is influenced by whether the question is male or female. We gave the questions and answers to two different people, one a man and the other a woman, separately and asked them to classify the questions as male, neutral or female, in terms of how likely men and women are to give a correct answer.[11] Instructions received by these raters are available upon request.

Table 2.1 presents the mean values and standard deviations for all the demographic and control variables for the proposers and the demographic variables for responders, overall (column 1), and for female (column 2) and male (column 3) participants, separately. Column 4 reports the *p*-values for the F-Test of equality of variable means across genders. As the data show, the only significant difference between male and female proposers is that men are less attractive and more likely to hold a low-level occupation, while women are more likely to hold a high-level occupation. We control for all these characteristics when analyzing the behavior of proposers. For responders, the only significant difference is that female responders are more attractive than male responders.

The rest of the columns in Table 2.1 compare the characteristics separated by the gender combinations of the bargaining matches. With 73 female and 61 male proposers, and 157 female and 271 male responders, we end up with 139 female-male, 92 female-female, 65 male-female, and 132 male-male bargaining matchings. For proposers, this enables us to see whether male/female proposers who choose a male responder differ in their characteristics from male/female proposers who choose a female responder. Similarly, for responders, this enables us to see whether male/female responders who are chosen by male proposers differ in their characteristics from male/female responders who are chosen by female proposers. The most important variable is that of the male/neutral/female nature of the question. Around 70% of questions are classified as neutral, 16% as male and 16% as female. Male and female proposers do not show significant differences in being asked male or female questions, see *p*-values in column 4. More importantly, as expected, the male or female nature of the question significantly affects the gender choice of the bargaining partner. When asked a male question, contestants look for a male responder, but when faced with a female question they look for a female responder, see *p*-values in columns 7 and 10. We find no evidence of important differences in the rest of the variables, with two minor exceptions: female

[11] The raters initially agreed on their classification in 70% of questions. Among the questions over which they disagreed, all but 5 were questions that one rater classified as neutral but the other classified as male or female. After the initial perceptions of each rater were collected, they discussed the questions over which they disagreed face to face and reached an agreement on all of them.

TABLE 2.1– DESCRIPTIVE STATISTICS: PROPOSERS AND RESPONDERS

Proposers	Overall	Female	Male	p-value	Female-Male	Female-Female	p-value	Male-Male	Male-Female	p-value
Obs.	134	73 (54%)	61 (46%)		139	92		132	65	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age	34.37	33.08	35.90	0.23	31.55	33.24	0.30	34.63	35.88	0.52
	(13.45)	(13.12)	(13.8)		(11.97)	(12.49)		(12.07)	(13.73)	
Student	0.19	0.23	0.13	0.13	0.31	0.16	0.01	0.11	0.14	0.62
	(0.39)	(0.42)	(0.34)		(0.46)	(0.37)		(0.32)	(0.35)	
Retired	0.05	0.07	0.03	0.36	0.06	0.05	0.92	0.02	0.06	0.08
	(0.22)	(0.25)	(0.18)		(0.23)	(0.23)		(0.12)	(0.24)	
Unemployed	0.09	0.07	0.11	0.35	0.08	0.07	0.69	0.14	0.08	0.18
	(0.29)	(0.25)	(0.32)		(0.27)	(0.25)		(0.35)	(0.27)	
Low-Level Occupation	0.35	0.23	0.49	0.00	0.19	0.27	0.17	0.45	0.45	0.99
	(0.48)	(0.42)	(0.5)		(0.4)	(0.45)		(0.5)	(0.5)	
High-Level Occupation	0.32	0.40	0.23	0.04	0.36	0.45	0.19	0.28	0.28	0.96
	(0.47)	(0.49)	(0.42)		(0.48)	(0.5)		(0.45)	(0.45)	
Perceived Age	2.41	2.32	2.51	0.34	2.15	2.35	0.17	2.35	2.63	0.08
	(1.13)	(1.16)	(1.09)		(1.06)	(1.12)		(1.02)	(1.02)	
Perceived Status	0.36	0.38	0.35	0.47	0.41	0.38	0.36	0.35	0.41	0.10
	(0.24)	(0.23)	(0.26)		(0.23)	(0.22)		(0.27)	(0.24)	
Perceived Attractiveness	3.44	3.92	2.87	0.00	4.14	3.91	0.19	3.01	2.77	0.18
	(1.37)	(1.35)	(1.18)		(1.32)	(1.32)		(1.2)	(1.05)	
Pie	345.84	337.39	355.95	0.57	421.05	366.68	0.33	385.46	445.75	0.33
	(186.23)	(189.45)	(183.35)		(425.52)	(387.49)		(397.67)	(429.45)	
(Pie-Mean) by Stage	-0.80	0.16	-1.96	0.51	-0.83	3.74	0.26	-1.20	-1.08	0.97
	(18.58)	(19.90)	(16.96)		(36.45)	(17.58)		(22.07)	(25.93)	
Stage	1.93	1.90	1.97	0.42	2.14	1.98	0.25	2.05	2.25	0.21
	(0.53)	(0.53)	(0.52)		(1.02)	(0.99)		(1)	(1.02)	
Bargaining Time	103.82	104.34	103.20	0.81	110.96	107.34	0.57	105.05	100.82	0.55
	(27.54)	(31.52)	(22.08)		(47.09)	(47.52)		(48.18)	(44.2)	
Male Question	0.16	0.17	0.14	0.55	0.26	0.03	0.00	0.21	0.05	0.00
	(0.23)	(0.23)	(0.23)		(0.44)	(0.18)		(0.41)	(0.21)	
Neutral Question	0.68	0.70	0.67	0.63	0.68	0.73	0.40	0.68	0.62	0.36
	(0.32)	(0.31)	(0.33		(0.47)	(0.45)		(0.47)	(0.49)	
Female Question	0.16	0.13	0.19	0.22	0.06	0.24	0.00	0.11	0.34	0.00
	(0.24)	(0.23)	(0.24)		(0.25)	(0.43)		(0.31)	(0.48)	
Responders	Overall	Female	Male	p-value	Male-Female	Female-Female	p-value	Male-Male	Female-Male	p-value
Obs.	428	157 (37%)	271 (63%)		65	92		132	139	
Perceived Age	2.80	2.70	2.86	0.20	2.75	2.67	0.65	2.96	2.76	0.17
	(1.17)	(1.15)	(1.18)		(1.14)	(1.16)		(1.19)	(1.18)	
Perceived Status	0.45	0.45	0.44	0.81	0.47	0.44	0.47	0.45	0.43	0.55
	(0.27)	(0.26)	(0.28)		(0.27)	(0.25)		(0.28)	(0.27)	
Perceived Attractiveness	3.17	3.63	2.90	0.00	3.54	3.69	0.47	3.05	2.76	0.02
	(1.19)	(1.26)	(1.07)		(1.5)	(1.07)		(1.07)	(1.05)	

Notes The table shows the mean values and the standard deviations of the main outcome and control variables. *Age* describes the age in years. *Student*, *Retired* , *Unemployed*, *Low-Level Occupation* and *High-Level Occupation* take the value of 1 when the proposer is a student, retired, unemployed and holding a low and high occupation, respectively. *Perceived Age*, *Perceived Status* and *Perceived Attractiveness* are elicited measures of age, status and attractiveness, in a scales between, 1 (below 30) to 6 (above 70), 0 (medium-low) and 1 (medium high), and 0 (very unattractive) to 8 (very attractive), respectively. *Pie* refers to the amount in euro to bargain over. *Stage* refers to the number of stage and can take values between 1 and 4. *(Pie-Mean) by stage* shows the deviation of the pie by stage. *Bargaining Time* summarizes the time left in seconds for the bargaining. Finally, *Male/Neutral/Female question* take the value of 1 when the question is classified as male, neutral and female. The *p*-value are for the F-Test of equality of variable means across gender.

TABLE 2.2– DESCRIPTIVE STATISTICS: OUTCOMES VARIABLES

	Obs.	Overall (1)	Female-Male 139 (32%) (2)	Female-Female 92 (21%) (3)	Male-Female 65 (15%) (4)	Male-Male 132 (31%) (5)	p-value (6)
Bargaining Outcomes:							
Prob. of No Agreement	428	0.12	0.10	0.15	0.06	0.14	0.24
		(0.32)	(0.30)	(0.36)	(0.24)	(0.35)	
Proposer's Outcome	377	367.14	378.29	325.31	403.59	364.02	0.68
		(394.09)	(404.99)	(354.89)	(418.33)	(396.69)	
No. of Rounds	377	3.58	3.72	3.81	3.77	3.16	0.09
(when agreement)		(2.12)	(2.31)	(2.06)	(2.2)	(1.86)	
No. of Rounds	51	2.71	2.57	2.71	2.75	2.79	0.96
(when no agreement)		(1.12)	(0.94)	(1.33)	(1.50)	(1.08)	
Bargaining Behavior when agreement:							
Offers	1283 (376)	34.11	35.62	33.25	29.90	35.63	0.23
		(36.96)	(38.53)	(34.05)	(23.05)	(43.88)	
Prob. Responder Accepts	1283 (376)	0.14	0.14	0.13	0.10	0.17	0.16
		(0.34)	(0.34)	(0.34)	(0.30)	(0.37)	
Demands	871 (321)	105.31	127.64	112.96	60.23	100.99	0.00
		(182.85)	(200.86)	(247.37)	(40.09)	(142.11)	
Prob. Proposer Accepts	871 (321)	0.23	0.22	0.21	0.25	0.25	0.62
		(0.42)	(0.42)	(0.40)	(0.43)	(0.44)	

Notes The table shows the mean values and the standard deviations of the main outcome variables. *Prob. of No Agreement* takes the value of 1 when the bargaining partners do not reach an agreement and 0 otherwise. *Proposer's Outcome* refers to the amount in euro agreed for the proposer and *No. of Rounds* summarizes the duration of the bargaining process. *Offer* and *Demand* refer to the offers and demands in euro by the proposer and responder, respectively, and *Prob. Responder(Proposer) Accepts* take the value of 1 when an offer(demand) is accepted and 0 otherwise. The p-value are for the F-Test of equality of variable means across gender combinations.

proposers who are students are more likely to choose male responders, and male responders who are chosen by male proposers are more attractive than those who are chosen by female proposers.

Table 2.2 shows the descriptive statistics of all the outcome variables we analyze, overall (column 1) and for the gender combinations of the bargaining matches, respectively. The last column in Table 2.2 reports the p-values for the F-Test of equality of variable means across all four gender matchings. We distinguish between variables that describe bargaining outcomes and bargaining behavior.

The main bargaining outcome variables of interest are *Prob. of No Agreement*, *Proposer's Outcome* and *No. of Rounds*.[12] *Prob. of No Agreement* is a dummy variable that takes the value of 1 if the proposer and the responder do not reach an agreement and 0 otherwise.[13] Overall, only 12% of the negotiations failed (51 out of 428), while 88% of the time proposers and responders reached an agreement (377 out of 428). For the rest of the outcome variables, we restrict the sample to successful bargaining matchings (377 matchings). Proposers on average earn 367 Euro. Given that the size of the

[12] We have also considered an alternative measure for number of rounds, such as time elapsed since bargaining started. Results remain unchanged.

[13] As explained in Section 2.2.1, there are 2 possible cases in which bargaining partners do not reach an agreement. Either the proposer drops the negotiation to look for another possible responder, or the three-minute limit is reached while negotiating. The former is the most common case (96% of breakdowns), while the latter is very rare (4% of breakdowns).

pie to be shared in successful negotiations averages 417 Euro, proposers take 88% of it.[14] On average, successful negotiations take longer (about 4 rounds) than unsuccessful ones (about 3 rounds). The final column shows the *p*-value for the comparison between the four cases of different gender pairings. The matchings that prove most beneficial for the proposer are those between a male proposer and a female responder (404 Euro), while the lowest outcome for the proposer is in bargaining between two women (325 Euro). These differences, however, are not significant. Interestingly, only the number of rounds when the negotiation did not fail shows significant differences, with negotiations between male contestants being the fastest. Notice however that a priori important variables can differ significantly from one matching to another, e.g. the size of the pie. Regression analysis shows that controls are important to identify gender differences and gender interaction effects.

When analyzing bargaining behavior in successful negotiations, taking into account the panel structure of the database, we look at the offers made by the proposers (*Offers*), the demands made by the responders (*Demands*), and their respective probabilities of accepting (*Prob. Responder Accepts*, *Prob. Proposer Accepts*). Note that these are round by round data, so there are several observations per bargaining matching, as long as the negotiation took more than one round. Moreover, there are matchings that have no demands or offers, so one bargaining role remained silent. Therefore, the column for the number of observations includes both the round by round observations and, in parenthesis, the number of bargaining matchings. The average offer by proposers is 34 Euro and the average demand is 105 Euro, with the probabilities of accepting being 14% by responders and 23% by proposers. Demands among different matchings show significant differences, demands from female responders to male proposers being the lowest, of about 60 Euro.

2.2.3 Research Questions and Identification Strategy

The database from the TV show enables us to answer two interesting research questions. First, we test whether male and female proposers differ in their choice of the gender of their bargaining partners. Second, we test for gender differences and gender interaction effects in bargaining. In other words, we do not only show the effect of the gender of the proposer and the responder on bargaining, but we also compare the four different gender combinations (male-female, female-male, male-male and female-female) to test for gender interaction effects in bargaining.[15] Here, we distinguish between bargaining outcomes and bargaining behavior: bargaining outcomes include whether the bargaining fails to reach an agreement (*Prob. of No Agreement*), the bargain-

[14] Shares of the pie and proportional offers and demands can be also analyzed. However, we decided to focus on absolute values (*Proposer's Outcome*, *Offers* and *Demands*) instead of relative ones for two main reasons: First, the pie varies significantly from matching to matching. It ranges between 100 Euro and 1600 Euro. Second, responders do not know the size of the pie, so they could not possibly care or negotiate in terms of shares of the pie.

[15] In all regressions the omitted category is female-female.

ing outcome for the proposer (*Proposer's Outcome*), and the duration of the bargaining process (*No. of Rounds*). Bargaining behavior includes offers made by proposers (*Offers*), demands made by responders (*Demands*), and their respective probabilities of accepting (*Prob. Proposer/Responder Accepts*).

In particular, to test whether female and male proposers behave differently when choosing the gender of their bargaining partners, we estimate the following regression:

$$MaleResponder_j = \alpha + \beta MaleProposer_i + \gamma X_i + \epsilon_i \tag{2.1}$$

To test whether gender and gender interactions matter in explaining bargaining outcomes and behavior, we estimate the following regressions for gender differences and gender interactions, respectively:

$$Y_{ij} = \alpha + \beta_1 MaleProposer_i + \beta_2 MaleResponder_j + \gamma X_{ij} + \epsilon_{ij} \tag{2.2}$$

$$Y_{ij} = \alpha + \beta_1 Male_i Female_j + \beta_2 Female_i Male_j + \beta_3 Male_i Male_j + \gamma X_{ij} + \epsilon_{ij} \tag{2.3}$$

In studying the choice of the gender of the bargaining partner and bargaining outcomes we use the collapsed data at the responder level, given that these variables, and indeed the independent variables, remain constant round by round. We therefore have 428 matchings and observations for the *Male Responder* and *Prob. of No Agreement* dependent variables. For *Proposer's Outcome* and *No. of Rounds* we constrain the sample to the matchings that reached an agreement. This results in 377 matchings and observations. Given that the same proposer is matched with different responders, we always cluster the standard errors at the proposer level.

When analyzing bargaining behavior in successful negotiations, we exploit the panel structure of the database, i.e. we use the round by round bargaining data but specify the identification of the responder as the panel variable, and estimate a random effects model. We also cluster the standard errors at the proposer level.[16]

Two types of control are applied in all regressions: First there are control variables that refer to the proposers' and responders' socio-demographic characteristics. Second there are controls specific to the bargaining matching, such as the stage, variation in the pie within the stage and bargaining time. Finally, we also include controls specific to each dependent variable, such as controlling for previous offers when explaining the probability of the responder accepting.

[16] When analyzing bargaining behavior with 0-1 outcome variables and exploiting the panel structure of the data, such as *Prob. Proposer Accepts*, *Prob. Responder Accepts*, we estimate a random effects probit model. In these cases, we use bootstrapped standard errors.

2.3 Results

2.3.1 Choice of Bargaining Partner

Proposers choose responders to bargain with. This choice is limited in two important ways. First, proposers choose responders from among the people on the streets at the time of the TV show.[17] Second, and more importantly, they make their choice under time pressure, as they have only three minutes to find a responder and negotiate a prize after they are given the question. Given these constraints, the choice of a responder seems more intuitive than meditated. It is, nevertheless, interesting to analyze the determinants of the gender of the responder chosen and in particular, whether male and female participants have different preferences for a particular gender of responder.

As seen above, although the sample of proposers is balanced in terms of gender, the sample of responders is not. In particular, there are significantly more male than female responders, which suggests that both male and female proposers show a preference for bargaining with men rather than with women. The over-representation of men in the responder sample is also consistent with women being more reluctant to participate in the show than men. However, given the actual gender composition of the samples of proposers and responders, the hypothesis that the existing gender combinations of the matchings are as expected under random matching cannot be rejected.[18]

We therefore study the determinants of the gender of the responder. In particular, we test whether the gender of the proposer has any effect on the gender of the responder chosen. In other words, we test whether male and female proposers have different preferences as regards the gender of responders. The results of this estimation are shown in Table 2.3. The first column shows the results with no controls, while the second column shows the results when controls are added.[19]

Once controls are added, evidence emerges that male participants have a stronger

[17] Note that the availability of potential responders, and whether the available responder population shows a balanced gender composition, could be important factors. However, as the show took place in the main streets of large cities in Spain, we assume that the availability of people is not determinant and that the gender composition is balanced. Hence, these factors are not a concern in analyzing the choice of the gender of the bargaining partner.

[18] With 73 female and 61 male proposers, and 157 female and 271 male responders, under random matching one would expect 147 female vs. male, 86 female vs. female, 74 male vs. female, and 125 male vs. male bargaining matchings. In the sample, we end up with 139 female vs. male, 92 female vs. female, 65 male vs. female, and 132 male vs. male actual bargaining matchings. The Chi-Square Goodness of Fit test cannot reject the null hypothesis that the gender matchings are as expected under random matching (p-value 0.34).

[19] Estimations with alternative specifications are shown in Table B.1, in Appendix B.1. We also consider linear probability and logit estimation models (in columns 1 and 2, respectively). Alternatively, we also treat data as a panel, as we observe the same proposer matched with different responders, and estimate a random effects probit model (column 3). These three specifications yield results that are similar both qualitatively and quantitatively.

TABLE 2.3– CHOICE OF THE SEX OF THE RESPONDER

	Prob. Male Responder (1)	Prob. Male Responder (2)
Male Proposer	0.0683	0.136***
	(0.0543)	(0.0527)
Age Proposer		0.00222
		(0.00333)
Student Proposer		0.108
		(0.0746)
Retired Proposer		-0.124
		(0.189)
Unemployed Proposer		0.138*
		(0.0742)
Low-Occupation Proposer		0.00293
		(0.0681)
Perc. Attractiveness Proposer		0.0534**
		(0.0261)
Perc. Status Proposer		-0.151
		(0.124)
(Pie-Mean) by Stage		-0.000261
		(0.000789)
Remaining Time		0.000529
		(0.000471)
Male Question		0.302***
		(0.0520)
Female Question		-0.285***
		(0.0626)
Stage FE	YES	YES
Observations	428	428

Notes The dependent variable takes the value 1 if the selected responder is male and 0 otherwise. The table shows the marginal effect values of the coefficients using the probit model. The control variables are described in the notes of Table 2.1. Clustered standard errors at the proposer level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

preference for male bargaining partners. Men are more likely to choose a male bargaining partner than women.[20] The stronger preference of male proposers for male responders may be explained by gender differences in preferences, such as taste-based discrimination (Becker (1971)) or, taking a more rational approach, may potentially be explained by gender differences in beliefs. First, men and women might have different beliefs about the likelihood of men and women knowing the correct answer. In particular, men might assign a higher probability than women to the notion of men knowing the correct answer. In the sample, we find no evidence that male and female responders are more/less likely to know the correct answer.[21] Second, men and women might

[20]Note that this cannot be explained by any bias in the gender composition in the sample of available potential responders, as any limitation by this type should affect both male and female proposers equally.

[21]Table B.2, in Appendix B.1, shows that when we do not control for the male/neutral/female nature of the question, men are less likely to know the correct answer (significant at 10%), shown in column 1. However, once that feature is controlled for, men and women are equally likely to know the correct answer, shown in column 2. Furthermore, responders are less likely to get the correct answers to male questions.

have different beliefs about the bargaining behavior of male and female responders. In particular, men may believe more strongly than women that male responders will be less aggressive in their bargaining behavior. As shown below, the bargaining outcomes coming from matchings between different genders indicate that this belief is not correct. Third, men and women might have also different beliefs about gender differences in the willingness to participate in the TV show as responders.

As expected, the most important determinant of the gender of the opponent is whether the question is perceived to be male or female (neutral being the omitted category). When presented with a male question (e.g. a sports related question), proposers look for men, while when presented with a female question (e.g. questions related to fashion or celebrities), proposers look for female responders. Furthermore, we have tested whether the male preference for male responders is independent of the perception (male/female/neutral) of the question, which is supported by the data.[22] Also, attractive proposers are more likely to choose male responders. Further analysis shows that this is not different for male and female proposers.[23]

From this analysis it is clear that responders are chosen by the proposers, and that male and female proposers show differences in the gender of their chosen bargaining partners. This yields an unbalanced sample of different gender combinations (male-male, male-female, female-male and female-female matchings), and a biased sample due to selection (male proposers are more likely to choose male responders). Section 4 addresses these two problems by carrying out additional regression analysis. First, we carry out a regression with probability weights that correct the unbalanced sample. Second, we carry out regressions in a matched sample using probability score matching to address the selection problem. Third, we reject the underlying hypothesis that male and female responders chosen by male and female proposers differ in their observable characteristics. We show that the main results are sound and robust.

2.3.2 Outcome Variables: Probability of No Agreement, Proposer's Outcome and Number of Rounds

Bargaining outcomes can be described by three main variables: First, whether the bargaining partners reach an agreement or not, summarized by *Prob. of No Agreement*; Second, by the amount of money agreed for the proposer, described by *Proposer's Outcome*[24]; and third, by the duration of the bargaining process, that is, the number of rounds (*No. of Rounds*).

[22] When interacting *Male Proposer* with the *Male/Female Question*, the interactions are insignificant. These results are shown in column 4 of Table B.1, in Appendix B.1.

[23] When interacting *Male Proposer* with the *Proposer's Attractiveness*, the interaction is insignificant. These estimation results are shown in column 5 of Table B.1, in Appendix B.1.

[24] Notice that analyzing the amount of money agreed for the proposer yields the same conclusions than analyzing the amount of money agreed for the responder.

Table 2.4 summarizes the regression results for these three outcome variables.[25] The first three columns look at the effect of the gender of the proposer and of the responder, while the last three columns also include the interaction between the gender of the bargaining partners. When estimating gender interaction effects we also include the *p*-values for the hypothesis test that conducts a pairwise comparison of the effects of different gender combinations, shown at the bottom of Table 2.4. We find neither gender difference nor gender interaction effects on the probability of no agreement and on the duration of bargaining. In contrast, there are important differences in terms of earnings. Although there is no evidence for any gender difference in proposers and responders' behavior (column 2), we find that negotiations between male proposers and female responders result in about 11 Euro more for the proposer (column 5). Furthermore, as shown by the hypothesis tests, bargaining between men and women stands out as the most beneficial for proposers, compared to negotiations with any other gender combination. This means that while male proposers earn 3% more when negotiating with women than with men, female responders earn 22% less when negotiating with men than with women. This result shows that gender interactions are crucial: it is not just that men and women behave differently when bargaining but, more importantly, differences depend on the gender of the bargaining partner. Later analysis of gender differences and gender interaction effects in the bargaining process clarify whether this is due to male proposers discriminating against female responders, to female responders behaving differently when interacting with male proposers, or to a combination of both.

Many controls are significant in explaining the bargaining outcomes. The first offer made by the proposer has been found to be an important determinant for bargaining outcomes, i.e. the quantity offered in the first round (see for example Van Poucke and Buelens (2002)).[26] We confirm that this is indeed an important determinant: the higher the first offer, the more likely it is that a successful agreement will be reached, the lower the proposer's outcome is and the shorter the negotiation is. Moreover in terms of the probability of there being no agreement, the longer the bargaining goes on the lower the probability of failure is. Also, those negotiations that have more time left are more likely to fail, given that the proposers have still time to find alternative responders. Interestingly, older proposers, those who hold low-level occupations, and students,

[25]For the variable *Prob. of No Agreement*, we show the estimation results using the probit model while for *No. of Rounds* we use a Poisson regression. We also consider alternative specifications, shown in Table B.3, in Appendix B.1. For *Prob. of No Agreement* we consider linear probability and logit estimation models (shown in columns 1 and 2, respectively). For *No. of Rounds*, we also consider OLS, shown in column 5. For the three outcome variables, we also consider the data as a panel, as we observe the same proposer matched with different responders, and estimate a random effects model, shown in columns 3, 4 and 6, for *Prob. of No Agreement*, *Proposer's Outcome*, and *No. of Rounds*, respectively. All these alternative specifications yield results that are both qualitatively and quantitatively the same.

[26]The empirical literature on bargaining highlights 3 different internal reference points that affect bargaining outcomes: reservation prices, aspiration prices and opening offers. It has been found that they are all positively, strongly correlated, so opening offers -the only observable internal reference point in our database- may contain the effect of the other two internal reference points.

TABLE 2.4– GENDER DIFFERENCES AND GENDER INTERACTION EFFECTS IN BARGAINING OUTCOMES

	Prob. No Agreement (1)	Proposer's Outcome (2)	No. of Rounds (3)
Male Proposer	0.0218	1.899	-0.313
	(0.0285)	(3.641)	(0.230)
Male Responder	0.00286	-3.873	-0.192
	(0.0315)	(3.111)	(0.248)
Age Proposer	-0.00764***	0.170	0.00990
	(0.00252)	(0.236)	(0.0130)
Student Proposer	-0.0641*	6.204	0.520
	(0.0367)	(4.888)	(0.351)
Retired Proposer		-3.331	-0.754
		(9.903)	(0.543)
Unemployed Proposer	0.0567	-5.383	-0.984***
	(0.0833)	(11.28)	(0.325)
Low-Occup. Proposer	-0.0664**	-2.643	-0.451*
	(0.0315)	(4.474)	(0.262)
Perc. Status Proposer	0.145*	7.265	0.886*
	(0.0831)	(6.479)	(0.530)
Perc. Attractiveness Proposer	-0.00758	0.442	-0.153
	(0.0131)	(1.275)	(0.103)
Perc. Age Responder	-0.00309	4.311**	-0.0310
	(0.0158)	(1.848)	(0.110)
Perc. Status Responder	-0.0966	-13.91**	0.208
	(0.0592)	(6.752)	(0.413)
Perc. Attractiveness Responder	0.00468	3.255	0.0933
	(0.0173)	(2.011)	(0.113)
(Pie-Mean) by Stage	0.00122	1.206***	-0.00167
	(0.000788)	(0.0998)	(0.00323)
Remaining Time	0.000663**	0.0829*	0.00601**
	(0.000280)	(0.0478)	(0.00242)
No. Of Rounds	-0.0307***	-7.214***	
	(0.00786)	(1.028)	
First Round Offer	-0.00204*	-1.116***	-0.0485***
	(0.00120)	(0.214)	(0.00893)
Proposer Starts		19.32***	0.925**
		(6.723)	(0.404)
Proposer Accepts		-11.11***	
		(3.330)	
Constant		62.36***	
		(17.77)	
Stage FE	YES	YES	YES
Observations	428	377	377

	Prob. No Agreement (4)	Proposer's Outcome (5)	No. of Rounds (6)
MF	-0.0447	10.78***	-0.177
	(0.0397)	(3.822)	(0.337)
FM	-0.0397	2.995	-0.0868
	(0.0384)	(3.689)	(0.332)
MM	0.0160	-0.523	-0.476
	(0.0389)	(5.037)	(0.305)
Age Proposer	-0.00767***	0.199	0.0102
	(0.00253)	(0.239)	(0.0130)
Student Proposer	-0.0553	5.251	0.500
	(0.0377)	(4.636)	(0.356)
Retired Proposer		-5.042	-0.981***
		(9.842)	(0.325)
Unemployed Proposer	0.0618	-5.153	-0.447*
	(0.0838)	(11.34)	(0.262)
Low-Occup. Proposer	-0.0621*	-2.545	-0.771
	(0.0320)	(4.456)	(0.543)
Perc. Status Proposer	0.160*	4.834	0.849
	(0.0820)	(6.654)	(0.538)
Perc. Attractiveness Proposer	-0.00930	0.708	-0.148
	(0.0130)	(1.278)	(0.104)
Perc. Age Responder	-0.00596	4.729**	-0.0241
	(0.0156)	(1.888)	(0.112)
Perc. Status Responder	-0.0924	-15.14**	0.189
	(0.0588)	(6.607)	(0.414)
Perc. Attractiveness Responder	0.00163	3.981*	0.105
	(0.0176)	(2.046)	(0.113)
(Pie-Mean) by Stage	0.00110	1.212***	-0.00157
	(0.000791)	(0.0999)	(0.00327)
Remaining Time	0.000653**	0.0835*	0.00602**
	(0.000274)	(0.0470)	(0.00242)
No. Of Rounds	-0.0298***	-7.263***	
	(0.00765)	(1.039)	
First Round Offer	-0.00211*	-1.108***	-0.0483***
	(0.00120)	(0.211)	(0.00883)
Proposer Starts		19.12***	0.920**
		(6.930)	(0.404)
Proposer Accepts		-11.41***	
		(3.324)	
Constant		54.92***	
		(19.21)	
Stage FE	YES	YES	YES
Observations	428	377	377
H_0: MF=FM	0.91	0.05	0.79
H_0: MF=MM	0.18	0.02	0.33
H_0: FM=MM	0.13	0.50	0.17

Notes The dependent variables refer to: *the Prob.of No Agreement*, which takes the value of 1 when the bargaining partners do not reach an agreement and 0 otherwise (column 1 and 3); *Proposer's Outcome*, which summarizes the outcome in euro obtained by the proposer from the bargaining (columns 2 and 4); and *No. of Rounds* describes the duration of the bargaining process (columns 3 and 6). *First Round Offer* summarizes the offer in euro made in the very first round. *Proposer Starts* is a dummy variable that takes the value of 1 when the bargaining starts with the proposer making an offer, and 0 otherwise. *Proposer Accepts* is a dummy variable that takes the value of 1 when the bargaining ends with the proposer accepting responder's demand, and 0 otherwise. Columns 1 and 4 show the marginal effect values of the coefficients using the probit model. Columns 2 and 5 show the coefficients for OLS and columns 3 and 6 show the marginal effect values of the coefficients using the Poisson regression model. At the bottom, *p*-values for the hypothesis testing are shown where pairwise comparisons are made for the different gender combinations, where FF refers to bargaining between women, MF refers to the bargaining between a male proposer and a female responder, FM refers to the bargaining between a female proposer and a male responder and MM refers to the bargaining between men. Clustered standard errors at the proposer level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

are less likely to break up the bargaining. As expected, we find that the bigger the pie is –the higher the stage and the greater the deviations from the mean pie at each stage–, the bigger the outcome is for the proposer as the responders are unaware of the size of the pie. Interestingly, the coefficient of the deviation from the mean pie at each stage is close to 1, which suggests that the increases in the pie are absorbed by the proposer, who plays the strong role in bargaining given the information asymmetry regarding the size of the pie. Also, as expected, the more rounds there are the lower the outcome is for the proposer. When looking at the duration of the negotiation (*No. of Rounds*) it is found, as expected, that the longer bargaining goes on the more rounds there are; and the bigger the pie is, measured in terms of the different stages, the longer bargaining goes on. Finally, when proposers are unemployed or hold a low-level occupation the process is shorter.[27]

Two control variables are of special interest. First, *Proposer Starts* describes how the negotiation starts, taking the value of 1 when it is the proposer who starts. Second, *Proposer Accepts* describes how the negotiation ends by taking the value of 0 when it is the responder who accepts the proposer's offer and 1 when it is the proposer who ends up accepting a responder's demand. Estimated coefficients show that initiating the negotiation pays off, while being the party who accepts the other's demand/offer does not. Further analysis, where we split the bargaining outcomes into those that end with the proposer accepting a responder's demand (202 cases out of 377), and those that end with the responder accepting a proposer's offer (175 cases out of 377) show that the difference found in the male-female interaction stems from those deals that end with proposers accepting responders' demands. This suggests that the bargaining outcome between male proposers and female responders is the most beneficial for proposers must be driven by females demanding less from male proposers rather than by male proposers offering less to female responders.

Three final remarks are noteworthy. First, the results shown in Table 2.4 are not driven by extremely high and low bargaining outcomes for the proposer. We have replicated the regressions on proposer's outcome deleting the 5% of highest and lowest outcomes, and the estimation results remain unchanged (results available upon request). Second, we have also considered other controls. In particular, we have controlled for whether the question is male or female, as one might consider situations in which a particular perception affects participants' bargaining power. We find that they are never significant, and more importantly, the results on the gender interactions of Table 2.4 remain unchanged (results available upon request). Third, given that gender interaction effects and the hypothesis testing shown at the end of the tables also inform about gender differences, from now on we only show the gender interaction effect regressions.

[27] The control *Retired* cannot be estimated in columns 1 an 4 because it turns out *Retired* predicts perfectly whether the negotiations ended successfully or not.

2.3.3 BARGAINING BEHAVIOR: OFFERS, DEMANDS, AND PROBABILITIES OF ACCEPTING

We now analyze bargaining behavior regarding offers, demands and the likelihood of accepting them.[28] In an alternating-offer bargaining round the proposer starts with an offer, which can be accepted or rejected by the responder. If the responder rejects the offer, he/she can respond with a demand, which may then be accepted or rejected by the proposer. This type of bargaining round can be repeated until an agreement is reached. Hence, the variables of interest in this section are offers, demands, and their likelihood of being accepted by responders and proposers. We analyze offers, and probabilities of responders accepting offers in two separate regressions, one for the first round and the other for subsequent rounds. Opening offers are exogenous and cannot be influenced by interaction with the responder, as they come first. Furthermore, they are important determinants of subsequent behavior in bargaining. We therefore decided to show the regressions for the opening offer and the remaining offers separately. For the rest of the variables, we test whether behavior in the first and subsequent rounds is indeed different. This analysis leads us to show regression analysis for the likelihood of responders accepting separately for initial and subsequent offers, but not for demands and the likelihood of proposers accepting demands.

Table 2.5 shows the results for offers (columns 1 and 2), for the likelihood of responders accepting them (columns 3 and 4), for demands (column 5) and for the likelihood of proposers accepting them (column 6).[29]

We start by looking at the offers made by proposers and the likelihood of responders accepting them. Opening offers (column 1) do not show any significant gender-related effect. This is in sharp contrast with the findings of other authors, e.g. by Dittrich et al. (2014), who report that offers from men to women are lower. Moreover, the signs suggest, consistently with previous literature, that women are more likely to

[28]We have also analyzed gender differences and gender interaction effects in the use of different bargaining strategies, available in Table B.4 in Appendix B.1. We classify four types of bargaining strategy: First, one can actively make offers or demands or remain passively silent and wait for the other person to do so. Active bargaining strategies can then be classified into increasing, decreasing or maintaining offers/demands from round to round. We find no gender differences or gender interaction effects in remaining silent when bargaining in either role. Interestingly, we find that men are more likely to increase their offers from round to round, while women are more likely to stick to an offer. We find neither gender differences nor gender interaction effects in the use of bargaining strategies among responders. Nevertheless, these findings are not enlightening the main result found in male-female matching being the most beneficial for proposers and the least for the responders.

[29]Table B.5 in Appendix B.1 shows the estimation results for alternative specifications. For *Offers* in subsequent rounds, column 1 shows the OLS estimation results using collapsed data (RE model yields the same estimates). Columns 2 and 3 show the probit and RE probit, respectively for *Prob. Responder Accepts*. For *Demands*, Column 4 shows the OLS estimation results using the collapsed data (RE model yields the same estimates). Finally, for *Prob. Proposer Accepts*, columns 5 and 6 show the probit and RE probit model estimation results, respectively. As shown by the estimation results, the magnitudes are slightly lower, due to collapsing the data, and the weak result on offers loses significance. More importantly, the result on demands is robust and sound.

TABLE 2.5– GENDER INTERACTION EFFECTS IN OFFERS, DEMANDS AND PROBABILITIES OF ACCEPTANCE

	Opening Offers (round=1)	Offers (round>1)	Prob. Responder Accepts (round=1)	Prob. Responder Accepts (round>1)	Demands	Prob. Proposer Accepts
	(1)	(2)	(3)	(4)	(5)	(6)
MF	-0.874	-4.706*	0.0968	-0.0515	-62.92***	0.00631
	(2.872)	(2.768)	(0.0656)	(0.0545)	(22.77)	(0.08892)
FM	-0.214	-3.051	0.0794	0.00532	2.403	0.0189
	(1.666)	(2.435)	(0.0573)	(0.0471)	(17.79)	(0.0731)
MM	0.433	-0.157	0.0742	0.0617	-9.280	0.0629
	(2.489)	(3.748)	(0.0517)	(0.0492)	(18.88)	(0.0911)
(Pie-Mean) by Stage	-0.218***	-0.0230	0.00164*	0.000450	0.0800	-0.000563
	(0.0496)	(0.0523)	(0.000841)	(0.000450)	(0.250)	(0.00124)
Bargaining Time	-0.00484	-0.0978***	-0.000907***	-0.000390	-0.201	-0.000446
	(0.0160)	(0.0351)	(0.000280)	(0.000347)	(0.195)	(0.000638)
No. of Rounds	-1.637***	-1.094**			29.39***	
	(0.303)	(0.492)			(6.832)	
Round		8.425***		0.0235	-7.185	0.0851*
		(1.407)		(0.0160)	(15.85)	(0.0442)
Demand at Current Round						-0.00333**
						(0.00188)
Previous Demand		0.0834***				
		(0.0281)				
(Previous Demand)^2		-4.57e-05**				
		(1.95e-05)				
First Round Offer		1.277***				0.00928**
		(0.223)				(0.00403)
Offer at Current Round			0.00381***	0.000215	0.977***	
			(0.000959)	(0.000679)	(0.234)	
Proposer Starts		-18.22***		-0.121**	-13.62	-0.0573
		(5.396)		(0.0679)	(21.61)	(0.138)
No Previous Demand					57.94**	-0.0414
					(24.12)	(0.0549)
Constant	11.40*	29.03**			13.31	
	(6.648)	(14.58)			(53.48)	
Controls For						
Stage FE	YES	YES	YES	YES	YES	YES
Socio-Demographics of Proposer	YES	YES	NO	NO	NO	YES
Perceptions about Proposer	YES	YES	YES	YES	YES	YES
Perceptions about Responder	YES	YES	YES	YES	YES	YES
Observations	357	926	357	926	871	871
Number of Responders	357	316	357	316	321	321
H_0: MF=FM	0.79	0.59	0.74	0.29	0.00	0.89
H_0: MF=MM	0.47	0.24	0.61	0.02	0.01	0.53
H_0: FM=MM	0.74	0.50	0.90	0.16	0.50	0.60

Notes The dependent variable *Offers* refer to the offers in euro made by the proposer (columns 1 and 2); *Prob. Responder Accepts* takes the value of 1 when the responder accepts the offer made by the proposer and 0 otherwise (columns 3 and 4); *Demands* refer to the demands in euro made by the responder (column 5); and*Prob. Proposer Accepts* takes the value of 1 when the proposer accepts the demand made by the responder and 0 otherwise (column 6). Columns 3, 4, and 6 show the marginal effects of the coefficients using the probit model. Except for round 1 regressions (column 1 and 3), we use random effects model. At the bottom, *p*-values for the hypothesis testing are shown where pairwise comparisons are made for the different gender combinations, where FF refers to bargaining between women, MF refers to the bargaining between a male proposer and a female responder, FM refers to the bargaining between a female proposer and a male responder and MM refers to the bargaining between men. Clustered standard errors at the proposer level (columns 1, 2, 3 and 5) and bootstrapped standard errors (columns 4 and 6), in parentheses. *** p<0.01, ** p<0.05, * p<0.1

accept an offer in the first round when it comes from a male proposer (column 3), but the differences are not statistically significant. Regarding offers and likelihood of responders accepting them in subsequent rounds (columns 2 and 4), once initial offers and past demands are controlled for, show no evidence of important gender differences nor gender interaction effects. Two minor differences can be mentioned. First, regarding offers, men offer to female responders about 5 Euro less than female proposers do, although it is significant at the 10%. Second, regarding responders' likelihood of accepting the offers, offers made by male proposers are less likely to be accepted by female responders.

We now turn our attention to demands and the likelihood of proposers accepting a given demand. We find one important gender interaction effect. Consistent with our findings when looking at raw mean values in demands (Table 2.2), the matching that shows a difference in behavior is again that of a male proposer and a female responder (column 5 in Table 2.5). Women's demands differ depending on whether they are interacting with a male or a female proposer. When conveying a demand to a male proposer, women demand about 63 Euro less. We find no differential behavior when looking at the likelihood of accepting demands (column 6).

All controls go in the expected directions. First, negotiations in which there is more time left have lower offers and are less likely to be accepted initially. Also, on the one hand negotiations that last longer have lower offers and higher demands, but on the other hand the more advanced a negotiation is the higher the offers are.[30] Third, as expected, the higher (lower) the offers (demands) are, the more likely it is that they will be accepted by the responder (proposer). Moreover, the higher the opening offer by the proposer the more likely he/she is to accept responders' demand. Also, the opening offer is an important determinant for subsequent offers, and previous offers positively affect subsequent demands. Fourth, past demands also prove important in explaining offers.[31] Furthermore some non-linearities appear, as the squared term is also significant.[32] Finally, and interestingly, the size of the pie to be divided up, measured through the stage variable, is positively related to offers and demands. This finding is not totally intuitive. Note that responders do not know how big the pie is, so proposers could pretend to be at the first stage in all negotiations, and offers should

[30]For the analysis of *Offers* and *Demands* we include both the total number of rounds and round variables as controls. Note that *No. of Rounds* measures the difference in rounds between different bargaining processes, while the *Round* variable measures the variation within the same process. For offers, the determinant variable is *Round*, while for demands it is *No. of Rounds*. For the likelihood of accepting we cannot have them both so we include only Round.

[31]Note that for responders' demands and the likelihood of proposers accepting, the control *Offers* includes all offers round by round, including those made in round 1, which affect demands and the likelihood of proposers accepting in round 1 and subsequent ones, which affect responders' and proposer's behavior in later rounds.

[32]Although past demands affect current offers, the existence of non-linearities makes the relation between demands and offers ambiguous. Thus we cannot ensure whether the fact that female responders demand less from male proposers aggravates the situation of these female responders receiving even lower offers.

not therefore depend on the size of the pie. However, proposers do adjust their offers to the size of the pie, and consequently responders do adjust their demands to the proposers' offers. Finally, regarding socio-demographics, responder' perceived status seem to be positively correlated with demands.

Four comments must be made. First, the main effect is observed when male contestants act as proposers, which is the strong role, and female participants act in the role of responders, which is the weak role. Second, this effect is sizable. As the average demand in a female-female matching is 113 Euro, female responders demand about 55% less from male proposers. Third, differential behavior depending on gender is not initiated by the proposer's opening offers but by the responder's demands. We can therefore directly relate this result with the difference found in earnings obtained from bargaining. Male proposers bargaining with female responders being the most beneficial for proposers is explained by female responders demanding less from male proposers. Finally, further analysis on subsequent offers suggests that when there is a previous demand, the weak effect we found in column 2 becomes slightly stronger, while when there is no previous demand, that is, when they reject an offer but post no demand (remain silent), we find no evidence of any gender difference.[33] This further confirms that the difference in bargaining outcomes is driven by differences in demands.

2.4 Robustness Tests: Unbalancedness and Selection

Contrary to what would occur in a perfectly randomized setting, in our setting proposers and responders are not randomly matched: proposers choose their bargaining partners. Section 2.3.1 above shows that although both male and female proposers are more likely to choose male bargaining partners, male proposers show a stronger preference for male bargaining partners. This results in an unbalanced sample of different gender matchings and in a selection problem. In this section we perform three robustness tests that address these issues, focusing on the three main bargaining outcome variables (*Prob. of No Agreement*, *Proposer's Outcome*, *No. of Rounds*). Columns 1, 4, and 7 in Table 2.6 replicate columns 4-6 in Table 2.4.

First, as can be seen clearly in Tables 2.1 and 2.2, the different gender combinations in the matched partners are represented in an unbalanced way. In particular, female-male and male-male matchings are over-represented while male-female and female-female matchings are under-represented. We carry out regressions that weight each observation within each gender combination by the inverse of its probability in the sample. The idea behind this analysis is to weight each observation within each matching

[33] Notice that opening offers and opening demands play quite different roles in our database. First, very few interactions start with a demand rather than an offer (20 out of 377), so opening offers are the real starting point of bargaining, while demands come after the first offer has been rejected.

TABLE 2.6– ROBUSTNESS TESTS: GENDER INTERACTION EFFECTS IN BARGAINING OUTCOMES

	Prob. No Agreement	Weights Prob. No Agreement	Matched Sample Prob. No Agreement	Proposer's Outcome	Weights Proposer's Outcome	Matched Sample Proposer's Outcome	Number of Rounds	Weights Number of Rounds	Matched Sample Number of Rounds
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MF	-0.0447	-0.0475	-0.0485	10.78***	10.41***	8.988**	-0.177	-0.236	-0.237
	(0.0397)	(0.0365)	(0.0382)	(3.822)	(3.591)	(3.600)	(0.337)	(0.338)	(0.350)
FM	-0.0397	-0.0328	-0.0403	2.995	3.217	0.524	-0.0868	-0.0565	0.00804
	(0.0384)	(0.0367)	(0.0385)	(3.689)	(3.637)	(3.990)	(0.332)	(0.339)	(0.367)
MM	0.0160	0.0158	0.0111	-0.523	-0.208	-0.479	-0.476	-0.522*	-0.694**
	(0.0389)	(0.0380)	(0.0396)	(5.037)	(4.886)	(5.260)	(0.305)	(0.304)	(0.343)
Age Proposer	-0.00767***	-0.00804***	-0.00816***	0.199	0.111	0.178	0.0102	0.00705	0.00762
	(0.00253)	(0.00240)	(0.00264)	(0.239)	(0.203)	(0.190)	(0.0130)	(0.0144)	(0.0154)
Student Proposer	-0.0553	-0.0704**	-0.0867**	5.251	4.728	4.404	0.500	0.508	0.458
	(0.0377)	(0.0343)	(0.0354)	(4.636)	(4.266)	(5.731)	(0.356)	(0.362)	(0.419)
Retired Proposer				-5.042	-1.642	-5.728	-0.771	-0.697	-0.828
				(9.842)	(8.247)	(7.944)	(0.543)	(0.598)	(0.598)
Unemployed Proposer	0.0618	0.0691	0.0775	-5.153	-4.932	-0.593	-0.981***	-0.943***	-1.082***
	(0.0838)	(0.0869)	(0.0966)	(11.34)	(10.04)	(6.329)	(0.325)	(0.352)	(0.379)
Low-Occup. Proposer	-0.0621*	-0.0540*	-0.0628*	-2.545	-2.438	-4.729	-0.447*	-0.412	-0.553*
	(0.0320)	(0.0318)	(0.0375)	(4.456)	(3.878)	(5.172)	(0.262)	(0.279)	(0.289)
Perc. Status Proposer	0.160*	0.131	0.178*	4.834	5.705	2.655	0.849	0.804	0.978
	(0.0820)	(0.0805)	(0.0965)	(6.654)	(6.109)	(8.482)	(0.538)	(0.552)	(0.622)
Perc. Attractiveness Prop.	-0.00930	-0.00415	-0.00836	0.708	0.209	-1.284	-0.148	-0.210**	-0.211*
	(0.0130)	(0.0122)	(0.0141)	(1.278)	(1.130)	(1.147)	(0.104)	(0.104)	(0.120)
Perc. Age Responder	-0.00596	-0.00396	-0.0285*	4.729**	4.524***	4.701**	-0.0241	-0.0183	-0.0915
	(0.0156)	(0.0158)	(0.0163)	(1.888)	(1.722)	(1.876)	(0.112)	(0.117)	(0.130)
Perc. Status Responder	-0.0924	-0.0954	0.0108	-15.14**	-15.32***	-22.32***	0.189	0.298	0.227
	(0.0588)	(0.0607)	(0.0686)	(6.607)	(5.836)	(8.263)	(0.414)	(0.415)	(0.457)
Perc. Attractiveness Resp.	0.00163	0.00591	-0.0205	3.981*	4.030**	4.542**	0.105	0.123	0.0324
	(0.0176)	(0.0174)	(0.0175)	(2.046)	(1.812)	(2.199)	(0.113)	(0.115)	(0.126)
(Pie-Mean) by Stage	0.00110	0.000638	0.000288	1.212***	1.223***	1.296***	-0.00157	-0.00211	-0.00299
	(0.000791)	(0.000802)	(0.000831)	(0.0999)	(0.0893)	(0.120)	(0.00327)	(0.00350)	(0.00584)
Remaining Time	0.000653**	0.000588**	0.00113***	0.0835*	0.0864**	0.0504	0.00602**	0.00650***	0.00737***
	(0.000274)	(0.000262)	(0.000317)	(0.0470)	(0.0403)	(0.0331)	(0.00242)	(0.00240)	(0.00285)
No. Of Rounds	-0.0298***	-0.0300***	-0.0324***	-7.263***	-6.614***	-6.563***			
	(0.00765)	(0.00760)	(0.00846)	(1.039)	(0.904)	(0.781)			
First Round Offer	-0.00211*	-0.00206*	-0.00309**	-1.108***	-1.098***	-1.235***	-0.0483***	-0.0494***	-0.0499***
	(0.00120)	(0.00114)	(0.00146)	(0.211)	(0.194)	(0.274)	(0.00883)	(0.00967)	(0.0107)
Proposer Starts				19.12***	16.94***	12.10	0.920**	0.862*	0.516
				(6.930)	(5.913)	(7.390)	(0.404)	(0.463)	(0.530)
Proposer Accepts				-11.41***	-10.74***	-8.884***			
				(3.324)	(2.965)	(3.313)			
Constant				54.92***	58.62***	73.75***			
				(19.21)	(16.09)	(14.56)			
Stage FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	428	428	314	377	377	278	377	377	278
R-squared				0.994	0.995	0.996			
H_0: MF=FM	0.91	0.70	0.85	0.05	0.06	0.06	0.79	0.61	0.51
H_0: MF=MM	0.18	0.13	0.14	0.02	0.03	0.06	0.33	0.36	0.17
H_0: FM=MM	0.13	0.18	0.24	0.50	0.51	0.85	0.17	0.11	0.05

Notes The dependent variables, *Prob. of No Agreement*, *Proposer's Outcome* and *No. of Rounds* are defined in the notes of Table 2.4. Columns 1, 4 and 7, replicate columns 4-6 in Table 2.4. Columns 2, 5 and 8, show estimation results using a regression that weights each observation within each gender combination by the inverse of its probability. Columns 3, 6 and 9 show estimation results restricted to the matched sample. That matching is done following a nearest neighbor without replacement, where the treatment variable is defined as the dummy variable that takes the value of 1 if the responder is female and 0 otherwise. At the bottom, *p*-values for the hypothesis testing are shown where pairwise comparisons are made for the different gender combinations, where FF refers to bargaining between women, MF refers to the bargaining between a male proposer and a female responder, FM refers to the bargaining between a female proposer and a male responder and MM refers to the bargaining between men. Clustered standard errors at the proposer level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

with a view to balancing the sample in terms of the different matchings.[34] Estimation results shown in Table 2.6 (columns 2, 5 and 8) show that the results are the same in both quantitative and qualitative terms.

Second, we perform a regression analysis on a matched sample using probability score matching (see Rosenbaum and Rubin (1985), and Caliendo and Kopeinig (2008), for a practical guide). As there are fewer women in the sample of responders, we define our treatment variable as having a female responder and we estimate the probability score using the regression shown in Table 2.3.[35] We then use the nearest neighbor matching method to match the sample of female responders to that of male responders, such that they have similar propensity scores. The distributions of probability scores for proposers choosing a male and female responder, both for the unmatched and matched samples, are shown in Figure B.1.[36] We carry out the regression analysis within the matched sample only. Results are shown in Table 2.6 (columns 3, 6 and 9).[37] As expected, some observations are lost when the analysis is restricted to the matched sample, as some observations cannot be matched. More importantly, the results remain both quantitatively and qualitatively the same. In particular the effect on the main bargaining outcome, *Proposer's Outcome*, remains positive, significant, and very similar in size.

Third, given we observe important characteristics of both proposers and responders, we can also check how male proposers who choose female responders differ from male proposers who choose male responders, and how female responders who are chosen by male proposers differ from female responders who are chosen by female proposers. Columns 8 to 10 in Table 2.1 present a comparison of male proposers and clearly show that those who happen to choose a male responder do not differ significantly from those who happen to select a female responder. A similar comparison of female responders in columns 5 to 7 in Table 2.1 clearly shows that female responders who happen to be chosen by male proposers do not differ from female responders who are chosen by female proposers. Finally, an examination of the analysis in the matched sample shows that the minor differences in all the characteristics of both the proposers

[34] This approach is successful as long as the coefficient of *Male Proposer* is insignificant when we regress the gender of the responder on that of the proposer, and the constant is 0.5.

[35] We eliminate the independent variables that have to do with the male/female nature of the questions, as they cannot possibly influence the outcome variables of *Prob. of No Agreement*, *Proposer's Outcome* and *No. of Rounds*.

[36] As expected, if we replicate the analysis on Table 2.3 on the matched sample, we find that the only determinant of the responder's gender is given by the male/female perception of the question. Furthermore, when replicating Table 2.1 on the matched sample all significant differences within all matchings disappear with the exception of the effect of the question's perception. These results are available upon request.

[37] The results shown in Table 2.6 use matchings without replacement. We also used matching with replacement: Compared to those without replacement for the *Proposer's Outcome*, the outcome for male-female matchings is not significantly different from that male-male matchings, and for the *No. of Rounds*, the male-male matching takes significantly shorter than the female-female matching. It is known that standard errors increase with replacement, which lowers the significance of some results.

and the responders that appear in Table 2.1 become insignificant when all four different matchings are compared. This further confirms that the results are not driven by male proposers who choose a female responder being of a particular type, or by female responders chosen by a male proposer being of a particular type, but by women in the role of responders demanding less from men in the role of proposers. Nevertheless, as we mentioned in the introduction, we cannot rule out selection based on characteristics that are unobservable.

2.5 Conclusions

We use bargaining behavior in a TV show, where proposers choose individuals to bargain with, to study gender differences in the choice of bargaining partners, and gender differences and gender interaction effects in negotiations involving sizable stakes.

We find evidence that men show a stronger preference for male bargaining partners. This is consistent with taste-based discrimination but also with gender differences in beliefs. First, men might put more weight on the probability that men know the correct answer. Second, men might put more weight on the notion that they will get better deals when negotiating with men. Future work should be guided to distinguish between these hypotheses. Third, men might put more weight on the notion that women will be more reluctant to participate in a TV show.

Moreover, we find significant gender interaction effects in both bargaining behavior and bargaining outcomes. The male-female matching is found to be different from all the others. Contrary to the findings in previous publications, we find no evidence of differences in opening offers between male and female proposers or male and female responders. More importantly, it is women who demand less from male proposers. These results cannot be explained by selection, as male proposers who choose female responders do not differ from male proposers who choose male responders; nor do female responders who are chosen by male proposers differ significantly from female responders who are chosen by female proposers.

Three main conclusions can be drawn: First, gender interaction effects have proved to be crucial in understanding gender differences. Women demand less *only from men*. All these differences result in negotiations that are more favorable to men and less favorable to women when men negotiate with women. Second, we find no differential behavior in opening offers. Third, the most relevant gender differential results are found in the behavior of responders, who hold what is a priori a weaker position in this setting. Accordingly, it is only when men take the role of the strong player (the proposer) and women the role of the weak player (the responder) that strong gender differences are found. This result highlights the importance of the role played in bargaining. Dittrich et al. (2014), using a employer-employee setting, and Andersen et al. (2013), using a seller-buyer framing, also find gender interaction effects and gender differences, respectively, that depend on the roles played. Future work should be directed at understanding the *interaction* between gender interaction effects and the role played in bargaining.

Chapter 3

The Role of Gender and Asymmetries in Alternating-Offer Bargains

3.1 Introduction

The gender wage gap has long been an important object of study in economics. Although it has shown a decreasing trend over time, its persistence in developed countries challenges the classical explanations based on differences in human capital, preferences or statistical discrimination (Blau and Kahn (2000)).

Gender differences in negotiation have been put forward as an alternative explanation for the gender gap. Starting wages are often the result of bilateral negotiation. Moreover, wages are also affected by negotiations that come later in one's career, e.g. for pay increases. If women are less likely to negotiate starting salaries and to ask for a pay increase, and/or if women obtain worse deals when negotiating, this would clearly go some way towards explaining the gender wage gap (Azmat and Petrongolo (2014); Card et al. (2016)). The focus so far has been on gender differences in the likelihood of starting a negotiation. The influential book by Linda Babcock and Sara Laschever "Women Don't Ask" reveals important gender differences in the likelihood of negotiating. A study mentioned in the book shows that among graduates of Carnegie Mellon University 57% of men negotiated the starting salary offered to them, while only 8% of women did so. Exley et al. (2016) propose a controlled environment such as the laboratory to show that women are less likely to start a negotiation.[1] Studying gender differences in bargaining is equally important.

This paper seeks to shift the focus onto studying gender differences and gender

[1] Gender interaction effects have received no attention in studies of entry into negotiation. An exception is the paper by Eriksson and Sandberg (2012), who find that women are less likely to initiate a negotiation if they are matched with a female partner.

interaction effects in alternating-offer bargaining in asymmetric situations. We use a symmetric bargaining setting as a benchmark, where bargaining parties show equal strength so that a 50:50 split of the pie emerges, and where we expect neither gender differences nor gender interaction effects. We then introduce asymmetries, making one bargaining party stronger (*Proposer* in our setting) and the other weaker (*Responder* in our setting), in such a way that men and women are ex-ante equally likely to be allocated to either the strong or the weak position. The asymmetries are expected to make the division of the pie shift away from 50:50, giving the proposer a larger share of the pie. We investigate three types of asymmetry: empowerment (only the proposer has a positive outside option), entitlement (the proposer is entitled to a greater share than the responder) and informational power (only the proposer knows the actual size of the pie). We study gender differences and gender interaction effects both in bargaining outcomes (probability of reaching an agreement, time taken to reach an agreement and the responder's share of the pie) and bargaining behavior (offers and demands).

Our laboratory study consists of three main stages. Subjects first perform a real effort task, where each subject obtains a productivity which then determined the pie to be shared. In the second stage, subjects are randomly matched and have 3 minutes to bargain over the pie. The bargaining stage consists of 10 bargaining periods with a different matched participant each time. Finally, in the third stage we elicit a set of beliefs to measure their self-assessed ability in the task and in bargaining, as well as risk and social preferences.

In the symmetric bargaining benchmark we find that, as expected, the split is not different from 50:50. Also as expected, we find no gender differences or gender interaction effects. When asymmetries are introduced we find both gender differences and gender interaction effects. When the proposer is empowered, male responders do worse than female ones (by 6 percentage points), which is explained by male responders getting lower initial offers. This is the only exception, as in the other two asymmetries women obtain smaller shares. When the proposer is entitled, male bargainers obtain a greater share in both roles than female bargainers (4 and 6 percentage points less for the proposer and responder role, respectively). This is explained by men tending to offer less and demand more than women. When the proposer has more information than the responder, male proposers negotiating with female responders get the greatest share (by 6 percentage points), which shows that gender interactions are important when there are informational asymmetries. This difference in bargaining outcomes is explained by the demand side, where female responders demand less only from men. Interestingly, these gender differences are important and active in the first half of the experiment (first 5 periods) but vanish in the second half of the experiment (last 5 periods), so when all periods are considered together the gender differences are found to be much weaker.

We also find an important deadline effect, given the 3-minute time limit, where almost 25% of the successful bargains are reached within the last 10 seconds. This is consistent with previous findings in bargaining (Roth et al. (1988) and Gneezy et al. (2003)). We show that those agreements that are closed within the last 10 seconds, referred to as *ultimatums*, are indeed different from non-*ultimatum* deals. Ideally, we would like to get rid of the deadline effect, as it is more of an artifact given the exogenous time limit implemented in the laboratory.

Gender differences in bargaining have been studied by economists. For example, male proposers' behavior has been analyzed in studies of discrimination by carrying out field experiments in which the gender of potential scripted buyers is varied (Ayres (1991); Ayres and Siegelman (1995); Castillo et al. (2013)). To study gender differences in wage negotiation, Säve-Söderbergh (2007) uses wage bids and wage offers of recent graduates and finds that women post lower wage bids, and receive lower offers. More recently, Leibbrandt and List (2014), using a field experiment, find that women are less likely to negotiate wages when they are not described explicitly as negotiable, but that the difference disappears when they are described as negotiable. Economists have also studied gender differences in controlled settings such as the laboratory, mostly using the ultimatum game, which represents a reduced-form bargaining setting, as it allows for a single offer (or demand) and the response to it. Rigdon (2012) finds that women demand less than men in a demand-ultimatum-game in the laboratory, and Andersen et al. (2013) find that gender differences in bargaining depend on culture. More recently, Exley et al. (2016) study gender differences in the choice to negotiate, and in their baseline treatment, where subjects are *forced* to negotiate, they find that men and women achieve similar returns.

Gender interaction effects in bargaining have received less attention. Given that bargaining requires interaction between two agents, gender differences in one role may crucially depend on the gender of the interlocutor. Existing studies based on field data or field experiments do not study gender interaction effects, either because the gender of the person in one role is not known (e.g. Leibbrandt and List (2014)) or because there is not enough variation (e.g. Castillo et al. (2013)). Economists are thus limited to the use of laboratory experiments. Using mostly face-to-face ultimatum games, Eckel and Grossman (2001) and Solnick (2001) show that offers to women are lower than offers to men and that women are more likely to accept offers.[2] Sutter et al. (2009) find more competition and retaliation between same gender matchings than mixed gender matchings using the power-to-take game.

Note that all these bargaining settings mentioned above, both in the field and the laboratory, involve one or more types of asymmetries between bargaining roles, and they all find some type of gender differences when bargaining, with the exception of Exley et al. (2016). Two papers are more closely related to our setting because they allow to test for gender differences and gender interaction effects in asymmetric

[2] To be precise, Eckel and Grossman (2001) find that women are more likely to accept offers, while Solnick (2001) finds that women are more likely to accept offers from male proposers than from female proposers.

alternating-offer bargaining settings. Dittrich et al. (2014), using a laboratory face-to-face alternating-offer wage-bargaining game where the firm is empowered, find that starting salaries offered by men to women are lower than those offered by women to men, resulting in significant gender interaction effects on wage-bargaining outcomes. Hernandez-Arenaz and Iriberri (2016), using data from a TV-show, where bargaining parties show major asymmetries in all three dimensions (empowerment, entitlement and information), find that the matching between a male proposer (strong) and a female responder (weak) is the only one that differs from the rest, yielding higher profits for the proposer. Remarkably, we replicate this exact finding but only under the informational asymmetry, which shows that informational asymmetry is the main source for this gender interaction effect. This line of research shows that observed gender differences depend crucially on the gender of the interacting individual, so such interactions deserve equal attention, and that gender differences and gender interaction effects are directly connected to asymmetric bargaining roles.

This paper offers three main contributions. First, it is the first study that looks at different types of asymmetric bargaining situations in connection to gender differences. Asymmetric bargaining settings are the rule rather than the exception in economically relevant situations such as the labor market. The fact that all the studies mentioned above include some type of asymmetry confirms that the relevant settings indeed involve asymmetries between player roles. Unfortunately, most of the time different sources of asymmetry are confounded. This paper isolates each of the three sources of asymmetry in order to study whether gender differences and gender interaction effects are different for each type of asymmetry (empowerment, entitlement, and informational asymmetry). Second, although gender differences in bargaining outcomes have been studied using both dictator and ultimatum games, alternating-offer bargaining has received less attention. This paper studies structured bargaining which allows us to study not only bargaining outcomes but also bargaining behavior such as offers and demands. Third, studies that can test for gender interaction effects on top of gender differences are scarce. This laboratory study uses gender avatars to inform subjects about the gender of the bargaining party such that gender differences and gender interaction effects can be studied.

The rest of the paper is organized as follows. Section 3.2 describes the design and procedures of the laboratory experiment. Section 3.3 describes the data, the identification strategy, and the main results. Section 3.4 concludes.

3.2 Experimental Design

A laboratory experiment was run at the Bilbao Laboratory of Experimental Analysis (Bilbao Labean) at the University of the Basque Country on a computer based form using z-Tree experimental software (Fischbacher (2007)) to explore gender differences and gender interaction effects in asymmetric bargaining situations.[3] Subjects were recruited through ORSEE (Greiner (2015)), with a total of 162 participants –80 (49.4%) men and 82 (50.6%) women– split into four different sessions. Recruiting was carried out in such a way that the gender balance in each session was assured while subjects were unaware of this at the time of recruiting.

3.2.1 Design and Procedures

At the beginning of each session, subjects were provided with written general instructions which were also read aloud to guarantee that they were public knowledge. In particular, these general instructions informed subjects that the experiment consisted of 3 different stages and that the detailed instructions would be displayed on their computer screens before the start of each stage. A translation of the instructions can be found in Appendix C.1. Each session lasted for about one and a half hours, including payment. Average earnings were 14.7 Euro (s.d. 5.6) including a show-up fee of 3 Euro, and total earnings ranged from 5 Euro to 34.5 Euro.

All sessions included three different stages: a real effort task, an alternating-offer bargaining task and a set of elicitation tasks. The real effort task and the elicitation tasks were identical in all sessions, but we varied the bargaining setting form one session to another in order to generate four different treatments: *Symmetric*, *Empowerment*, *Entitlement*, and *Information*. These treatments differ from one another in the source of the induced bargaining asymmetry. In the rest of this section we explain each stage of the experiment in detail and outline the differences between the treatments.

Real Effort Task: Subjects were presented with a matrix filled with "0"s and "1"s similar to that in Figure 3.1 and asked to count the number of ones.[4] Once a number was entered for a matrix and the subject confirmed the input a new matrix appeared on the screen. Subjects performed this task for 5 minutes and the performance measure was the total number of matrices for which the correct number of "1"s was provided.[5] This task was not directly incentivized but subjects were informed that their performance in this task would determine their earnings in the bargaining stage.[6] Consistently with previous findings, this task proved to be gender neutral in performance, in

[3] The experimental design was registered at the *AEA RCT registry* before any sessions were run, under the reference AEARCTR-0002029. The preplan analysis can be checked at `https://www.socialscienceregistry.org/trials/2029/history/15499`.

[4] A similar task was used, for example in Abeler et al. (2011) and Mengel (2015).

[5] The z-Tree program was designed such that the maximum number of matrices that could be attempted was 60. This was explicitly sated in the instructions. Data show that this constraint is not binding as the maximum number of matrices that a subject faced was 33 with an average of 22.7.

[6] As will become clear in the explanation of the bargaining stage, the relationship between performance and the pie to be bargained over in the bargaining stage may induce competitive attitudes. To preclude any feeling of competition while subjects performed the real effort task, the instructions for the real effort explained that higher individual performances increased the pie that would be available in the bargaining stage on average but did not explain the exact link.

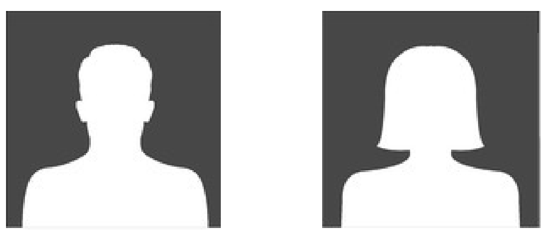
FIGURE 3.1– EXAMPLE OF A MATRIX SHOWN TO SUBJECTS DURING THE REAL EFFORT TASK

0	1	1	0	1	1
0	1	1	0	0	1
0	0	0	1	1	1
1	0	0	0	0	0
1	1	0	0	1	0
0	0	0	1	0	0

the number of matrices attempted, and in the failure rate.[7]

At the end of this stage and just before entering into the bargaining stage, subjects indicate their gender. In particular, they were presented with two avatars representing the silhouettes of a man and a woman and explicitly asked “Are you a man or a woman?”. As can be checked in Figure 3.2, these avatars were chosen to elicit subjects’ gender in the most aseptic and neutral wat possible, without giving any further cues such as facial expressions.[8]

FIGURE 3.2– GENDER AVATARS



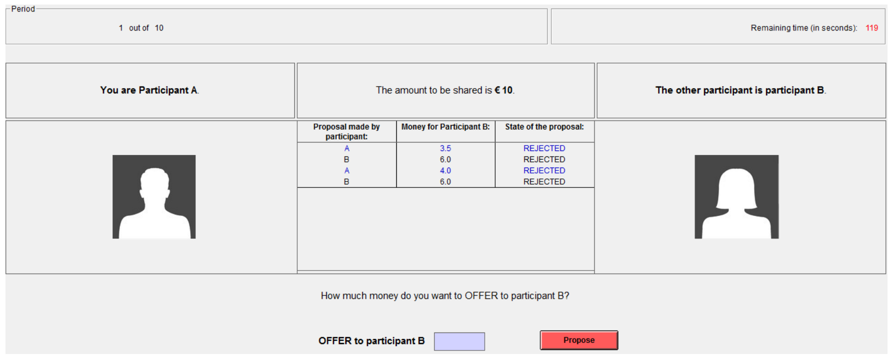
Bargaining Stage: Symmetric. Based on their relative performances in the real effort task, subjects were assigned a *productivity*, which determined the pie to be bargained over. In particular, the top third of performers were endowed with a productivity of €15, the middle third with a productivity of €10, and the bottom third with a productivity of €5. Subjects were informed about this protocol but no information about their actual productivity was provided.

Each subject was then randomly matched with another subject. One is assigned the role of participant A (hereafter the *Proposer*) and the other that of Participant B (here-

[7]Moreover, this gender neutrality in performance, effort, and productivity remains within each treatment. In addition, the three measures are similar in all four treatments.

[8]One potential concern is that this feature of the design could yield some type of experimental demand effect. Once stage 2 was finished but before the elicitation stage started, we explicitly asked "What do you think is the objective of this experiment?". This question could be answered using a free format. Only 13% of the subjects mentioned a gender related objective and there was no gender difference on this. These two results alleviate any concern about potential experimental demand effects.

FIGURE 3.3– SCREEN SEEN BY PROPOSERS DURING A BARGAINING IN THE SYMMETRIC TREATMENT



after the *Responder*). The role of *Proposer* was assigned to the subject in the matching with the higher score in the real effort task, although this protocol was not revealed.[9] Within each match the pie to be bargained over was randomly drawn from the productivity of the proposer and that of the responder with equal probabilities. This means that the potential pie size was 5, 10, or 15 Euro. Once the pie size was determined, this information was made public and each matching had 3 minutes to reach a deal on how to split the resulting pie through an alternating-offer bargaining process. During the bargaining proposers decided on offers to responders while responders decided on demands from proposers. In other words, the whole bargaining process took place in terms of the amount of money that the responder would get. During the bargaining, the information available to all subjects consisted of their own avatar and that of the opponent (their gender and that of their matched partner), the size of the pie to be shared and the bargaining history of offers and demands. See Figure 3.3 for an illustration. Importantly, subjects did not observe their own productivity or their opponent's. If they reached a deal within the 3-minute limit, the agreed split was implemented. Otherwise they got 0.

The whole bargaining process was repeated for 10 periods in all, with a different matched participant each time.[10] Importantly, from one period to the next the role in the bargaining matching (proposer or responder) and the pie to be split could change. For payment, subjects were informed that the computer would take two periods randomly –one from periods 1-5 and another from periods 6-10– and the resulting outcomes would be implemented.

[9]In particular, the subjects were just told that they would be given a bargaining role. Roles were assigned in this way in order to facilitate comparison across treatments. In case of ties, roles were randomly assigned.

[10]In the *Symmetric* treatment there was a technical problem and the z-Tree program stopped at the second repetition. We ran the bargaining module again and everything worked fine the second time. Thus, for the *Symmetric* treatment we gather data from 12 bargaining periods instead of 10 but given that periods 1 and 3 and periods 2 and 4 involve exactly the same matchings, we only consider periods 1-2 and 5-12 for the analysis of this treatment.

Bargaining Stage: Empowerment. Everything is as in the *Symmetric* treatment except that there was an outside option for the proposer. In particular, if a deal was not reached within the 3-minute limit the proposer had an outside option while the responder got 0. The outside option available to the proposer was a random amount drawn from a uniform distribution of between 50% and 85% of the pie. Both parties know about the outside option but neither knew its exact value when bargaining.[11]

Bargaining Stage: Entitlement. Everything is as in the *Symmetric* treatment except that the subjects were able to observe their own productivity and that of their partners. This was public knowledge. This treatment thus informed subjects of whose productivity determined the size of the pie. This was intended to generate a feeling of positive or negative entitlement.[12] In case of a tie, there is no entitlement effect, so that we do not consider those bargaining matchings in the analysis in the rest of the paper (note the lower number of observations in the entitlement treatment).

Bargaining Stage: Information. Everything is as in the *Symmetric* treatment except that only the proposer could observe the actual size of the pie, while the responder only knew that it could be 5, 10 or 15 Euro. This was public knowledge.

Elicitation Tasks. After completing the 10 bargaining periods, subjects entered the third and last stage of the experiment. In this stage we elicited beliefs about self-assessed ability and gender differences in performing both the real effort task and the bargaining task. Regarding the real effort task, subjects were asked to indicate which quartile of the performance distribution they thought they were in and to state which gender they believed had performed better (or whether there were no gender differences). Similarly, for the bargaining task subjects were asked to indicate which quartile of the distribution they thought they were in based on the relative surplus obtained during the 10 negotiations and to state which gender on average had obtained a greater share of the pie over the 10 periods (or whether there were no gender differences). All these measures were incentivized. In this stage we also elicited risk attitudes following the methodology in Eckel and Grossman (2002) and social preferences via the primary Slider Measure items described in Murphy et al. (2011) and implemented for z-Tree by Crosetto et al. (2012). These are used as additional controls.

[11]We decided not to provide the exact value of the outside option so as not to make that amount too salient. In addition, we decided to guarantee that the outside option would be at least 50% of the pie in order to properly implement a bargaining asymmetry through the introduction of an outside option. Notice that in this case, the Nash bargaining solution (Nash Jr (1950)) and the *deal-me-out* solution (Binmore et al. (1989)) return the same and, more importantly, agree on the effect of the outside option. By contrast, if the outside option is lower than 50% these two solution concepts disagree on whether the existence of an outside option has any effect.

[12]Notice that, by design, the productivity of the proposer is at least as high as that of the responder. So, we argue that when the size of the pie is the proposer's productivity the proposer feels a positive entitlement –the pie is high thanks to the proposer's productivity– while when the pie size is the responder's productivity the responder feels a negative entitlement –the pie is low because of the responder's productivity.

At the end of the experiment subjects completed a non-incentivized questionnaire that asked for standard demographics and for the big five personality traits (Gosling et al. (2003)).

3.2.2 Assessing the Design

We now check whether the *Empowerment*, *Entitlement*, and *Information* treatments generated the asymmetry we aimed for, and whether our design generated balanced gender combinations within the bargaining matchings. These two issues will confirm the validity of our design in testing for gender differences and gender interaction effects in asymmetric bargaining situations.

We first compare the *Symmetric* bargaining setting with the three asymmetric treatments. The *Symmetric* bargaining setting is the natural benchmark as it shows the most balanced bargaining situation between the proposer and the responder, so bargains can be expected to end in a 50:50 split. The data confirms this intuition: In the 164 bargaining matchings that reached a deal within the 3-minute limit the responder's average share of the pie is 0.502 (s.d. 0.103), which is clearly not statistically different from 0.5. In addition, the data show that 72.6% of the successful bargains in this treatment end up in the 50:50 split. We take this as evidence of the symmetry that exists in this treatment.

Table 3.1 shows the average treatment effect of each asymmetric treatment in comparison with the *Symmetric* treatment for the three main bargaining outcome variables: The likelihood of reaching an agreement within the 3-minute limit (*Success*), the time in seconds from the reaching of an agreement to the end of the 3-minute limit (*Remaining Time*), and the share of the pie that is obtained by the responder in each successful negotiation (*Responder's Pie Share*).[13]

As intended in the design, the results in columns (3) and (6) of Table 3.1 show that the *Empowerment*, *Entitlement*, and *Information* treatments give significant advantages to the proposer, enabling her/him to obtain a greater share of the pie. This asymmetry in bargaining roles is further evidenced by the fact that the proportion of divisions other than a 50:50 split in each asymmetric treatment is 100% for *Empowerment*, 75% for *Entitlement*, and 72% for *Information*. The three asymmetric treatments break the symmetry between the bargaining parties and make the proposer's role stronger than the responder's.

In addition, columns (1) and (4) of Table 3.1 also show that neither the existence of asymmetries nor their nature seem to have much effect on the probability of reaching a deal within the 3-minute limit, with the exception of a slight difference between

[13] Alternatively, the number of offers and demands can be used as an indicator of bargaining length instead of the *Remaining Time* variable. The results are qualitatively the same.

TABLE 3.1– AVERAGE TREATMENT EFFECT ON THE OUTCOME VARIABLES

	Success (1)	Remaining Time (2)	Responder's Pie Share (3)	Success (4)	Remaining Time (5)	Responder's Pie Share (6)
Empowerment	0.0293	-30.05***	-0.238***	0.0381	-25.15**	-0.217***
	(0.0529)	(11.53)	(0.0217)	(0.0511)	(11.31)	(0.0250)
Entitlement	0.0634	-21.83**	-0.0542***	0.0616*	-19.93**	-0.0475***
	(0.0439)	(9.523)	(0.0157)	(0.0347)	(9.610)	(0.0159)
Information	-0.0106	-40.13***	-0.109***	-0.0218	-38.44***	-0.0960***
	(0.0532)	(10.13)	(0.0221)	(0.0472)	(9.613)	(0.0207)
Constant	0.880***	113.0***	0.550***	1.005***	140.9***	0.569***
	(0.0486)	(9.883)	(0.0209)	(0.137)	(20.78)	(0.0526)
Individual Controls	NO	NO	NO	YES	YES	YES
Observations	748	626	626	748	626	626
Clusters Prop	42	42	42	42	42	42
Clusters Resp	42	42	42	42	42	42
R-squared	0.024	0.106	0.394	0.058	0.131	0.446
H_0: Emp=Ent	0.4118	0.3936	0.0000	0.5529	0.5925	0.0000
H_0: Emp=Inf	0.3271	0.3071	0.0000	0.1036	0.1481	0.0000
H_0: Ent=Inf	0.0461	0.0251	0.0353	0.0326	0.0293	0.0529

Notes: OLS for the mean effect of each treatment on the main outcome variables. The omitted treatment is *Symmetric*. All regressions control for *Pie Size* and include Period Fixed Effects. *Success* takes a value of 1 if the subjects reach a deal within the 3-minute limit and 0 otherwise. *Remaining Time* is the number of seconds from the time when agreement is reached until the three-minute limit expires. *Responder's Pie Share* is the share obtained by the responder in each successful bargain. *Entitlement* considers only those matchings in which an effective entitlement is implemented, so ties between subjects' productivities are disregarded. *Individual Controls* include subjects' gender, risk and social preferences and their self-assessed ability levels in the real effort task and in the bargaining ability, separately for both *Proposers* and responders. Standard errors are clustered at the proposer and responder level using two-way clustering. *** p<0.01, ** p<0.05, * p<0.1

the *Entitlement* and the *Information* treatments. This shows that asymmetries do not affect the efficiency –measured by the likelihood of failing to reach a deal and thus destroying the pie– of the bargaining. Columns (2) and (5) of Table 3.1 also suggest that asymmetries do affect the time required to reach an agreement, where all asymmetric treatments require longer than in the *Symmetric* setting. This suggests that introducing asymmetries, whatever their nature, into the bargaining setting increases the conflict with respect to a symmetric situation.

We now assess whether the matching protocol generated a balanced gender matching distribution. Since this study aims to look at potential gender differences and gender interaction effects under the presence of different bargaining asymmetries, a crucial step is to look at whether all possible gender matchings are balanced across and within treatments. While the treatment assignment as well as the matchings made are completely random, the role assigned to each party is not. In particular, within each matching the party with the higher score in the real effort task is the one that is assigned the role of proposer. However, given the gender neutrality of the real effort task the ex-ante expectation is that all matchings should be evenly represented. This is confirmed in Table 3.2, where it can be checked that, within each treatment, each

TABLE 3.2– DISTRIBUTION OF GENDER MATCHINGS WITHIN EACH TREATMENT

	Symmetric	Empowerment	Entitlement	Information	Total
FF	24.5%	24.0%	26.1%	27.1%	25.4%
MF	27.0%	22.5%	26.1%	22.9%	24.5%
FM	24.0%	29.5%	22.5%	27.6%	26.2%
MM	24.5%	24.0%	25.4%	22.4%	23.9%
Male Proposer	51.50%	46.50%	51.45%	45.24%	48.40%
Male Responder	48.50%	53.50%	47.83%	50.00%	50.13%
Observations	200	200	138	210	748

different matching accounts for close to 25%, the figure expected under full randomization. It can also be checked in Table 3.2 that within each treatment close to 50% of the matchings have a male proposer and 50% a male responder.

Thus, Table 3.2 shows that the procedure implemented in the experimental design and the gender neutrality in performance of the real effort task used in stage one generated a distribution of matchings which was balanced both across and within treatments. In order words, men and women had ex-ante equal probabilities of being assigned the strong and weak bargaining roles.

3.3 RESULTS

3.3.1 DATA, VARIABLES OF INTEREST, AND IDENTIFICATION STRATEGY

From the experiment we gather data on 738 different bargains coming from 162 different experimental subjects. This data includes variables of interest of two different types: *Bargaining Outcomes* and *Bargaining Behavior*. Regarding *Bargaining Outcomes*, we look at the probability of reaching an agreement within the 3-minute limit (*Success*), the time required to reach an agreement (*Remaining Time*), and the share of the pie that is obtained by the responder (*Responder's Pie Share*).[14] Regarding *Bargaining Behavior* we study opening offers (*1^{st} Offer*), *Subsequent Offers*, and *Demands*.[15] We differentiate between opening and subsequent offers because the very first offer is the only action in the bargaining process that can be considered to be exogenous: All other actions in the bargaining are affected by the past bargaining history.

Given this data and the research question of the present study, we performed two

[14] Needless to say, we could have looked at the share of the pie captured by the proposer. However, we decided to use the responder's share because offers and demands were made based on the responder's share.

[15] We have also analyzed the probabilities of offers and demands being accepted. However, given the higher endogeneity involved in these probabilities, the results are harder to interpret, so we relegate this analysis to the Appendix (See Table C.1 in Appendix C.2)

different analyses for each variable of interest Y and for each treatment as follows:

$$Y_{ij} = \alpha + \beta_1 MaleProp_i + \beta_2 MaleResp_j + \gamma X_{yij} + \epsilon_{yij} \quad (3.1)$$

$$Y_{ij} = \alpha' + \beta_1' MaleProp_i + \beta_2' MaleResp_j + \beta_3' MaleProp_i * MaleResp_j + \gamma' X_{yij} + \epsilon_{yij}' \quad (3.2)$$

where $MaleProp_i$ ($MaleResp_j$) takes a value of 1 if the *Proposer* i (*Responder* j) is a man and 0 for a woman and X_{yij} is a set of control variables specific to variable Y and the matching between *Proposer* i and *Responder* j. Specification (3.1) enables us to test whether gender differences in bargaining can be detected, i.e. whether men and women in the role of *Proposer*/*Responder* obtain different outcomes from bargaining and/or behave differently while bargaining. In this specification our coefficients of interest are β_1 and β_2. Additionally, for each analysis we show the results from the specification in (3.2) to test whether the potential gender effect detected under specification (3.1) is independent of the gender of the other bargainer, i.e. whether there is any gender interaction effect. In this specification our coefficient of interest is β_3'. Thus in the rest of the analysis we focus on the sign and significance of the coefficients β_1, β_2, and β_3'.

In studying bargaining outcomes we use an OLS specification and perform a two-way clustering at the proposer and responder level (Cameron et al. (2011); Thompson (2011)). When analyzing bargaining behavior in successful negotiations we exploit the panel structure of the database, i.e. we use the round by round bargaining data but specify each individual matching as the panel variable and estimate a random effects model. In this specification we cluster standard errors at the identity of the decision maker (i.e. at the proposer/responder level when analyzing offers/demands). Two types of control variable are applied in all regressions: First there are control variables that refer to the characteristics of the proposers and responders, such as risk preferences, social value orientation, self-assessed ability in the real effort task, and self-assessed ability in bargaining. All these variables are incentivized elicited measures from the experimental design. Second there are controls specific to the bargaining matching. In this regard, all regressions control for the *Pie Size* –with the exception of the responder's behavior in the *Information* treatment– and include *Period* fixed effects to account for potential common trends in learning.

3.3.2 Estimation Results

Table 3.3 shows the results for *Bargaining Outcomes* (Panel A) and *Bargaining Behavior* (Panel B) for the full sample. First, Panel A shows that all the figures in the symmetric bargaining setting are closer to zero than the corresponding ones in the asymmetric treatments. This suggests that gender differences play a more important role

TABLE 3.3– GENDER AND GENDER INTERACTION EFFECTS

Panel A: Bargaining Outcomes

	Symmetric			Empowerment			Entitlement			Information		
	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	-0.00329	-18.78	-0.00920	-0.0581	-23.40**	-0.0350	0.00272	-15.32*	-0.0194*	-0.0824*	-1.923	-0.0241
	(0.0662)	(13.70)	(0.0120)	(0.0682)	(11.56)	(0.0294)	(0.0607)	(8.167)	(0.0103)	(0.0428)	(9.768)	(0.0244)
β_2: Male Resp.	-0.0457	5.110	0.0110	-0.0724*	-33.15***	-0.0327**	0.00697	12.96	0.0230	-0.0644	-14.84	0.0394
	(0.0691)	(9.604)	(0.0186)	(0.0402)	(9.277)	(0.0151)	(0.0525)	(12.85)	(0.0229)	(0.0549)	(10.24)	(0.0244)
β_3': Male#Male	0.0185	38.46*	0.00195	0.0520	10.83	-0.0255	0.00865	7.093	-0.0129	-0.107	-34.89**	0.0179
	(0.0762)	(21.74)	(0.0309)	(0.102)	(11.45)	(0.0280)	(0.129)	(17.93)	(0.0311)	(0.101)	(16.09)	(0.0254)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	200	164	164	200	170	170	138	122	122	210	170	170
Clusters Prop	38	37	37	38	37	37	26	26	26	38	38	38
Clusters Resp	37	36	36	38	37	37	27	27	27	37	37	37

Panel B: Bargaining Behavior

	Symmetric			Empowerment			Entitlement			Information		
	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	-0.0363	-0.0163**	-0.00396	-0.0147	-0.0257**	0.00834	-0.0900**	-0.00624	-0.0100	0.0112	-0.00218	-0.218**
	(0.0305)	(0.00751)	(0.00693)	(0.0372)	(0.0114)	(0.00805)	(0.0377)	(0.00977)	(0.0125)	(0.0409)	(0.00852)	(0.0972)
β_2: Male Resp.	0.0390	0.00598	0.00246	-0.0584***	-0.00786	0.00720	-0.00991	-0.00145	0.0375*	0.0216	0.00818*	0.367**
	(0.0242)	(0.00552)	(0.00970)	(0.0178)	(0.00635)	(0.00718)	(0.0239)	(0.00832)	(0.0197)	(0.0148)	(0.00462)	(0.144)
β_3': Male#Male	0.0941*	-0.0161	0.0161	0.00259	-0.00308	-0.0163	-0.00124	-0.0120	0.000824	0.0213	0.0115	0.141
	(0.0473)	(0.0118)	(0.0133)	(0.0298)	(0.0109)	(0.0143)	(0.0424)	(0.0173)	(0.0292)	(0.0357)	(0.00855)	(0.202)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	164	664	608	170	985	877	122	706	644	170	1,133	1,054
No. Bargains	164	117	109	170	155	146	122	93	86	170	154	146
No. Clusters	37	33	33	37	34	36	26	26	27	38	38	36

Notes: All regressions control for *Pie Size* and include Period Fixed Effects. **Panel A**: OLS for the three main outcome variables for each treatment. *Success* takes a value of 1 if the subjects reach a deal within the 3-minute limit and 0 otherwise. *Remaining Time* is the number of seconds from the time when agreement is reached until the three-minute limit expires. *Responder's Pie Share* is the share obtained by the responder in each successful bargain. *Entitlement* considers only those matchings in which an effective entitlement is implemented, so ties between subjects' productivities are disregarded. *Individual Controls* include subjects' risk and social preferences and their self-assessed ability levels in the real effort task and in the bargaining ability, separately for both *Proposers* and responders. Standard errors are clustered at the proposer and responder level using two-way clustering. **Panel B**: GLS random-effects model for opening offers (1^{st} *Offer*), *Subsequent Offers*, and *Demands*. All variables represent the *Responder's Pie Share*, except *Demands* in the *Information* treatment, which shows absolute values. *Other Controls* include the round and the time remaining in seconds at the point when the offer (demand) has been made, the previous offer, the previous demand, and the individual controls for the *Proposer* (*Responder*). Clustered standard errors at the proposer level for offers and at the responder level for demands. *** p<0.01, ** p<0.05, * p<0.1

when asymmetries are introduced. Second, there are some common tendencies across treatments. Having a male bargainer seems to decrease the probability of reaching an agreement ((1), (4), (7), and (10)) and lengthen the time taken to reach one (columns (2), (5), (8), and (11)). This is especially true for male proposers. Nevertheless, these coefficients are almost never significantly different from zero. Regarding the share of the pie (columns (3), (6), (9), and (12)), male proposers tend to get a greater share than female proposers but this is only significant for the entitlement treatment. In this regard, Panel B shows that male proposers in the entitlement treatment offer significantly less from the very beginning. In view of the positive effect that previous offers have on new offers, this would explain why male proposers obtained a greater share of the pie. Male responders also tend to get a greater (but not significantly) share of the pie. The only exception is that of male responders in empowerment, where their share is significantly smaller than that of female responders. Panel B suggests that the main reason for this is that male responders receive lower initial offers, so that subsequent offers also decrease, which would hurt their final bargaining outcome.

Overall, this table shows that gender differences and gender interaction effects seem to be modest and not significantly different from zero. This seems to be in contrast with the more substantial gender differences observed in bargaining behavior, as we find major gender differences in offers in empowerment and entitlement, and in demands in the information treatment.

In our experiment, subjects bargain for 10 periods with a different bargaining partner each time. A relevant question is whether gender differences and gender interaction effects show any time trend or difference over different periods, referred here to as learning or experience effects. The simplest analysis replicates the main results in Table 3.3 splitting the sample into the first half (first 5 periods) and the second half (last 5 periods).[16] Table 3.4 presents the results for both *Bargaining Outcomes* for the first half of the experiment (the first 5 periods) and the second half of the experiment (the last 5 periods). Table 3.5 presents the equivalent results for *Bargaining Behavior*.[17]

Interestingly, we find that the gender differences and gender interaction effects are substantial and significant in the first half of the experiment but practically vanish in the second half of the experiment.[18] Most of our important findings are related to differences in the share of the pie, so we now concentrate on this variable (columns (3), (6), (9), and (12) in Table 3.4). The previous findings of Table 3.3 are again found but are now bigger and some new results also arise. When empowerment is in place, male responders obtain a smaller share of the pie (by 6.5 percentage points), significant

[16] One potential concern is that gender matchings might not be balanced within each of the two parts. Table C.2 shows that gender matchings are also balanced within the first and second halves.

[17] For the sake of completeness, we also look at gender differences and gender interaction effects on the probability of accepting offers and demands. This data is available in Appendix C.2, in Table C.4.

[18] The results remain qualitatively the same if the analysis is performed by taking the first 3 and last 3 periods. See Table C.5 in Appendix C.2.

TABLE 3.4– GENDER AND GENDER INTERACTION EFFECTS BY EXPERIENCE: BARGAINING OUTCOMES

Panel A: First Half (Periods 1–5)

	Symmetric			Empowerment			Entitlement			Information		
	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	-0.0139	-29.89	-0.00892	-0.0902	-35.62***	-0.0400	0.0195	-0.992	-0.0460**	0.0273	3.289	-0.0383
	(0.0828)	(20.91)	(0.0156)	(0.0698)	(11.77)	(0.0363)	(0.107)	(16.03)	(0.0201)	(0.0537)	(10.55)	(0.0320)
β_2: Male Resp.	-0.0269	6.905	0.0423*	-0.0396	-27.24*	-0.0647***	0.0380	17.70	0.0600**	-0.105	-18.49	0.0561**
	(0.0776)	(11.33)	(0.0246)	(0.0756)	(14.96)	(0.0206)	(0.0660)	(14.78)	(0.0298)	(0.0764)	(12.07)	(0.0251)
β_3': Male#Male	-0.0920	97.73***	0.0137	-0.107	16.22	-0.0283	0.141	27.35	-0.00796	-0.283***	-31.75**	0.0555***
	(0.138)	(23.17)	(0.0495)	(0.123)	(22.20)	(0.0485)	(0.125)	(26.23)	(0.0471)	(0.0981)	(16.00)	(0.0152)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	100	82	82	100	90	90	66	58	58	105	88	88
Clusters Prop	35	34	34	34	33	33	24	24	24	35	34	34
Clusters Resp	35	32	32	32	31	31	24	24	24	36	35	35

Panel B: Second Half (Periods 6–10)

	Symmetric			Empowerment			Entitlement			Information		
	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	0.0102	-12.36	0.000650	-0.0383	-10.43	-0.0289	-0.0351	-16.78	0.0157	-0.114	-6.128	0.0216
	(0.0709)	(11.66)	(0.0122)	(0.0844)	(13.69)	(0.0315)	(0.0901)	(12.67)	(0.0198)	(0.0712)	(10.33)	(0.0239)
β_2: Male Resp.	-0.0439	4.494	-0.0184	-0.0963	-25.47***	0.0106	-0.0148	6.879	-0.00697	0.0295	-12.11	0.0126
	(0.0825)	(15.38)	(0.0190)	(0.0826)	(8.542)	(0.0236)	(0.0687)	(11.50)	(0.0201)	(0.0595)	(12.21)	(0.0324)
β_3': Male#Male	0.127	-13.38	-0.00493	0.123	12.60	-0.0109	-0.128	-21.16	0.0191	0.0554	-43.35**	-0.0360
	(0.155)	(30.61)	(0.0311)	(0.141)	(20.04)	(0.0416)	(0.184)	(29.96)	(0.0234)	(0.166)	(18.75)	(0.0559)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	100	82	82	100	80	80	72	64	64	105	82	82
Clusters Prop	32	30	30	33	31	31	24	24	24	35	32	32
Clusters Resp	33	32	32	32	30	30	26	26	26	34	34	34

Notes: All regressions control for *Pie Size* and include Period Fixed Effects. OLS for the three main outcome variables for each treatment. *Success* takes a value of 1 if the subjects reach a deal within the 3-minute limit and 0 otherwise. *Remaining Time* is the number of seconds from the time when agreement is reached until the three-minute limit expires. *Responder's Pie Share* is the share obtained by the responder in each successful bargain. *Entitlement* considers only those matchings in which an effective entitlement is implemented, so ties between subjects' productivities are disregarded. *Individual Controls* include subjects' risk and social preferences and their self-assessed ability levels in the real effort task and in the bargaining ability, separately for both *Proposers* and responders. Standard errors are clustered at the proposer and responder level using two-way clustering.*** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 3.5– GENDER AND GENDER INTERACTION EFFECTS BY EXPERIENCE: BARGAINING BEHAVIOR

Panel A: First Half (Periods 1–5)

	Symmetric			Empowerment			Entitlement			Information		
	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	-0.0310	-0.0411***	-0.0151	-0.0291	-0.0371**	0.0185*	-0.0720**	-0.000664	-0.0221	0.0102	0.00303	-0.398**
	(0.0432)	(0.0145)	(0.0137)	(0.0375)	(0.0168)	(0.00958)	(0.0324)	(0.00795)	(0.0148)	(0.0498)	(0.0148)	(0.200)
β_2: Male Resp.	0.0453	0.0252**	0.00690	-0.0683**	-0.0325***	0.0141	0.0567	-0.00291	0.0419**	0.0268	0.00944	0.353
	(0.0360)	(0.0121)	(0.0135)	(0.0310)	(0.0109)	(0.0111)	(0.0394)	(0.0109)	(0.0207)	(0.0261)	(0.00878)	(0.242)
β_3': Male#Male	0.140*	0.0240	0.0209	0.0353	0.00727	-0.0366*	0.0493	0.00367	0.0243	0.0233	0.00420	0.322
	(0.0782)	(0.0271)	(0.0225)	(0.0491)	(0.0210)	(0.0197)	(0.0863)	(0.0199)	(0.0355)	(0.0533)	(0.0180)	(0.395)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	82	277	250	90	421	359	58	272	245	88	440	402
No. Bargains	82	53	48	90	81	74	58	41	38	88	78	72
No. Clusters	37	33	33	37	34	36	26	26	27	38	38	36

Panel B: Second Half (Periods 6–10)

	Symmetric			Empowerment			Entitlement			Information		
	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	-0.0331	0.00512	0.00156	0.0140	-0.0148***	-0.000765	-0.105**	-0.00567	0.00202	-0.000535	-0.00197	-0.233**
	(0.0342)	(0.00755)	(0.00801)	(0.0502)	(0.00453)	(0.0101)	(0.0494)	(0.0128)	(0.0151)	(0.0469)	(0.00709)	(0.110)
β_2: Male Resp.	0.0248	-0.0127	-0.0106	-0.0380	0.00598	0.00721	-0.0694**	-0.00546	0.0416	0.0231	0.00647	0.349**
	(0.0294)	(0.00852)	(0.0121)	(0.0259)	(0.00440)	(0.00744)	(0.0284)	(0.0141)	(0.0304)	(0.0217)	(0.00492)	(0.160)
β_3': Male#Male	0.0501	-0.0357*	0.0232	-0.0346	-0.00601	-0.0120	-0.0253	-0.0299	-0.00999	-0.0261	0.0157*	-0.0245
	(0.0573)	(0.0183)	(0.0198)	(0.0605)	(0.00765)	(0.0175)	(0.0537)	(0.0235)	(0.0375)	(0.0512)	(0.00919)	(0.237)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	82	387	358	80	564	518	64	434	399	82	693	652
No. Bargains	82	64	61	80	74	72	64	52	48	82	76	74
No. Clusters	37	33	33	37	34	36	26	26	27	38	38	36

Notes: GLS random-effects model for opening offers (*1^{st} Offer*), *Subsequent Offers*, and *Demands*. All variables represent the *Responder's Pie Share*, except *Demands* in the *Information* treatment, which shows absolute values. *Other Controls* include the round and the time remaining in seconds at the point when the offer (demand) has been made, the previous offer, the previous demand, and the individual controls for the *Proposer* (*Responder*). Clustered standard errors at the proposer level for offers and at the responder level for demands. *** p<0.01, ** p<0.05, * p<0.1

at 1%. This is explained by male responders getting lower initial offers and lower following offers (even once previous offers are controlled for). When the proposer is entitled, male proposers are still observed to get a greater share (by 4.6 percentage points), which is now significant at 5%. In addition, we find that male responders also get a greater share (by 6 percentage points), which is significant at 5%. This is explained by male proposers making lower initial offers and by male responders demanding more. Finally, when informational asymmetries are implemented we find a strong gender interaction effect, as the matching given by a male proposer and a female responder is the only different matching. In this matching women as responders get a smaller share (by 6.5 percentage points) when matched with a male proposer than when matched with a female proposer, which is significant at 1%. Importantly, this male-female matching is different, not only from the female-female matching but also from any other matchings, but all the other matchings are indistinguishable from one another. This finding is explained by the demand side, as Panel A of Table 3.5 shows that women demand less only from men.[19] Interestingly, this replicates the findings of Hernandez-Arenaz and Iriberri (2016), who also find that women demand less only from men and that this is the only matching that is different in the share of the pie. In their bargaining setting there were three sources of asymmetry (empowerment, entitlement and information), so this finding suggests that the main driver of their results is the existence of informational asymmetry.

These results show that gender differences and gender interaction effects are strong and active among inexperienced bargainers but that they quickly disappear as bargainers gain experience.

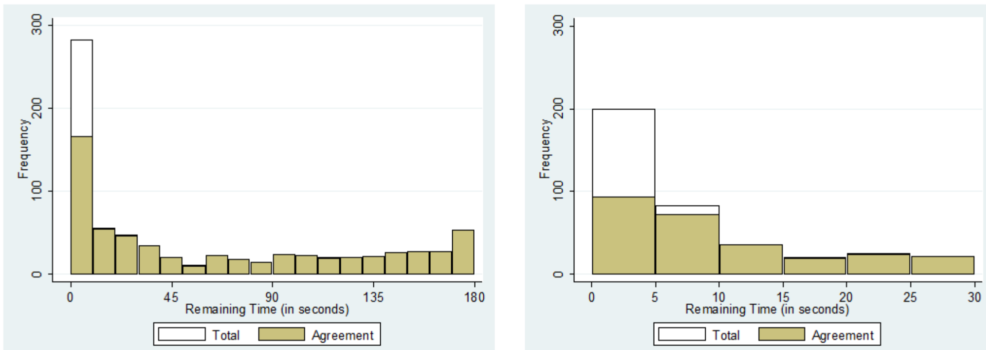
3.3.3 Deadline Effect

One consistent finding when looking at bargaining processes is the so called "*deadline effect*". This has been shown to shift a substantial amount of agreements toward the deadline, delaying the whole process. This effect has been widely documented both with field data (see for example Cramton and Tracy (1992)) and in the lab (see for example Roth et al. (1988) or Gächter and Riedl (2005)).[20] The existence of such a deadline effect is important, not only because delays in agreements may generate inefficiencies but also because they seem to be caused by bargainers for strategic reasons (Sterbenz and Phillips (2001); Gneezy et al. (2003)).

[19] Although this cannot be fully appreciated in Table 3.5, the analysis for gender interaction effects shows that the MF matching is different from the FF matching (Coef.: 0.554, *p*-value=.06), from the FM matching (Coef.: 0.765, *p*-value=.02), and from the MM matching (Coef.: 0.533, *p*-value=.07).

[20] The deadline effect is not exclusive to bargaining settings. It has also been documented in auctions, both in the field (Roth et al. (2002)) and in the lab (Ariely et al. (2005)). This is especially surprising because both Roth et al. (2002) in the field and Ariely et al. (2005) in the lab use data from second-bid auctions, in which there are no strategic reasons to delay as there could be the case –and indeed are as we document below– in a bargaining setting.

FIGURE 3.4– DEADLINE EFFECT



Notes: Distribution of the last proposals (offer or demand) over time in seconds. *Total* refers to the total number of matchings that make their last proposal in a given time window. *Agreement* refers to the number of matchings that make their last proposal in a given time window *and* which are accepted (Distribution of agreements over time).

As shown by Figure 3.4, our experimental data shows substantial deadline effects, which are consistent with previous findings. 37.8% of negotiations (283 out of 748) are still ongoing in the last 10 seconds –i.e., a bargaining party makes a new proposal within that time window. As pointed out above, this deadline effect has negative consequences for the efficiency of the bargaining process as delaying the negotiation can cause it to fail. In particular, our data shows that 41.3% among the bargains that are ongoing in the last 10 seconds end up failing to reach an agreement.[21] The left-hand picture in Figure 3.4 shows this effect clearly, enabling us to check that there is an atom of last proposals in the last 10 seconds of the negotiation while otherwise they are more or less uniformly distributed over time. The right-hand picture in Figure 3.4 confirms that the above pattern can be recognized even when we zoom-in on the last 30 seconds of the 3-minute limit.

Notice that as the timing of new proposals approaches the deadline, they can be considered as equivalent to take-it-or-leave-it proposals, as the chances of effectively making a counterproposal in the remaining time becomes very small so the receiver is obliged to accept the proposal or let the bargaining fail. Apart from this intuitive argument, the left-hand picture and more especially the right-hand picture in Figure 3.4 also support this idea by showing that proposals made within the last 10 seconds are either accepted or the negotiation ends in failure, while those proposals made with more than 10 seconds left are either accepted or are not the last proposal of the bargaining. This illustrates that the proposals that are made with more than 10 seconds left and are rejected give rise to counterproposals. Thus, we identify and refer here to proposals (independently of whether they are offers or demands) made within the last 10 seconds as *ultimatums*.

[21] As expected, most of the failed bargains are ongoing in the last 10 seconds. In particular, only 5 of the 122 (4%) bargains that ended up in disagreement were not ongoing in this time window.

TABLE 3.6– DISTRIBUTION OF ULTIMATUMS ACROSS TREATMENTS

	Symmetric	Empowerment	Entitlement	Information	Overall
Friendly	81.7%	75.9%	69.7%	65.9%	73.5%
Ultimatum	18.3%	24.1%	30.3%	34.1%	26.5%
Observations	164	170	122	170	626

Notes: Percentages computed out of the total of successful bargains. *Entitled* considers only those matching in which there is actual entitlement.

Table 3.6 shows the percentage of deals reached through an *ultimatum* and *friendly agreements* out of the total number of successful negotiations, overall and separately for each treatment. The proportion of deals reached under ultimatums is high in all four treatments although the data suggests that their proportion increases when there are asymmetries.[22]

Having established that ultimatums account for a large fraction of the successful negotiations in our data, it is natural to account for the effects that ultimatums have on the final bargaining outcome. Table C.3 in Appendix C.2 shows two things. First, it shows that the *Responder's Pie Share* is significantly affected by whether the deal is reached through friendly agreement or ultimatum agreements. Second, it also shows that whether the last proposal was a demand or an offer only affects *Responder's Pie Share* when the deal comes from an ultimatum.

Table 3.7 looks at gender differences and gender interaction effects on the propensity for closing a deal via an ultimatum in each of the four treatments.

The results suggest that a bargaining matching in which the proposer is a man is more likely in general to close a deal of this type. This is consistent with the findings in Table 3.3 that men as proposers tend to delay agreements longer. The only exception is the *Entitlement* treatment, in which the coefficient for male proposers has the opposite sign, though the size is modest. For responders, as in Table 3.3 with the *Remaining Time* variable, evidence is more mixed.[23]

Finally, we split the sample to test for gender differences and gender interaction effects separately for those matchings that reach a friendly agreement and those that reach an agreement via an ultimatum. Table 3.8 shows the results. Interestingly, a look at friendly agreements reveals, as expected, that gender differences are more in line with the effects found in bargaining behavior in Table 3.3. In *Empowerment* male responders obtain a smaller share, as before and in *Information*, male responders obtain a

[22]The proportion of deals closed in an ultimatum situation is similar to that typically found in the literature. For example, using data from 4 different experiments with a total of 1237 observations, Roth et al. (1988) found that the percentage of deals closed within the last 10 seconds was 28.3% which is similar to our overall figure of 26.5%.

[23]The results remain qualitatively the same if we analyze all last proposals, independently of whether they end up in agreement or not.

TABLE 3.7– PROBABILITY OF REACHING AN ULTIMATUM

	Symmetric (1)	Empowerment (2)	Entitlement (3)	Information (4)
β_1: Male Prop.	0.134**	0.277***	-0.0545	0.0761
	(0.0639)	(0.0792)	(0.0918)	(0.0790)
β_2: Male Resp	-0.0720	0.121**	-0.0458	0.136
	(0.0552)	(0.0577)	(0.0861)	(0.0963)
β_3': Male#Male	-0.0195	0.201	-0.120	0.0187
	(0.125)	(0.132)	(0.139)	(0.130)
Other Controls	YES	YES	YES	YES
Observations	164	170	122	170
Clusters Prop.	37	37	26	38
Clusters Resp.	36	37	27	37

Notes: OLS for the probability of closing a deal within the last 10 seconds. All regressions control for *Pie Size* and include *Period* Fixed Effects. *Other Controls* include risk and social preferences and self-assessment in the real effort task and in bargaining for both proposers and responders. Standard errors are clustered at the *Proposer* and *Responder* level using two-way clustering. *** p<0.01, ** p<0.05, * p<0.1

larger share. This last result was not significant before but becomes so when restricted to friendly agreements.[24]

To sum up, in this section we have shown that limiting bargaining to a fixed duration yields an important deadline effect, which shows important consequences in how the pie is split. Ideally, we would like to get rid of the deadline effect, as it is more an artifact given the exogenous time limit implemented in the laboratory. Future research for understanding gender differences and gender interaction effects should be directed to undermine the deadline effect.

3.4 DISCUSSION

Most, if not all, bargaining situations are asymmetric in economically relevant situations and it is hard to study gender differences and gender interaction effects in such real life settings. We thus propose a laboratory experiment to study gender differences and gender interaction effects in asymmetric bargaining situations.

We find that gender differences and gender interaction effects are absent in symmetric settings, where a 50:50 split is the norm, but become important when asymmetries between roles are introduced. The biggest differences are found in the role of the responder, where male responders obtain less when the proposer is empowered and

[24]Table C.6 in Appendix C.2 shows that if a distinction is drawn between the first and second halves of the experiment and only *friendly agreements* are considered then, consistently with the previous section, strong significant effects are found for the first half while nothing is found to be significant in the second. More importantly, this analysis of *friendly agreements* replicates the findings in the *Responder's Pie Share* of Table 3.4. Interestingly, coefficients are typically more significant and of a higher magnitude than in Table 3.4. This suggests that experience and deadline effects are capturing different things. Notice that the relatively low number of observations prevents this split from being performed on *ultimatum* deals.

TABLE 3.8– GENDER AND INTERACTION EFFECTS BY AGREEMENT'S TYPE: RESPONDER'S PIE SHARE

	Symmetric		Empowerment		Entitlement		Information	
	Friendly	Ultim.	Friendly	Ultim.	Friendly	Ultim.	Friendly	Ultim.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1: Male Prop.	0.0113	0.0816**	-0.0245	-0.0130	-0.0274	-0.0187	-0.0434	0.0119
	(0.0188)	(0.0372)	(0.0315)	(0.0195)	(0.0186)	(0.0459)	(0.0369)	(0.0201)
β_2: Male Resp.	0.00388	0.141***	-0.0498***	0.0105	0.00370	0.106***	0.0450**	0.0160
	(0.0165)	(0.0442)	(0.0155)	(0.00819)	(0.0183)	(0.0198)	(0.0222)	(0.0229)
β_3': Male#Male	0.00344	0.0703	-0.0492	-0.0459	-0.0390	0.0101	0.0263	0.0218
	(0.0280)	(0.0500)	(0.0369)	(0.0583)	(0.0310)	(0.0565)	(0.0384)	(0.0434)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	134	30	129	41	85	37	112	58
Clusters Prop.	37	18	35	20	25	17	33	29
Clusters Resp.	36	18	36	24	23	21	35	27

Notes: All regressions control for *Pie Size* and for who makes the last offer (Proposer or Responder) and include *Period* Fixed Effects. *Other Controls* include risk and social preferences and self-assessment in the real effort task and in bargaining for both, proposers and responders. Standard errors are clustered at the *Proposer* and *Responder* level using two-way clustering. Clustered standard errors at the individual level using two-way clustering. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

male responders get a greater share of the pie in both the entitlement and information cases. Furthermore, male proposers also obtain a greater share of the pie when entitled. Finally, when informational asymmetries are introduced, female responders demand less from male proposers, so they end up with the smallest share (consistent with Hernandez-Arenaz and Iriberri (2016)). Interestingly, the gender differences and gender interaction effects vanish in the second half of the experiment.

The deadline effect seems to be important and as this seems more of an artifact given the exogenous 3-minute time limit imposed in the laboratory, ideally we would like to get rid of it. Self-selection issues in this regard become important; are those subjects that reach an ultimatum different from others? A similar problem is found in auctions and some mechanism for avoiding the deadline effect in auctions (Roth et al. (2002); Ariely et al. (2005)) have proved to be useful. One important avenue for future research is using mechanisms of these types to avoid such a deadline effect.

Appendices

Appendix A

Stereotypes and Tournament Self-Selection

A.1 Technical Appendix

A.1.1 Formal derivation of expected utilities

Let $q_g : r \to [1,0]$ be a generic strategy profile for agents from social group $g \in \{S, N\}$ mapping each signal observed into the probability of entering tournament A and let $q_g^\alpha(r)$ be the corresponding one at equilibrium.

Given any arbitrary pair of strategy profiles $(q_s(r), q_n(r))$, the expected entry rate into tournament A of those agents whose ability is exactly a given the distributional assumptions $F^S(a)$ and $F^N(a)$ is

$$E\left[\lambda' q_s(r) + (1-\lambda')q_n(r)|a\right] = \int_{-\infty}^{\infty} \left[\lambda' q_s(a+\mu) + (1-\lambda')q_n(a+\mu)\right]\phi(\mu)d\mu$$

where λ' is the believed proportion of S-agents among those with ability a, which equals $\lambda' = \lambda f^S(a)/[\lambda f^S(a) + (1-\lambda)f^N(a)]$. Consequently, among the participants in tournament A, only those whose real ability exceeds a certain threshold a_A holding[1]

$$\int_{a_A}^{\infty} E\left[\lambda' q_s(r) + (1-\lambda')q_n(r)|a\right] dF^\alpha(a) \leq \delta_A \quad (= \text{ if } a_A > -\infty)$$

will expect to obtain a prize, where $F^\alpha(a) \equiv \lambda F^S(a) + (1-\lambda)F^N(a)$ is the CDF of the believed distribution of abilities in whole population. Therefore, given the pair of strategy profiles $(q_s(r), q_n(r))$, the believed expected return from participating in tournament A for an agent who observes signal r_i from social group $g \in \{S, N\}$ is

$$U[A|q_s(r), q_n(r), r_i, F^g] = W_A[1 - \hat{F}(a_A|r_i, F^g)] \tag{A.1}$$

[1] By replacing and doing the maths this expression can be transformed into

$$\lambda \int_{a_A}^{\infty}\left[\int_{-\infty}^{\infty} q_s(a+\mu)\phi(\mu)d\mu\right] dF^S(a) + (1-\lambda)\int_{a_A}^{\infty}\left[\int_{-\infty}^{\infty} q_n(a+\mu)\phi(\mu)d\mu\right] dF^N(a)$$

where $1-\hat{F}(a_A|r_i, F^g)$ is the self-assessed probability of having a real ability level higher than a_A.

Similarly, the expected entry rate in tournament B of those agents whose ability is exactly a is $1-E\left[\lambda' q_s(r)+(1-\lambda')q_n(r)|a\right]$ so only those with a real ability higher than a certain threshold a_B which holds

$$\int_{a_B}^{\infty} 1-E\left[\lambda' q_s(r)+(1-\lambda')q_n(r)|a\right] dF^{\alpha}(a) \leq \delta_B \quad (= \text{ if } a_B > -\infty)$$

will expect to obtain a prize. Therefore, given the pair of strategy profiles $(q_s(r), q_n(r))$, the expected return from participating in tournament B for an agent who observes signal r_i from social group $g \in \{S, N\}$ is

$$U[B|q_s(r), q_n(r), r_i, F^g] = W_B[1-\hat{F}(a_B|r_i, F^g)]$$

where $1-\hat{F}(a_B|r_i, F^g)$ is the self-assessed probability of having a real ability level higher than a_B.

A.1.2 CHARACTERIZATION OF THE EQUILIBRIUM

First I demonstrate formally that the equilibrium strategy profiles, $(q_s^{\alpha}(r), q_n^{\alpha}(r))$, should induce a situation in which tournament A is necessarily overcrowded. Consider any pair of strategy profiles $(q_s'(r), q_n'(r))$ such that the induced standard needed to beat for tournament A, a_A, is $-\infty$. Then it is easy to see that $\forall g \in \{S, N\}$

$$U[A|q_s'(r), q_n'(r), r_i, F^g] = W_A > W_B \geq U[B|q_s'(r), q_n'(r), r_i, F^g] \quad \forall r_i$$

which clearly contradicts the fact that the pair $(q_s'(r), q_n'(r))$ represent an equilibrium as all agents have incentives to choose tournament A. This has two immediate consequences. First, since tournament A is overcrowded the standard for winning in tournament A at equilibrium should be given by a value $a_A \equiv {}^{o}a^{\alpha} > -\infty$ holding

$$\int_{a^{\alpha}}^{\infty} E\left[\lambda' q_s^{\alpha}(r)+(1-\lambda')q_n^{\alpha}(r)|a\right] dF^{\alpha}(a) = \delta_A \tag{A.2}$$

which furthermore implies that $a^{\alpha} < \infty$. Second, since $\delta_A + \delta_B \geq 1$ in equilibrium tournament B should end up with vacancies, i.e. $\int_{a_B}^{\infty} 1-E\left[\lambda' q_s^{\alpha}(r)+(1-\lambda')q_n^{\alpha}(r)|a\right] dF^{\alpha}(a) < \delta_B$, $\forall a_B$ which implies the existence of a corner solution at $a_B = -\infty$, so $U[B|q_s^{\alpha}(r), q_n^{\alpha}(r), r_i, F^g] = W_B$, $\forall r_i, g$. This means that in equilibrium agents' decision can be simplified to two options: getting W_B for sure or getting W_A if his real ability happens to be greater than a^{α} and 0 otherwise.

Since the believed ability threshold needed to beat at tournament A to get the prize in equilibrium, a^{α}, is an interior point in $(-\infty, \infty)$, it can be checked from equation (A.1) that the boundary conditions for agents' expected utility in tournament A in equilibrium are $\lim_{r_i \to -\infty} U[A|q_s^{\alpha}(r), q_n^{\alpha}(r), r_i, F^g] = 0$ and $\lim_{r_i \to \infty} U[A|q_s^{\alpha}(r), q_n^{\alpha}(r), r_i, F^g] = W_A$.

Moreover, from equation (A.1) follows that $U[A|q_s(r), q_n(r), r_i, F^g]$ is differentiable and strictly increasing on r_i for any pair $(q_s(r), q_n(r))$ that induce an ability threshold $a^{\alpha} \in (-\infty, \infty)$. This strictly monotone increasing property of $U[A|q_s(r), q_n(r), r_i, F^g]$ together with its boundary conditions allows us to claim that the group-specific strategy profile in equilibrium must be also monotone increasing. In particular, for any pair of strategy profiles $(q_s^\alpha(r), q_n^\alpha(r))$ that constitutes an equilibrium there must be a signal value r_g^α for each $g \in \{S, N\}$ such that

$$U[A|q_s^\alpha(r), q_n^\alpha(r), r_g^\alpha, F^g] \equiv W_A[1 - \hat{F}(a^\alpha|r_g^\alpha, F^g)] = W_B \tag{A.3}$$

so that, by the strictly monotone property, $U[A|q_s^\alpha(r), q_n^\alpha(r), r_i, F^g] \gtrless U[B|q_s^\alpha(r), q_n^\alpha(r), r_i, F^g] = W_B$ if $r_i \gtrless r_g^\alpha$. Therefore it can be concluded that, provided a^α, the equilibrium strategy profiles for social group g is characterized by a signal contingent strategy profile such that $q_g^\alpha(r_i) = 1$ if $r_i > r_g^\alpha$ and $q_g^\alpha(r_i) = 0$ if $r_i < r_g^\alpha$. Substituting this strategy profiles into equation (A.2) follows that the believed real ability threshold at the equilibrium should hold[2]

$$\int_{a^\alpha}^{\infty} \lambda[1 - \Phi(r_S^\alpha - y)]f^S(y) + (1-\lambda)[1 - \Phi(r_N^\alpha - y)]f^N(y)dy = \delta_A \tag{A.4}$$

that together with equation (A.3) fully characterizes the equilibrium of this game.

A.1.3 Existence and uniqueness of the equilibrium

We know that the equilibrium is characterized by a cutoff strategy summarized by a signal value r_g^α for each social group (see Appendix A.1.2 above). Assume an arbitrary pair signal threshold $\bar{r}_S, \bar{r}_N$. Given imperfect self-knowledge, there are some agents from social group $g \in \{S, N\}$ who have very low (high) ability levels who received a signal $r > (<)\bar{r}_g$, and thus mistakenly opted for tournament A (B). Specifically, the probability of an agent from social group g with real ability a participating in tournament A is conditional on he/she receiving a signal $r > \bar{r}_g$. This probability can be computed as

$$Prob(r > \bar{r}_g|a) = Prob(a + \mu > \bar{r}_g) = 1 - \Phi(\bar{r}_g - a)$$

Therefore, given an arbitrary pair of threshold signals $(\bar{r}_S, \bar{r}_N)$, the mass of agents participating in tournament A in equilibrium can be computed as $\int_{-\infty}^{\infty} \lambda[1-\Phi(\bar{r}_S-y)]f^s(y)+(1-\lambda)[1-\Phi(\bar{r}_N - y)]f^N(y)dy$. Consequently the distribution of abilities *within* those participating in tournament A can be summarized by the CDF

$$\mathcal{A}^{\bar{r}_S,\bar{r}_N}(a) \equiv \frac{\int_{-\infty}^{a} \lambda[1 - \Phi(\bar{r}_S - y)]f^s(y) + (1-\lambda)[1 - \Phi(\bar{r}_N - y)]f^N(y)dy}{\int_{-\infty}^{\infty} \lambda[1 - \Phi(\bar{r}_S - y)]f^s(y) + (1-\lambda)[1 - \Phi(\bar{r}_N - y)]f^N(y)dy}$$

Thus, an agent with real ability a' believes (with certainty) that there is, within the agents participating in tournament A, a proportion of agents $\mathcal{A}^{\bar{r}_S,\bar{r}_N}(a')$ whose ability levels are lower than his/her own, while the remaining proportion $1 - \mathcal{A}^{\bar{r}_S,\bar{r}_N}(a')$ has ability levels strictly higher than a'.

On the other hand, notice that there is a mass of prizes of exactly δ_A for participants in tournament A. As the mass of agents in that tournament is $\int_{-\infty}^{\infty} \lambda[1 - \Phi(\bar{r}_S - y)]f^s(y) + (1-\lambda)[1 - \Phi(\bar{r}_N - y)]f^N(y)dy$, the proportion of agents participating in tournament A who obtain W_A is

$$\Delta_A^{\bar{r}_S,\bar{r}_N} = \frac{\delta_A}{\int_{-\infty}^{\infty} \lambda[1 - \Phi(\bar{r}_S - y)]f^s(y) + (1-\lambda)[1 - \Phi(\bar{r}_N - y)]f^N(y)dy} > \delta_A$$

[2] See footnote 1 for the intermediate step going from equation (A.2) to equation (A.4).

Thus the probability of obtaining a prize in tournament A is equal to the probability of being in the top $\Delta_A^{\bar{r}_S,\bar{r}_N}$ of the distribution $\mathcal{A}^{\bar{r}_S,\bar{r}_N}(a)$. But as we are working with a continuum of agents, for any pair $(\bar{r}_S,\bar{r}_N)$ there is an ability level $a(\bar{r}_S,\bar{r}_N)$ such that

$$1-\mathcal{A}^{\bar{r}_S,\bar{r}_N}(a(\bar{r}_S,\bar{r}_N))=\Delta_A^{\bar{r}_S,\bar{r}_N}$$

In other words, if an agent knows for sure that his/her ability level is higher than $a(\bar{r}_S,\bar{r}_N)$, he/she will think that his/her probability of winning the prize is 1.

Lemma A.1.1. *For any arbitrary pair $(\bar{r}_S,\bar{r}_N)$ there is a uniquely determined ability level $a(\bar{r}_S,\bar{r}_N)$ such that $1-\mathcal{A}^{\bar{r}_S,\bar{r}_N}(a(\bar{r}_S,\bar{r}_N))=\Delta_A^{\bar{r}_S,\bar{r}_N}$. This $a(\bar{r}_S,\bar{r}_N)$ solves*

$$\int_{a(\bar{r}_S,\bar{r}_N)}^{\infty}\lambda[1-\Phi(\bar{r}_S-y)]f^S(y)+(1-\lambda)[1-\Phi(\bar{r}_N-y)]f^N(y)dy\leq\delta_A\quad(=\text{ if } a(\bar{r}_S,\bar{r}_N)>-\infty)\tag{A.5}$$

Moreover, $\lim_{(\bar{r}_S,\bar{r}_N)\to(-\infty,-\infty)}a(\bar{r}_S,\bar{r}_N)=a^{\alpha,sd}$, $\lim_{(\bar{r}_S,\bar{r}_N)\to(\infty,\infty)}a(\bar{r}_S,\bar{r}_N)=-\infty$ *and* $\frac{\partial a(\bar{r}_S,\bar{r}_N)}{\partial\bar{r}_g}<0$ $\forall\bar{r}_g$, $g\in\{S,N\}$.

Proof: *The implicit solution for $a(\bar{r}_S,\bar{r}_N)$ in equation (A.5) is immediately apparent if we substitute $\mathcal{A}^{\bar{r}_S,\bar{r}_N}(a(\bar{r}_S,\bar{r}_N))$ and $\Delta_A^{\bar{r}_S,\bar{r}_N}$ by their extended expressions and do the math. On the other hand, uniqueness is easily proven once it is noted that equation (A.5) is monotonically decreasing in $a(\bar{r}_S,\bar{r}_N)$.*

If $(\bar{r}_S,\bar{r}_N)\to(-\infty,-\infty)$ equation (A.5) yields that $a(\bar{r}_S,\bar{r}_N)$ implicitly solves

$$\int_{\lim_{(\bar{r}_S,\bar{r}_N)\to(-\infty,-\infty)}a(\bar{r}_S,\bar{r}_N)}^{\infty}\lambda f^s(y)+(1-\lambda)f^s(y)dy=1-F^{\alpha}(a(\bar{r}_S,\bar{r}_N))=\delta_A$$

from what it emerges that the upper bound for the ability threshold is[3]

$$\lim_{(\bar{r}_S,\bar{r}_N)\to(-\infty,-\infty)}a(\bar{r}_S,\bar{r}_N)=(F^{\alpha})^{-1}(1-\delta_A)=a^{\alpha,sd}$$

Similarly, if $(\bar{r}_S,\bar{r}_N)\to(\infty,\infty)$ equation (A.5) shows that there is no interior solution for $a(\bar{r}_S,\bar{r}_N)$ as

$$\int_{\lim_{(\bar{r}_S,\bar{r}_N)\to(\infty,\infty)}a(\bar{r}_S,\bar{r}_N)}^{\infty}0dy=0<\delta_A,\quad\forall a(\bar{r}_S,\bar{r}_N)$$

so the corner solution applies, leading to $\lim_{(\bar{r}_S,\bar{r}_N)\to(\infty,\infty)}a(\bar{r}_S,\bar{r}_N)=-\infty$.

The last part of the lemma is proven straightforwardly by noting that for any change in the pair $(\bar{r}_S,\bar{r}_N)$ the value of $a(\bar{r}_S,\bar{r}_N)$ should also change such that the above equality holds. Thus, the total derivative with respect to $\bar{r}_g$, $g\in\{S,N\}$, in the foregoing expression should be zero, i.e.

$$-[1-\Phi(\bar{r}_g-a(\bar{r}_S,\bar{r}_N))]f(a(\bar{r}_S,\bar{r}_N))\frac{\partial a(\bar{r}_S,\bar{r}_N)}{\partial\bar{r}_g}-\int_{a(\bar{r}_S,\bar{r}_N)}^{\infty}\phi(\bar{r}_g-y)f(y)dy\equiv0\Rightarrow\frac{\partial a(\bar{r}_S,\bar{r}_N)}{\partial\bar{r}_g}<0$$

Thus, the mapping $a:(\bar{r}_S,\bar{r}_N)\to a$ can be represented by means of a monotonically decreasing, continuous function in both arguments whose codomain is the set $(-\infty,a^{\alpha,sd})$. □

[3]Notice that this value $a^{\alpha,sd}$ is just the ability threshold needed to beat when everyone enters tournament A. Therefore, it can be thought as to represent the minimum ability an agent should have in order to consider tournament A a strictly dominant strategy. This value $a^{\alpha,sd}$ also represents the minimum ability required to participate in tournament A under perfect self-knowledge.

The above lemma simply states that $a(\bar{r}_S, \bar{r}_N)$ can be considered as a function mapping each pair of potential signal levels behaving as a threshold to exactly one compatible real ability threshold level. Moreover, it implies that this function is upper bounded at $a^{\alpha,sd}$ and that it is inversely related to r such that $a(\bar{r}_S, \bar{r}_N) : \mathbb{R}^2 \rightarrow (-\infty, a^{\alpha,sd})$.

Now fix $a(\bar{r}_S, \bar{r}_N)$ and denote it as $\bar{a}$. Given $\bar{a}$, the assigned probability of an agent from social group $g \in \{S, N\}$ observing signal r_i for winning the prize in tournament A is just the probability that he/she assigns –given distributional assumptions– to having a real ability greater than $\bar{a}$ conditional on the signal he/she has observed, i.e. $1 - \hat{F}(\bar{a}|r_i, F^g)$.

Lemma A.1.2. *For any arbitrary $\bar{a}$ there is a uniquely determined $r_g(\bar{a})$ that solves*

$$W_A[1 - \hat{F}(\bar{a}|r_g(\bar{a}), F^g)] = W_B \tag{A.6}$$

Moreover, $\lim_{\bar{a}\rightarrow-\infty} r_g(\bar{a}) \rightarrow -\infty$, $\lim_{\bar{a}\rightarrow\infty} r_g(\bar{a}) \rightarrow \infty$ *and* $\frac{\partial r_g(\bar{a})}{\partial \bar{a}} > 0, \forall \bar{a}$.

Proof: *The proof is straightforward once it is noted that for any $\bar{a}$, $r_g(\bar{a})$ should be such that equation (A.6) holds. Thus, the total derivative with respect to $\bar{a}$ of equation (A.6) should be zero, i.e.*

$$W_A \left[-\hat{f}(\bar{a}|r_g(\bar{a}), F^g) - \frac{\partial \hat{F}(\bar{a}|r_g(\bar{a}), F^g)}{\partial r_g(\bar{a})} \frac{\partial r_g(\bar{a})}{\partial \bar{a}} \right] \equiv 0 \Rightarrow \frac{\partial r_g(\bar{a})}{\partial \bar{a}} > 0$$

The boundary conditions are straightforward and can be checked by applying the limits to equation (A.6). Thus, the mapping $r : \bar{a} \rightarrow r$ can be represented by means of a monotonically increasing, continuous function. □

Summing up, from lemmas A.1.1 and A.1.2 we find that any real ability value established as standard induces one signal threshold level for each social group that at the same time induce a new standard. Therefore, in equilibrium we should find that both the signal threshold levels (induced by the real ability standard) and the real ability level (induced by the signal threshold levels) should be compatible with the beliefs held by the agents. Summarizing, we are looking for a value a^α that represents a fixed point for the system $a(r_N(a^\alpha), r_S(a^\alpha)) = a^\alpha$. Now we can prove 1.2.2 very easily.

The first thing to notice is that, given the properties on lemmas A.1.1 and A.1.2

$$\frac{da(r_N(\bar{a}), r_S(\bar{a}))}{d\bar{a}} = \frac{\partial a(r_N, r_S)}{\partial r_N} \frac{\partial r_N(\bar{a})}{\partial \bar{a}} + \frac{\partial a(r_N, r_S)}{\partial r_S} \frac{\partial r_S(\bar{a})}{\partial \bar{a}} < 0$$

and that $a(r_N(\bar{a}), r_S(\bar{a}))$ is continuous on $\bar{a}$. Thus, if there is a^α holding $a(r_N(a^\alpha), r_S(a^\alpha)) = a^\alpha$, it must be unique. Next it can be seen that the limit of $a(r_N(\bar{a}), r_S(\bar{a}))$ when $\bar{a} \rightarrow -\infty$ is

$$\lim_{\bar{a}\rightarrow-\infty} a(r_N(\bar{a}), r_S(\bar{a})) = \lim_{(r_N, r_S)\rightarrow(-\infty,-\infty)} a(r_N, r_S) = a^{\alpha,sd}$$

and that the limit of $a(r_N(\bar{a}), r_S(\bar{a}))$ when $\bar{a} \rightarrow \infty$ is

$$\lim_{\bar{a}\rightarrow\infty} a(r_N(\bar{a}), r_S(\bar{a})) = \lim_{(r_N, r_S)\rightarrow(\infty,\infty)} a(r_N, r_S) = -\infty$$

Therefore, we find that the function $a(r_N(\bar{a}), r_S(\bar{a})) : \bar{a} \in \mathbb{R} \rightarrow (-\infty, a^{\alpha,sd})$ is continuous on $\bar{a}$ and it is strictly decreasing. Thus there is $a^\alpha > -\infty$ holding $a(r_N(a^\alpha), r_S(a^\alpha)) = a^\alpha$. □

A.1.4 Proof of propositions: Representativity and comparative statics

Proof of Proposition 1.2.4: Assume that the stereotype is negative so $\alpha < 0$. By corollary 1.2.3, $r_S^\alpha > r_N^\alpha$ so $1 - \Phi(r_S^\alpha - a) < 1 - \Phi(r_N^\alpha - a)$ $\forall a$ which actually means that the probability of an N-agent participating in A is greater than for an S-agent with the same real ability. With this it is easy to check that the denominator of the representativity index shown in equation (1.2) is

$$\int_{-\infty}^{\infty} \lambda[1 - \Phi(r_S^\alpha - y)]f(y) + (1-\lambda)[1 - \Phi(r_N^\alpha - y)]f(y)dy >$$
$$\int_{-\infty}^{\infty} \lambda[1 - \Phi(r_S^\alpha - y)]f(y) + (1-\lambda)[1 - \Phi(r_S^\alpha - y)]f(y)dy = \int_{-\infty}^{\infty} [1 - \Phi(r_S^\alpha - y)]f(y)dy$$

which proves that $1 > \frac{\int_{-\infty}^{\infty}[1-\Phi(r_S^\alpha - y)]f(y)dy}{\int_{-\infty}^{\infty} \lambda[1-\Phi(r_S^\alpha - y)]f(y)+(1-\lambda)[1-\Phi(r_N^\alpha - y)]f(y)dy} = \frac{\lambda_A^\alpha}{\lambda}$. The analogous analysis for $\alpha > 0$ proves that $\frac{\lambda_A^\alpha}{\lambda} > 1$ □

Proof of Proposition 1.2.5: Assume that the stereotype is negative. Since the stereotype is by assumption false, a^α does not represent the real minimum ability necessary to win a prize in tournament A. Instead, as S-agents are ex-ante more skillful than predicted by the stereotype the required minimum ability for obtaining W_A given the behavior rules (r_S^α, r_N^α) is $\hat{a} > a^\alpha$. Thus, the proportion of S-agents who win a prize W_A from social group S can be expressed as

$$\omega_A^\alpha = \frac{\lambda \int_{\hat{a}}^{\infty} [1 - \Phi(r_S^\alpha - y)]f(y)dy}{\int_{\hat{a}}^{\infty} \lambda[1 - \Phi(r_S^\alpha - y)]f(y) + (1-\lambda)[1 - \Phi(r_N^\alpha - y)]f(y)dy}$$

so the equivalent index to the one shown in equation (1.2) is $\frac{\omega_A^\alpha}{\lambda}$. Like the index for participation, this index reflects under/overrepresentation when $\frac{\omega_A^\alpha}{\lambda} \neq 1$. As in the proof of proposition 1.2.4 it is straightforward to see that

$$\int_{\hat{a}}^{\infty} [1 - \Phi(r_S^\alpha - y)]f(y)dy < \int_{\hat{a}}^{\infty} \lambda[1 - \Phi(r_S^\alpha - y)]f^S(y) + (1-\lambda)[1 - \Phi(r_N^\alpha - y)]f^N(y)dy$$

which comes directly from the fact that as $r_S^\alpha > r_N^\alpha$ so $1 - \Phi(r_S^\alpha - a) < 1 - \Phi(r_N^\alpha - a)$ $\forall a$. This implies that $\frac{\omega_A^\alpha}{\lambda} < 1$. If the stereotype is positive an analogous analysis takes place, but taking into account that $\hat{a} < a^\alpha$ and $r_S^\alpha < r_N^\alpha$. □

Proof of Proposition 1.2.6: Assume that the stereotype is false and negative. Then the actual distribution of signals for both social groups is the same. In particular, signals are distributed by a normal distribution with mean M^N and variance $\sigma_a^2 + \sigma_\mu^2$. Let the PDF of this distribution be $\Gamma(r)$. A negative stereotype, $\alpha < 0$, implies by corollary 1.2.3 $r_S^\alpha > r_N^\alpha$. Notice that except for a change in λ any comparative statics exercise will only affect the index $\frac{\lambda_A^\alpha}{\lambda}$ through a change in λ_A^α when the stereotype is false.

Therefore the proportion of participants in tournament A who come from social group S, λ_A^α, can be expressed as the mass of agents from social group S who observe signals greater than r_S^α as a proportion of the total mass of agents who enter tournament A, i.e.

$$\lambda_A^\alpha = \frac{\lambda \int_{r_S^\alpha}^{\infty} \Gamma(y)dy}{\lambda \int_{r_S^\alpha}^{\infty} \Gamma(y)dy + (1-\lambda)\int_{r_N^\alpha}^{\infty} \Gamma(y)dy} = \frac{\lambda \int_{r_S^\alpha}^{\infty} \Gamma(y)dy}{\int_{r_S^\alpha}^{\infty} \Gamma(y)dy + (1-\lambda)\int_{r_N^\alpha}^{r_S^\alpha} \Gamma(y)dy}$$

which after invoking corollary 1.2.3 proves to be

$$\frac{\lambda \int_{r_N^\alpha+k}^{\infty} \Gamma(y)dy}{\int_{r_N^\alpha+k}^{\infty} \Gamma(y)dy + (1-\lambda)\int_{r_N^\alpha}^{r_N^\alpha+k} \Gamma(y)dy}$$

with $k = r_S^\alpha - r_N^\alpha > 0$. Clearly this expression is lower than λ, which is what would be expected under perfect information. Taking the derivative of that expression with respect to r_N^α we find that the above expression is decreasing iff

$$\frac{\int_{r_N^\alpha+k}^{\infty} \Gamma(y)dy}{\Gamma(r_N^\alpha+k)} < \frac{\int_{r_N^\alpha}^{\infty} \Gamma(y)dy}{\Gamma(r_N^\alpha)}$$

which can be shown to hold by using the Abramowitz and Stegun (1964) approximation for cumulative normal distribution (algorithm 26.2.16).[4] Therefore it can be claimed that the relationship between the proportion of participants in tournament A who come from social group S and r_N^α is decreasing i.e.

$$\frac{\partial \lambda_A^\alpha}{\partial r_N^\alpha} < 0 \tag{A.7}$$

Thus, by observing that the direct effect of a change of δ_A on r_N^α is null (lemma A.1.2) it can be seen that

$$\frac{dr_N^\alpha}{d\delta_A} = \frac{\partial r_N^\alpha(a^\alpha)}{\partial a^\alpha}\frac{\partial a^\alpha}{\partial \delta_A} < 0$$

which establishes that $\frac{d\lambda_A^\alpha}{d\delta_A} = \frac{\partial \lambda_A^\alpha}{\partial r_N^\alpha}\frac{\partial r_N^\alpha}{\partial a^\alpha}\frac{\partial a^\alpha}{\partial \delta_A} > 0$ and so $d\left[\frac{\lambda_A^\alpha}{\lambda}\right]/d\delta_A > 0$ as claimed in point one of the proposition.

In addition, note that the marginal effect of an increase on $\frac{W_B}{W_A}$ on r_N^α is positive, so it can be concluded that $\frac{d\lambda_A^\alpha}{d(W_B/W_A)} = \frac{\partial \lambda_A^\alpha}{\partial r_N^\alpha}\frac{\partial r_N^\alpha}{\partial (W_B/W_A)} < 0$ and thus $d\left[\frac{\lambda_A^\alpha}{\lambda}\right]/d\left[\frac{W_B}{W_A}\right] < 0$ as claimed in point two of the proposition.

To account for the effect of the proportion of S-agents in the total population, notice that there is a direct effect and an indirect one such that

$$\frac{d(\frac{\lambda_A^\alpha}{\lambda})}{d\lambda} = \frac{\partial(\frac{\lambda_A^\alpha}{\lambda})}{\partial \lambda} + \frac{\partial a^\alpha}{\partial \lambda}\frac{\partial r_N^\alpha}{\partial a^\alpha}\frac{\partial(\frac{\lambda_A^\alpha}{\lambda})}{\partial r_N^\alpha} > 0$$

From equations (A.5), (A.6), and (A.7) it can be seen straightforwardly that the indirect effect is positive (an increase of λ decreases the real ability threshold a^α, which lowers the signal threshold r_N^α, which increases λ_A^α and ultimately the ratio λ_A^α/λ). It is easy to see from equation (1.2) that the direct effect is $\int_{-\infty}^{\infty}[\Phi(r_S^\alpha - y) - \Phi(r_N^\alpha - y)]f^N(y)dy > 0$.

[4] More precisely, using the Abramowitz and Stegun (1964) approximation the above holds iff $a_1 t(r_N^\alpha + k) + a_2[t(r_N^\alpha+k)]^2 + a_3[t(r_N^\alpha+k)]^3 < a_1 t(r_N^\alpha) + a_2[t(r_N^\alpha)]^2 + a_3[t(r_N^\alpha)]^3$ where a_1, a_2 and a_3 are parameters provided by the algorithm and $t : [0,\infty) \to (0,1]$ which is decreasing in its argument. The inequality presented above holds if the derivative of $a_1 t(x) + a_2[t(x)]^2 + a_3[t(x)]^3$ with respect to x is negative. Since $t(x) \in (0,1]$, $t'(x) < 0$, $a_1, a_3 > 0$, $a_2 < 0$ and $|a_1| > 2|a_2|$ it is immediately apparent that $\frac{\partial a_1 t(x) + a_2[t(x)]^2 + a_3[t(x)]^3}{\partial x} < 0$

Finally, to account for the effect of an increase in the signal informativeness, γ, it is easy to see that

$$\frac{d\lambda_A^\alpha}{d\gamma} = \frac{\partial \lambda_A^\alpha}{\partial k}\frac{\partial k}{\partial \gamma} > 0$$

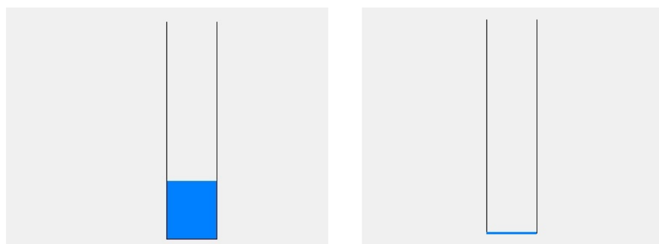
since $k = -\alpha M^N \left[\frac{(1-\gamma)}{\gamma}\right]$ together with $\alpha < 0$ makes $\frac{\partial k}{\partial \gamma} < 0$ and $\frac{\partial \lambda_A^\alpha}{\partial k} = \lambda(\lambda - 1)\Gamma(r_N^\alpha + k)\left[\int_{r_N^\alpha + k}^{\infty} \Gamma(y) dy\right] < 0$.

If the stereotype is positive, the same procedure should be carried out but taking into account that $k < 0$ and adapting the analysis accordingly. Notice that although the effects on the ratio λ_A^α/λ are of the opposite sign, the effects on the representativity problem, $|1 - [\lambda_A^\alpha/\lambda]|$, are of the same sign. $\square$

A.2 INSTRUCTIONS FOR THE EXPERIMENT

A.2.1 INSTRUCTIONS FOR STAGE 1

In this stage you will be asked to perform a task that involves two pictures. The first picture will show a glass containing some water (like the one in the left picture) and will be located at the top of the screen. This first picture will be displayed for 3 seconds. After that time, the picture of the full glass will disappear and, a second picture of an empty glass will appear at the bottom of your screen with a blue bar at its base (like the picture on the right).



Your task is to indicate on the second picture the water level from the first one. To do this, left click your mouse on the blue bar in the second picture and drag it until it matches the level of water shown in the first picture. Once you are satisfied with the location of the bar, press the "OK" button that will be displayed at the bottom right of your screen. This will take you automatically to a new picture of glass with water. The place where the bar is when you press "OK" determines the location of the bar for the purpose of computing your score.

You will perform this task 14 consecutive times with different initial pictures. The score you awarded for each repetition will be computed according to the following formula:

$$Score = 100 - 5 * distance(Real\_Level\_Water - Bar\_Position)$$

where 100 is the height of the glass and distance(Real_Level_Water-Bar_Position) is the absolute value of the distance between the water level in the first picture and the level where you place the bar in the second picture. If you do not understand the formula exactly, don't panic. The important thing is that the closer you leave the bar to the actual water level, the higher your score will be. NOTE: If the distance is greater than 20 your score will be negative.

PAYMENTS:
The computer will choose at random one of the 14 times that you perform the task and you will be paid according to the following rule

$$Payment = max(5 * Score/100, 0)$$

For example, if the distance is 0 in the randomly chosen repetition (i.e. you placed the bar at exactly the right water level), your score will be 100 (100=100-0) and you will get €5 (€5=€5*1). If the distance is 4 in the randomly chosen repetition your score will be 80 (80=100-5*4) and you will get €4 (€4=€5*0.8). If the distance between the water level and the place where you locate the bar in the randomly chosen repetition is greater than 20 your score will be negative and you will get €0.

A.2.2 Instructions for Stage 2 (Not Manipulated)

In this part of the experiment you will be asked to estimate your average score over the 14 repetitions of the task you performed previously.

This average represents your real ability in the task. If your estimation matches your actual average score you will be paid €1.5. Otherwise you will get €0.

Before you enter your estimation, we will provide you with certain information that may help you. Please press "OK" to continue.

(INFORMATION_SIGNAL) Now you can see on your screen the score chosen randomly by the computer from among the 14 times you have performed the task, i.e. the score that you see is your real score from one of the 14 times that you have completed the "glass task".

Please, press "OK" to continue.

(INFORMATION_OTHERS) Now you can see the distribution of the average scores have been distributed for the session in intervals of three. As you can see, this information is available in both chart and plot form.

Please, press "OK" to continue.

(SELF-ASSESSMENT) Now you can see at the bottom right of your screen a list of options. All you have to do is choose the group to which you believe you belong from among the 11 possibilities available. If you guess correctly you will get an extra €1.5. Once you have selected an option press "OK" to confirm your choice. You will then be asked another question that you must also answer. Once everyone has answered this additional question you will go on to stage 3 of the experiment.

A.2.3 Instructions for Stage 2 (Manipulated)

In this part of the experiment you will be asked to estimate your average score over the 14 repetitions of the task you performed previously.

This average represents your real ability in the task. If your estimation matches your actual average score you will be paid €1.5. Otherwise you will get €0.

Before you enter your estimation, we will provide you with certain information that may help you. Please press "OK" to continue.

(INFORMATION_FRAMING)
DESCRIPTION OF THE TASK: The task that you have just performed is closely linked to your ability to perceive the proportions of one object and transfer them to another. In particular, the previous task is based in your spatial vision capability and your subsequent handling of visual information. These abilities are essential for performing tasks in the field of engineering.

According to the description of the BACHELOR'S DEGREE IN ENGINEERING provided by the UPV/EHU one of the BASIC SKILLS required for engineers is:

"The capability for spatial vision and knowledge of graphic representation techniques, including both conventional methods of metric geometry and descriptive geometry, such as computer-aided design applications."

Please, press "OK" to continue.

(INFORMATION_SIGNAL) Now you can see on your screen the score chosen randomly by the computer from among the 14 times you have performed the task, i.e. the score that you see is your real score from one of the 14 times that you have completed the "glass task".

Please, press "OK" to continue.

(INFORMATION_OTHERS) Now you can see the distribution of the average scores have been distributed for the session in intervals of three. As you can see, this information is available in both chart and plot form.

Please, press "OK" to continue.

(SELF-ASSESSMENT) Now you can see at the bottom right of your screen a list of options. All you have to do is choose the group to which you believe you belong from among the 11 possibilities available. If you guess correctly you will get an extra €1.5. Once you have selected an option press "OK" to confirm your choice. You will then be asked another question that you must also answer. Once everyone has answered this additional question you will go on to stage 3 of the experiment.

A.2.4 Instructions for Stage 3 (Coordination)

In this part of the experiment, you will be presented with two options (option A and option B) under 9 different situations. Your task will be to choose between option A and option B for each of the 9 situations. Each of these two options will give you the chance to obtain different prizes. However, each option has a maximum number of prizes for those who choose that option. The maximum number of prizes awarded under each option varies from one situation to another but in all the situations the sum of the prizes for A and B is equal to the number of people in this room. That is, in all the situations the number of prizes from A and the number of prizes from B add up to —.

If more people choose one option than there are prizes for that option in the situation in question, prizes will be awarded only to those subjects who have chosen the option who showed the highest ability levels in the task on which you worked in the stage 1 (measured as the mean score over the 14 repetitions). For example, if an option offers 5 prizes and 7 people choose it, only the 5 with the highest ability levels will obtain the prize.

The prize in option B is always be €1.00. However, the prize available in option A varies in each of the 9 situations.

At the end of the experiment, the computer will choose randomly one of the 9 situations and you will be paid what you have won in that situation given your choice and the choice of the rest of the participants in that situation.

An example is shown below. Please press“OK” to go through the example.

EXAMPLE: Assume that there are 10 people participating (so the total number of prizes between options A and B is always 10). In this case, the next screen will show you a matrix similar to the one below where the 9 different situations are presented:

	Prize in A: 1.50€ Prize in B: 1.00€	Prize in A: 2.00€ Prize in B: 1.00€	Prize in A: 3.00€ Prize in B: 1.00€
Number of Prizes in A : 2 Number of Prizes in B : 8	A B	A B	A B
Number of Prizes in A : 5 Number of Prizes in B : 5	A B	A B	A B
Number of Prizes in A : 8 Number of Prizes in B : 2	A B	A B	A B

The first row of the table presents 3 different situations in which there are always 2 prizes in A and 8 in B. However, in each of the three situations in the first row the prize in option A is different.

Similarly, the first column shows three situations in which always the prize of option A is €1.50 and that in option B €1.00. However, the maximum number of prizes in options A and B change in each of the situations in column 1.
(CONTROL_QUESTION_COOR) To check whether you have understood how your decisions and those of others determine your payments in this stage, we will now run through an example, at the end of which you will be asked to answer some questions.

Once everyone has answered these questions correctly you will move on to the screen where you have to take the decisions.

Example: Assume that there are 4 participants in this session (yourself and 3 others). Your ability (average score) in the task was 65 and in the randomly chosen repetition you obtained a score of 90. The average scores of the other 3 participants were 60, 70, and 80 respectively. Assume that you are making your choice in a situation in which option A offers 1 prize and option B offers 3. Based on this information, answer the following questions.

The 4 questions that you must answer are shown below. Once you have answered them, press “OK”. If your answers are correct you will be moved on automatically to the decision-making screen as soon as all your partners have also answered the questions correctly.

1. If one of your partners had choose option A and the other two option B, will you win a prize if you choose option A?
    - Yes
    - No
    - **Depends**
2. If two of your partners choose option A and the other option B, will you win a prize if you choose option A?
    - Yes
    - **No**

- Depends

3. If you all choose option A, who will win the prize? (choose as many as you think are correct)
   - Me
   - The one whose ability level is 60
   - The one whose ability level is 70
   - **The one whose ability level is 80**
4. If you all choose option B, who will win the prize? (choose as many as you think are correct)
   - **Me**
   - The one whose ability level is 60
   - **The one whose ability level is 70**
   - **The one whose ability level is 80**

(CONTROL_QUESTION_OUT) To check whether you have understood how your decisions and those of others determine your payments in this stage, we will now run through an example, at the end of which you will be asked to answer some questions.

Once everyone has answered these questions correctly you will move on to the screen where you have to take the decisions.

Example: Assume that there are 4 participants in this session (yourself and 3 others). Your ability (average score) in the task was 65 and in the randomly chosen repetition you obtained a score of 90. The average scores of the other 3 participants were 60, 70, and 80 respectively. Assume that you are making your choice in a situation in which option A offers 1 prize. Based on this information, answer the following questions.

The 4 questions that you must answer are shown below. Once you have answered them, press "OK". If your answers are correct you will be moved on automatically to the decision-making screen as soon as all your partners have also answered the questions correctly.

1. If one of your partners had choose option A and the other two option B, will you win a prize if you choose option A?
   - Yes
   - No
   - **Depends**
2. If two of your partners choose option A and the other option B, will you win a prize if you choose option A?
   - Yes
   - **No**
   - Depends
3. If you all choose option A, who will win the prize? (choose as many as you think are correct)

- Me
- The one whose ability level is 60
- The one whose ability level is 70
- **The one whose ability level is 80**

4. If you all choose option B, who will win the prize? (choose as many as you think are correct)

   - **Me**
   - **The one whose ability level is 60**
   - **The one whose ability level is 70**
   - **The one whose ability level is 80**

A.2.5 Instructions for Stage 3 (Belief elicitation)

Next you will be asked 5 questions regarding this session. At the end of the experiment the computer will chose one of these questions at random and you will be paid €1.5 if your answer is correct according to the data collected during the session and €0 otherwise.

Question 1: Consider the situation above in which option A offered 16 prizes of €2.00 each and option B 17 prizes of €1.00 each. In this situation you have chosen option —.
Given your choice and your partners' choices in this situation, in what range do you think the minimum ability required to win tournament A's prize lies?
Note: If you think that in this situation there are fewer participants than prizes in option A, you should answer "<70"

Available options: <70, 70-73, 73-76, 76-79, 79-82, 82-85, 85-88, 88-91, 91-94, 94-97 and 97-100

Question 2: Consider the situation above in which option A offered 16 prizes of €2.00 each and option B — prizes of €1.00 each. In this situation you have chosen option —.
Given your choice and your partners' choices in this situation, in what range do you think the minimum ability required to win tournament A's prize lies?
Note: If you think that in this situation there are fewer participants than prizes in option A, you should answer "<70"

Available options: <70, 70-73, 73-76, 76-79, 79-82, 82-85, 85-88, 88-91, 91-94, 94-97 and 97-100

Question 3: In this session, who do you think has performed better on average in the initial task?

Available options: Men/No Differences/Women

Question 4: In this session the average ability across all the — participants has been —. In what range do you think men's average ability level in the initial task lies?

Available options: <70, 70-73, 73-76, 76-79, 79-82, 82-85, 85-88, 88-91, 91-94, 94-97 and 97-100

Question 5: In this session the average ability across all the — participants has been —. In what range do you think women's average ability level in the initial task lies?

Available options: <70, 70-73, 73-76, 76-79, 79-82, 82-85, 85-88, 88-91, 91-94, 94-97 and 97-100

A.2.6 Instructions for Stage 4 (Ability estimation, single sex)

This part of the experiment is very similar to stage 3. The difference is that now you will interact only with persons of your own sex

On the next screen you will be asked again to estimate your ability level in the initial task (average over the 14 repetitions). To that end, you will be shown the distribution of abilities of all the participants of your own gender. That is, if you are a man all the other participants shown will also be men and if you are a woman they will all be women.

If your estimation matches your real ability you will get €1.00. Otherwise you will get €0.

A.2.7 Instructions for Stage 4 (Coordination, single sex)

In this part of the experiment you will again be presented with two options (A and B) under 9 different situations. The decision matrix that you will see will be similar to the one in stage 3A.

The difference from stage 3A is that now you will only interact with participants of your own gender, i.e. if you are a man all the other participants will also be men and if you are a woman they will be women.

In this session there are — participants of your own gender. Therefore, the total number of prizes between option A and option B will in all cases be —.

As in part 3A, the computer will randomly pick one of the situations and you will be paid according to your choice and the choices of those with whom you interact.

A.2.8 Instructions Stage 5 (Risk Elicitation)

On the next screen you will be presented with 8 different options, each of which offers two different quantities that you can win by choosing that option. In all the options, each outcome has a probability of 50%, i.e., the result of choosing an option depends exclusively on luck. At the end of the experiment the computer will randomly pick one result from the option you have chosen and you will be paid accordingly.

Below this text you will find the 8 available options. To see in more detail how to read this table, consider option 5. In this option the possible results are €0.7 and €2.7. Both are equally likely, which means that the computer will choose €0.7 as the payment on one of every 2 occasions and €2.7 the other.

You must choose one of the 8 possible options. To that end, an empty box will appear where you must enter the number of the option (from 1 to 8) that you want to choose.

	Probability 50%	Probability 50%
1	€1.5	€1.5
2	€1.3	€1.8
3	€1.1	€2.1
4	€0.9	€2.4
5	€0.7	€2.7
6	€0.6	€2.8
7	€0.4	€2.9
8	€0	€3

A.3 SUPPLEMENTARY ANALYSES

TABLE A.1– ALTERNATIVE ANALYSES FOR RELATIVE ASSESSMENT

Sample:	*All*	*All+Inconsistent*	*Not Outliers*	*Not Manipulated*	*Manipulated*
	(1)	(2)	(3)	(4)	(5)
Signal	0.966***	0.958***	2.241***	0.965*	1.070***
	(0.195)	(0.216)	(0.335)	(0.556)	(0.240)
Female	0.673*	0.0337	0.00431	-0.0248	0.113
	(0.340)	(0.0421)	(0.0409)	(0.0537)	(0.0699)
Rel.Maleness	0.317***	0.303***	0.202*	0.259**	0.550**
	(0.0994)	(0.113)	(0.105)	(0.125)	(0.228)
Rel.Maleness∗Female	-0.550***	-0.535***	-0.320**	-0.593***	-0.728***
	(0.141)	(0.151)	(0.145)	(0.208)	(0.256)
Taste Comp.	0.0530**	0.0449**	0.0562***	0.0146	0.0836***
	(0.0212)	(0.0223)	(0.0193)	(0.0313)	(0.0278)
Risk Pref.	0.00832	0.00356	0.0108	0.00805	-0.00975
	(0.0116)	(0.0106)	(0.0106)	(0.0120)	(0.0177)
Age	0.0116**	0.0113**	0.0162**	0.00558	0.0146**
	(0.00488)	(0.00534)	(0.00716)	(0.00713)	(0.00664)
Difficulty	-0.0574	-0.0554	-0.0289	-0.101**	0.0134
	(0.0361)	(0.0335)	(0.0367)	(0.0400)	(0.0480)
Extroversion	0.0202				
	(0.0204)				
Agreeableness	0.0294				
	(0.0230)				
Conscientiousness	0.00780				
	(0.0159)				
Emotional Stability	0.00397				
	(0.0158)				
Openness	-0.0207				
	(0.0242)				
Social Risk	0.150**				
	(0.0596)				
(Social Risk)∗Female	-0.189**				
	(0.0941)				
Constant	-1.235***	-0.497	-1.871***	-0.141	-1.027**
	(0.402)	(0.346)	(0.352)	(0.652)	(0.415)
Session FE	YES	YES	YES	YES	YES
Observations	120	120	96	62	58
R-squared	0.530	0.489	0.570	0.438	0.592

Notes: Inconsistent subjects are those whose state perception and maleness are not compatible. *Not Outliers* consider consistent agents between the top 10% and the bottom 10% of the signal distribution. *Not Manipulated* subjects are those participating in sessions 1 and 2. *Manipulated* are those participating in sessions 3 and 4. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE A.2– OLS REGRESSION FOR CONFIDENCE IN ESTIMATION

	Conf.Est. (1)	*Conf.Est.* (2)
Signal	2.241***	0.924
	(0.604)	(0.695)
Relative Assessment		1.244***
		(0.403)
Female	-0.351*	-0.358*
	(0.185)	(0.183)
Rel.Maleness	0.377	-0.0468
	(0.466)	(0.420)
Rel.Maleness∗Female	-0.871	-0.175
	(0.599)	(0.532)
Constant	2.876***	3.245***
	(0.543)	(0.485)
Session FE	YES	YES
Observations	120	120
R-squared	0.158	0.222

Notes: Conf.Est consists of a 7 level scale with higher values indicating greater confidence in estimation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A.3– ALTERNATIVE ANALYSES FOR THE PROBABILITY OF ENTERING TOURNAMENT A

Sample:	*All*	*Marginal*	*All+Inconsistent*	*Not Manipulated*	*Manipulated*	*All*
Method:	*Probit*	*Probit*	*Probit*	*Probit*	*Probit*	*OLS*
	(1)	(2)	(3)	(4)	(5)	(6)
Signal	0.404**	0.972	0.467***	0.591*	0.463**	0.525**
	(0.165)	(0.720)	(0.177)	(0.308)	(0.185)	(0.211)
Female	0.354**	0.404**	-0.0198	-0.0318	0.0185	-0.0164
	(0.155)	(0.170)	(0.0409)	(0.0535)	(0.0655)	(0.0426)
Rel.Maleness	0.233*	0.458**	0.221	0.0963	0.656**	0.220
	(0.140)	(0.191)	(0.144)	(0.152)	(0.271)	(0.143)
Rel.Maleness∗Female	-0.391**	-0.984***	-0.368**	-0.287*	-0.765**	-0.378**
	(0.169)	(0.231)	(0.174)	(0.172)	(0.302)	(0.177)
Risk Pref	0.0222**	0.00818	0.0191*	0.0207	0.0155	0.0198*
	(0.0109)	(0.0144)	(0.0106)	(0.0162)	(0.0172)	(0.0110)
Min.Ab.Win	0.0150	0.0303*	0.0184	0.0265	0.00417	0.0209
	(0.0131)	(0.0170)	(0.0128)	(0.0224)	(0.0126)	(0.0142)
Taste Comp.	0.0236	0.00375				
	(0.0201)	(0.0302)				
Age	0.00365	-0.00802				
	(0.00603)	(0.00802)				
Difficulty	-0.0409	-0.157***				
	(0.0315)	(0.0389)				
Extroversion	-0.0213	-0.0555**				
	(0.0185)	(0.0228)				
Agreeableness	0.0141	0.00274				
	(0.0268)	(0.0319)				
Conscientiousness	0.0305*	0.0357*				
	(0.0156)	(0.0196)				
Emotional Stability	-0.0120	-0.0335				
	(0.0177)	(0.0213)				
Openness	-0.00712	-0.0559*				
	(0.0237)	(0.0297)				
Social Risk	0.141**	0.185**				
	(0.0644)	(0.0796)				
(Social Risk)∗Female	-0.132	-0.140				
	(0.0866)	(0.103)				
Constant						-0.394*
						(0.211)
Tournament Controls	YES	YES	YES	YES	YES	YES
Session FE	YES	YES	YES	YES	YES	YES
Number of Clusters	120	97	140	62	58	120
Observations	1,080	342	1,080	558	522	1,080

Notes: Inconsistent subjects are those whose state perception and maleness are not compatible. *Not Manipulated* subjects are those participating in sessions 1 and 2. *Manipulated* are those participating in sessions 3 and 4. Marginal effects are reported. Clustered standard errors at subject level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE A.4– DETERMINANTS OF THE OPTIMAL SIGNAL-CONTINGENT STRATEGY

Sample:	*All* (1)	*Marginal* (2)
Signal	1.275***	0.989
	(0.251)	(1.145)
Female	0.0130	-0.00361
	(0.0639)	(0.0963)
Rel.Maleness	0.0299	-0.476*
	(0.176)	(0.247)
Rel.Maleness∗Female	-0.200	0.0546
	(0.232)	(0.367)
Tournament Controls	YES	YES
Session FE	YES	YES
Number of Clusters	120	97
Observations	1,080	342

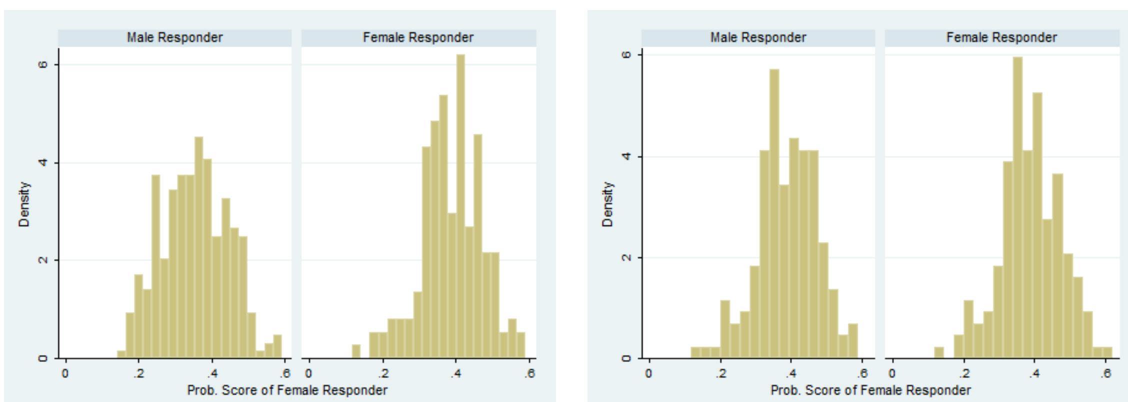
Notes: Marginal effects for the probability of obtaining W_A unconditional on the behavior of others (probability that tournament A represents a dominant strategy). Tournament controls include the allocation of prizes. Clustered standard errors at subject level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix B

Women ask for less (only from men)

B.1 Robustness and Additional Results

FIGURE B.1– PROBABILITY OF CHOOSING A FEMALE RESPONDER



(a) (b)

Notes: (a) Histogram for Probability Score of Proposers choosing a Female Responder in the whole data. (b) Histogram for Probability Score of Proposers choosing a Female Responder in the Matched Sample

TABLE B.1– ALTERNATIVE SPECIFICATIONS AND ADDITIONAL RESULTS FOR THE CHOICE OF THE SEX OF THE RESPONDER

	LPM Prob. Male Resp. (1)	Logit Prob. Male Resp. (2)	RE Probit Prob. Male Resp. (3)	Probit Prob. Male Resp. (4)	Probit Prob. Male Resp. (5)
Male Proposer	0.142**	0.138***	0.145**	0.154**	0.0552
	(0.0555)	(0.0525)	(0.0605)	(0.0647)	(0.153)
Age Proposer	0.00252	0.00257	0.00166	0.00221	0.00212
	(0.00356)	(0.00339)	(0.00372)	(0.00337)	(0.00338)
Student Proposer	0.117	0.110	0.104	0.106	0.106
	(0.0819)	(0.0739)	(0.0917)	(0.0750)	(0.0749)
Retired Proposer	-0.119	-0.135	-0.0934	-0.133	-0.128
	(0.192)	(0.192)	(0.285)	(0.186)	(0.187)
Unemployed Proposer	0.146*	0.145**	0.138	0.140*	0.134*
	(0.0843)	(0.0700)	(0.0905)	(0.0724)	(0.0741)
Low-Occupation Proposer	0.00276	0.00403	0.0103	0.00423	0.00322
	(0.0726)	(0.0677)	(0.0794)	(0.0689)	(0.0684)
Perc. Attractiveness Proposer	0.0534*	0.0551**	0.0537*	0.0539**	0.0433
	(0.0277)	(0.0272)	(0.0306)	(0.0265)	(0.0285)
Perc. Status Proposer	-0.148	-0.157	-0.163	-0.149	-0.152
	(0.126)	(0.125)	(0.151)	(0.124)	(0.124)
(Pie-Mean) by Stage	-0.000435	-0.000434	0.0000334	-0.000277	-0.000274
	(0.000779)	(0.000773)	(0.00105)	(0.000773)	(0.000786)
Remaining Time	0.000507	0.000491	0.000502	0.000534	0.000543
	(0.000486)	(0.000482)	(0.000504)	(0.000473)	(0.000467)
Male Question	0.290***	0.308***	0.324***	0.343***	0.301***
	(0.0539)	(0.0535)	(0.0550)	(0.0586)	(0.0521)
Female Question	-0.303***	-0.282***	-0.314***	-0.275***	-0.289***
	(0.0638)	(0.0623)	(0.0720)	(0.0965)	(0.0639)
Male Proposer*Male Question				-0.154	
				(0.157)	
Male Proposer*Female Question				-0.0193	
				(0.120)	
Male Proposer*Perc. Attract. Proposer					0.0249
					(0.0419)
Constant	0.274				
	(0.184)				
Stage FE	YES	YES	YES	YES	YES
Observations	428	428	428	428	428

Notes: The dependent variable takes the value 1 if the selected responder is male and 0 otherwise. The table shows the marginal effect values of the coefficients using the probit model. The control variables are described in the notes of Table 2.1. Clustered standard errors at the proposer level (columns 1, 2, 4 and 5) and bootstrapped standard errors (column 3), in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE B.2– ARE MALE RESPONDERS MORE LIKELY TO KNOW THE CORRECT ANSWER?

	Prob. Correct Answer (1)	Prob. Correct Answer (2)
Male Responder	-0.0648*	-0.0482
	(0.0343)	(0.0362)
Male Question		-0.123**
		(0.0612)
Female Question		-0.0353
		(0.0507)
Observations	427	427

Notes: The dependent variable takes the value 1 if the responder provided the correct answer and 0 otherwise. The table shows the marginal effect values of the coefficients using the probit model. The control variables are described in the notes of Table 1. Clustered standard errors at the proposer level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE B.3– ALTERNATIVE SPECIFICATIONS FOR GENDER DIFFERENCES AND GENDER INTERACTION EFFECTS IN BARGAINING OUTCOMES

	LPM Prob. No Agreement (1)	Logit Prob. No Agreement (2)	RE Probit Prob. No Agreement (3)	RE Proposer's Outcome (4)	OLS No. of Rounds (5)	RE No. of Rounds (6)
MF	-0.0624	-0.0521	-0.0440	10.78***	-0.209	-0.183
	(0.0471)	(0.0407)	(0.129)	(3.822)	(0.375)	(0.374)
FM	-0.0532	-0.0372	-0.0422	2.995	-0.0952	-0.0851
	(0.0499)	(0.0403)	(0.0462)	(3.689)	(0.361)	(0.361)
MM	0.00145	0.0117	0.0183	-0.523	-0.520	-0.510
	(0.0466)	(0.0392)	(0.0497)	(5.037)	(0.320)	(0.322)
Constant	0.413***			54.92***	2.231**	2.198**
	(0.137)			(19.21)	(1.063)	(1.073)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	428	428	428	377	377	377
R-squared	0.117				0.160	
Number of Proposers			131	131		131
H_0: MF=FM	0.85	0.74	0.99	0.05	0.77	0.80
H_0: MF=MM	0.19	0.18	0.64	0.02	0.35	0.31
H_0: FM=MM	0.20	0.21	0.22	0.50	0.15	0.15

Notes: The dependent variables refer to: *the Prob.of No Agreement*, which takes the value of 1 when the bargaining partners do not reach an agreement and 0 otherwise (columns 1 to 3); *Proposer's Outcome*, which summarizes the outcome in euro obtained by the proposer from the bargaining (column 4); and *No. of Rounds* describes the duration of the bargaining process (columns 5 to 6). All controls, as shown in Table 3 in the paper, are included. At the bottom, *p*-values for the hypothesis testing are shown where pairwise comparisons are made for the different gender combinations, where FF refers to bargaining between women, MF refers to the bargaining between a male proposer and a female responder, FM refers to the bargaining between a female proposer and a male responder and MM refers to the bargaining between men. Clustered standard errors at the proposer level (columns 1 ,2 ,4 ,5 and 6) and bootstrapped standard errors (column 3), in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE B.4– GENDER INTERACTION EFFECTS IN PROPOSERS' AND RESPONDERS' BARGAINING STRATEGIES

	Proposers				Responders			
	Silent	Increasing Offers	Decreasing Offers	Maintaining	Silent	Increasing Demands	Decreasing Demands	Maintaining
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MF	0.0220	0.0789**	-0.0172**	-0.0610	0.0410	0.0106	-0.0720	0.0959
	(0.0279)	(0.0393)	(0.00845)	(0.0387)	(0.0702)	(0.0580)	(0.0902)	(0.0826)
FM	0.00388	0.0259	-0.0336**	0.00381	0.0752	-0.0607	0.0430	0.0278
	(0.0170)	(0.0389)	(0.0142)	(0.0372)	(0.0635)	(0.0467)	(0.0815)	(0.0730)
MM	0.0119	0.116***	-0.0228**	-0.0847**	0.0352	-0.0538	0.0223	0.0659
	(0.0214)	(0.0372)	(0.0103)	(0.0393)	(0.0612)	(0.0431)	(0.0797)	(0.0812)
Age Proposer	0.000779	0.00109	5.46e-05	-0.00120				
	(0.000735)	(0.00189)	(0.000533)	(0.00177)				
Student Proposer	-0.0140	0.0656*	-0.0110	-0.0527				
	(0.0149)	(0.0388)	(0.00899)	(0.0408)				
Retired Proposer	-0.0266	-0.0436	-0.0153***	0.0782				
	(0.0299)	(0.105)	(0.00437)	(0.126)				
Unemployed Proposer	-0.0101	0.00654	-0.000952	0.000285				
	(0.0342)	(0.0674)	(0.0169)	(0.0874)				
Low-Occup. Proposer	-0.00647	-0.0727*	0.00308	0.0667*				
	(0.0146)	(0.0418)	(0.0119)	(0.0403)				
Perc. Age Proposer					0.00322	-0.0232	0.00414	0.0166
					(0.0202)	(0.0214)	(0.0250)	(0.0252)
Perc. Status Proposer	-0.0483	0.0130	0.00650	0.00499	0.0266	-0.0822	0.128	-0.0600
	(0.0317)	(0.0662)	(0.0259)	(0.0649)	(0.0883)	(0.0859)	(0.136)	(0.112)
Perc. Attractiveness Proposer	0.00592	-0.00121	-0.00329	0.00350	0.0141	0.00453	-0.0158	0.0170
	(0.00577)	(0.0118)	(0.00463)	(0.0120)	(0.0183)	(0.0171)	(0.0255)	(0.0216)
Perc. Age Responder	-0.00177	0.00610	-0.00785	8.83e-05	0.00421	0.0113	-0.0102	-0.00512
	(0.00533)	(0.0177)	(0.00559)	(0.0175)	(0.0214)	(0.0180)	(0.0320)	(0.0300)
Perc. Status Responder	0.0171	-0.108*	-0.00692	0.0917	-0.135*	0.0812	0.0493	-0.134
	(0.0236)	(0.0627)	(0.0196)	(0.0605)	(0.0805)	(0.0753)	(0.109)	(0.102)
Perc. Attractiveness Responder	-0.00725	0.0153	-0.00941*	-0.00317	0.0242	-0.0347	0.00394	0.0297
	(0.00663)	(0.0175)	(0.00531)	(0.0172)	(0.0253)	(0.0222)	(0.0328)	(0.0328)
(Pie-Mean) by Stage	0.000555	0.000295	-4.86e-05	-0.000334	0.000613	-0.000844	0.00166	-0.000752
	(0.000348)	(0.000570)	(0.000153)	(0.000518)	(0.000858)	(0.000514)	(0.00106)	(0.000879)
Remaining Time	0.000332**	5.31e-05	0.000132	-0.000132	-0.000441	-0.000370	0.000504	-0.000237
	(0.000169)	(0.000344)	(0.000110)	(0.000320)	(0.000486)	(0.000393)	(0.000630)	(0.000511)
No. Of Rounds	0.00312	0.0336***	0.00322	-0.0360***	0.0558***	0.0482***	-0.0287**	-0.0450***
	(0.00345)	(0.0127)	(0.00241)	(0.0117)	(0.0151)	(0.00960)	(0.0129)	(0.0153)
Round	-0.00338	-0.0864***	0.00220	0.0842***	-0.0945***	-0.0395***	-0.0104	0.0730***
	(0.00305)	(0.0122)	(0.00216)	(0.0128)	(0.0133)	(0.0135)	(0.0146)	(0.0148)
Stage FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,349	911	911	911	1,174	551	551	551
Number of Responders	377	308	308	308	354	230	230	230
H_0: MF=FM	0.43	0.19	0.15	0.09	0.64	0.24	0.20	0.42
H_0: MF=MM	0.63	0.30	0.43	0.51	0.93	0.23	0.24	0.71
H_0: FM=MM	0.64	0.02	0.35	0.03	0.49	0.88	0.78	0.60

Notes: Dependent variables refer to the different types of strategies used in the bargaining process. *Silent* takes the value of 1 when the proposer or responder remains silent. *Increasing* takes the value of 1 when the proposer or responder increases the offer or demand from one round to the next. *Decreasing* takes the value of 1 when the proposer or responder decreases the offer or demand from one round to the next. Finally, *Maintain* takes the value of 1 when the proposer or responder maintains the same offer or demand from one round to the next. All columns show the marginal effect values of the coefficients using the probit random effects model. At the bottom, *p*-values for the hypothesis testing are shown where pairwise comparisons are made for the different gender combinations, where FF refers to bargaining between women, MF refers to the bargaining between a male proposer and a female responder, FM refers to the bargaining between a female proposer and a male responder and MM refers to the bargaining between men. Bootstrapped standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE B.5– ALTERNATIVE SPECIFICATIONS WITH COLLAPSED DATA FOR GENDER INTERACTION EFFECTS IN OFFERS, DEMANDS AND PROBABILITIES OF ACCEPTANCE

	OLS or RE Offers (round>1) (1)	Probit Prob. Responder Accepts (round>1) (2)	RE Probit Prob. Responder Accepts (round>1) (3)	OLS or RE Demands (4)	Probit Prob. Proposer Accepts (5)	RE Probit Prob. Proposer Accepts (6)
MF	-0.673	-0.127	-0.127	-42.71***	0.0446	0.044
	(3.090)	(0.0960)	(0.091)	(14.18)	(0.0783)	(0.0963)
FM	-2.959	0.0480	0.048	-6.507	-0.0565	-0.053
	(2.445)	(0.0887)	(0.0892)	(12.79)	(0.0737)	(0.085)
MM	-1.815	0.109	0.109	-2.557	-0.120*	-0.120
	(3.375)	(0.0807)	(0.0963)	(12.48)	(0.0685)	(0.0874)
Constant	22.21**			87.75		
	(11.02)			(58.13)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	316	316	316	321	321	321
H_0: MF=FM	0.50	0.09	0.14	0.02	0.24	0.34
H_0: MF=MM	0.70	0.00	0.01	0.00	0.01	0.04
H_0: FM=MM	0.76	0.45	0.41	0.79	0.35	0.43

Notes: The dependent variable *Offers* refer to the offers in euro made by the proposer (column 1); *Prob. Responder Accepts* takes the value of 1 when the responder accepts the offer made by the proposer and 0 otherwise (columns 2 and 3); *Demands* refer to the demands in euro made by the responder (columns 4); and*Prob. Proposer Accepts* takes the value of 1 when the proposer accepts the demand made by the responder and 0 otherwise (columns 5 and 6). At the bottom, *p*-values for the hypothesis testing are shown where pairwise comparisons are made for the different gender combinations, where FF refers to bargaining between women, MF refers to the bargaining between a male proposer and a female responder, FM refers to the bargaining between a female proposer and a male responder and MM refers to the bargaining between men. Clustered standard errors at the proposer level (columns 1,2,4,and 5) and bootstrap standard errors (column 3 and 6), in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix C

The Role of Gender and Asymmetries in Alternating-Offer Bargains

C.1 Instructions

C.1.1 General Instructions

THANK YOU FOR PARTICIPATING IN THE EXPERIMENT!

We are going to start the experiment. From now on you are not permitted to talk, look at what other participants are doing or to leave your seat. Please switch off your mobile phone. If you have any questions or you need help, raise your hand and one of the researchers will assist you. If you do not follow these instructions YOU WILL BE ASKED TO LEAVE THE EXPERIMENT AND YOU WILL NOT GET ANY PAYMENT. Thank you.

The University of the Basque Country has provided the funds for this experiment. How much you can earn depends on your decisions, on the decisions of other participants, and on luck.

Stages of the experiment and tasks: The experiment consists of 3 stages:

In the first stage you will be shown matrices with "0"s and "1"s for 5 minutes. Your task is to count the number of "1"s in each matrix. The number of correct answers that you provide will determine your productivity which will be relevant for the next part of the experiment.

In the second stage of the experiment the computer will randomly match you with a partner and your task will consist of dividing an amount of money through a bargaining process. This quantity depends on your own productivity and the productivity of the participant with whom you are matched. You will have 3 minutes for each negotiation. There will be 10 bargaining rounds in which you will be matched with a different participant each time.

In the third stage you will be presented with three short tasks in which you can earn more money.

Earnings:

You are guaranteed 3 Euro. In addition, once the experiment has concluded, the computer will choose two bargaining rounds randomly from in the second stage of the experiment and you will be paid the amount that you earned in each of them. Finally, in the third stage you can earn extra money for each of the three short tasks, so at the end of the experiment your final earnings will be the sum of the 3 Euro that you get for participating, your earnings in the two bargaining rounds randomly selected, and your earnings in each of the short tasks from stage 3. Your earnings will be paid in cash privately at the end of the experiment.

We will now start the experiment. At the beginning of each stage, detailed information about the task, about the decisions, and about earning is provided.

C.1.2 Real Effort Task

In this stage, you will be shown matrices with "0"s and "1"s, similar to the ones displayed below for 5 minutes.

Your task is to count the number of "1"s in each matrix. The size of the matrices will vary. Once you have entered an answer for a matrix and clicked on "OK" the next matrix will appear. All participants will see the same matrices in the same order. There is a maximum of 60 matrices.

Example 1: 8x8 Matrix, Solution = 30

1	1	0	1	0	1	1	0
1	0	1	0	1	0	1	1
0	0	1	0	1	0	1	1
0	0	0	0	0	1	1	0
0	0	0	0	1	0	1	0
1	0	0	0	0	0	1	1
0	0	0	1	1	0	1	0
0	0	1	1	1	1	1	1

Example 2: 6x6 Matrix, Solution = 16

1	0	0	0	1	1
1	0	0	0	1	0
0	0	0	1	0	0
1	0	0	1	0	0
1	1	0	0	1	1
1	1	1	1	0	0

The number of correct answers that you provide will determine your productivity. The higher your productivity is, the higher the average amount of money that you have to divide in the next stage will be.

C.1.3 Bargaining Stage: Symmetric

In this stage you will be matched randomly with another participant and your task consists of dividing an amount of money through a bargaining process. This amount may be €5, €10 or €15.

HOW IS THE AMOUNT OF MONEY TO BE DIVIDED COMPUTED?

1. The number of correct answers in the first stage determines the productivity of each participant as follows:

- Bottom third: Those participants with a fewest number of correct answers will be assigned a productivity of €5
- Middle third: Those participants with an intermediate number of correct answers will be assigned a productivity of €10
- Top third: Those participants with the highest number of correct answers will be assigned a productivity of €15

2. In each round you will be randomly matched with another participant and the amount to be divided will be:

   - YOUR PRODUCTIVITY with a 50% chance
   - THE PRODUCTIVITY OF THE PARTICIPANT YOU ARE MATCHED WITH with a 50% chance

For example, if your productivity is €5 and the productivity of the other participant is €15, there is a 50% likelihood that the amount to be divided will be €5 and a 50% chance that will be €15. Finally, if you and the participant with whom you are matched have the same productivity then the amount to be divided will be 5, 10 or 15 Euro, whichever amount you both have.

WHAT DECISIONS CAN BE TAKEN DURING A BARGAINING PROCESS?

Before starting, you will be told whether you are participant A or participant B in each matching. During the bargaining you will have to decide HOW MUCH MONEY PARTICIPANT B WILL GET, so if you are participant A you will make offers to participant B and if you are participant B you will make demands of participant A.

The negotiation works in the following way:

- Participant A starts the negotiation with an opening offer, deciding how much money he/she wants to offer to participant B.
- Participant B can accept or reject that offer. If the offer is accepted participant B gets the amount offered and participant A gets the rest of the pie (i.e. the full amount to be divided minus the amount offered to participant B).
- If the offer is rejected the bargaining continues and it is the turn of participant B to make a demand of participant A, deciding how much money he/she wants to get.
- Participant A can accept or reject that demand. If the demand is accepted participant B gets the amount demanded and participant A gets the rest of the pie (i.e. the full amount to be divided minus the amount demanded by participant B).
- If the demand is rejected the bargaining continues and it is the turn of participant A to make a further offer to participant B. And so on.

Offers and demands must be multiples of €0.1 (10 cents). You have 3 minutes to reach a deal. If you fail to reach a deal within that time, both participants get €0.

There will be 10 different bargaining rounds, in which you will be matched with a different participant each time. During each negotiation you will be informed of how much money you have to divide, of whether you are participant A or participant B, of how much of the 3-minute time limit is left for the 3 minutes, and of the complete bargaining record: offers made by A, demands made by B and whether they have been accepted or rejected.

For payment, at the end of the experiment the computer will choose two bargaining rounds randomly, one from rounds 1–5 and another from rounds 6–10, and you will be paid the amount of the deal you reach in those rounds or €0 if you fail to reach a deal.

C.1.4 Bargaining Stage: Empowerment

[. . .] If you fail to reach a deal within that time,
participant A is guaranteed an amount of money, while participant B gets €0. The amount of money that participant A gets is randomly chosen between 50% and 85% of the amount to be divided.
That is, if not deal is reached within the 3 minutes, participant A gets:

- Between €2.5 and €4.25 if the amount to be divided is €5
- Between €5 and €8.5 if the amount to be divided is €10
- Between €7.5 and €12.75 if the amount to be divided is €15

The exact amount is randomly chosen by the computer once the negotiation has finished.

[. . .]For payment, at the end of the experiment the computer will choose two bargaining rounds randomly, one from rounds 1–5 and another from rounds 6–10, and you will be paid the amount of the deal you reach in those rounds or
a positive amount if you are participant A and €0 if you are participant B if you fail to reach a deal.

C.1.5 Bargaining Stage: Entitlement

[. . .]and of the complete bargaining record: offers made by A, demands made by B and whether they have been accepted or rejected.
In addition, you will be told your productivity and the productivity of the participant with whom you are matched, so you can learn whether the amount to be divided is that of your productivity or that of the participant with whom you are matched.

C.1.6 Bargaining Stage: Information

[. . .] There will be 10 different bargaining rounds, in which you will be matched with a different participant each time.
During each bargaining process only participant A observes the amount to be divided: Participant B only knows that it may be 5, 10 or 15 Euro, but is unaware of the exact amount. Participant A cannot accept demands that are higher than the amount of money to be divided.

C.1.7 Elicitation Tasks

This stage of the experiment consists of three short tasks, with which you can earn extra money. The first is to answer four questions about this session. In the second and third you must choose between different options.

As you progress through this third stage of the experiment, we will provide you with more detailed instructions about each task.

TASK I:

Next you will be asked 4 questions about this session. At the end of the experiment the computer will choose one of them randomly and you will be paid €1 if your answer is correct according to the data that we have gathered during the session and €0 otherwise.

QUESTION 1: If all participants in this session are sorted from the lowest to the highest number of correct answers in stage 1 (counting "1"s), and all subjects are divided into 4 segments of equal sizes such that the participants with highest scores are in the first segment, the next in the second, the next in the third, and those with lowest scores in the fourth segment, which segment do you think you will be in?

Options: 1^{st} segment/2^{nd} segment/3^{rd} segment/4^{th} segment

QUESTION 2: On average, who do you think has performed better in the task in stage 1 (counting "1"s)?

Options: Men/No difference/Women

QUESTION 3: In each negotiation a participant could get between 0% and 100% of the amount of money to be divided. If all participants in this session are sorted from the lowest to the highest share of money obtained on average over the 10 rounds, and all the subjects are divided in 4 segments of equal sizes such that the participants who obtained the highest average share of the money are in the first segment, the next in the second, the next in the third, and those with lowest shares in the fourth segment, which segment do you think you will be in?

Options: 1^{st} segment/2^{nd} segment/3^{rd} segment/4^{th} segment

QUESTION 4: On average, who do you think has obtained a greater share of money during the negotiations?

Options: Men/No difference/Women

TASK II:

On the next screen you will be given 8 different options, each of which offers two different amounts that you can win by choosing that option. In all the options, each outcome has a probability of 50%, i.e. the outcome depends exclusively on luck. At the end of the experiment the computer will randomly pick one result from the option that you have chosen and you will be paid accordingly.

Below this text you will find the 8 available options. For more details of how to read this table, consider option 5. In this option the possible results are €0.7 and €2.7. Both are equally likely, which means that the computer will choose €0.7 as the payment half the times and €2.7 the other half.

	Probability 50%	Probability 50%
1	€1.5	€1.5
2	€1.3	€1.8
3	€1.1	€2.1
4	€0.9	€2.4
5	€0.7	€2.7
6	€0.6	€2.8
7	€0.4	€2.9
8	€0	€3

You must choose one of the 8 possible options. To that end, an empty box will appear where you must enter the number of the option (from 1 to 8) that you want to choose.

TASK III:

Next you will be matched randomly with another participant in this room. You will be presented with 6 situations in which you will have to choose one of 9 options. Each option represents the amount of money that you can earn from this task and the amount that the participant with whom you are matched can earn.

At the end of the task one participant in the matching will be randomly selected as the Decisor and the other as Receptor. The computer will randomly select one of the 6 situations and the payment that you will get is the following:

- If you are the Decisor, you will get what you have chosen for yourself in the situation selected by the computer
- If you are the Receptor, you will get what the other participant has chosen for you in the situation selected by the computer

The amounts shown are in Euro cents.

C.2 Robustness and Additional Results

TABLE C.1– GENDER INTERACTION EFFECTS FOR BEHAVIOR: PROB. OF ACCEPTING

	Symmetric			Empowerment			Entitlement			Information		
	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	-0.0356	-0.0745***	0.00744	-0.0351	-0.0133	-0.0298	5.36e-05	0.0480	-0.00264	-0.0350	0.000781	-0.0338*
	(0.0604)	(0.0271)	(0.0389)	(0.0413)	(0.0300)	(0.0266)	(0.0498)	(0.0398)	(0.0361)	(0.0376)	(0.0189)	(0.0197)
β_2: Male Resp.	-0.0108	0.0278	0.00870	-0.0187	-0.0472	-0.0545*	0.0746*	0.0941*	0.0115	-0.0285	-0.0280	0.0552**
	(0.0390)	(0.0377)	(0.0415)	(0.0391)	(0.0289)	(0.0280)	(0.0365)	(0.0527)	(0.0211)	(0.0328)	(0.0269)	(0.0253)
β_3': Male#Male	0.110	0.0186	0.0113	0.0978	-0.0449	0.0224	-0.0648	0.0850	0.0398	-0.0304	-0.0758*	-0.0668
	(0.110)	(0.0732)	(0.109)	(0.0838)	(0.0510)	(0.0599)	(0.102)	(0.0777)	(0.0501)	(0.0784)	(0.0426)	(0.0450)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	164	664	608	170	985	877	122	706	644	170	1,133	930
No. Bargains	164	117	109	170	155	146	122	93	86	170	154	143
No. Clusters	36	33	32	37	37	34	27	27	26	37	37	37

Notes: GLS random-effects model for the probability of accepting opening offers (*1^{st} Offer*), *Subsequent Offers* and *Demands*. *Other Controls* include the round and the time remaining in seconds at the point when the offer (demand) is made, the current offer (demand), the previous offer, the previous demand, and the individual controls for the responder (proposer). Clustered standard errors at the responder level for the probability of accepting offers and at the proposer level for demands. *** p<0.01, ** p<0.05, * p<0.1

TABLE C.2– DISTRIBUTION BY GENDER MATCHINGS BY EXPERIENCE

First Half (Periods 1-5)

	Symmetric	Empowerment	Entitlement	Information	Total
FF	24.0%	24.0%	27.3%	28.6%	25.9%
MF	27.0%	28.0%	27.3%	20.0%	25.3%
FM	25.0%	24.0%	22.7%	27.6%	25.1%
MM	24.0%	24.0%	22.7%	23.8%	23.7%
Observations	100	100	66	105	371

Second Half (Periods 1-5)

	Symmetric	Empowerment	Entitlement	Information	Total
FF	25.0%	24.0%	25.0%	25.7%	24.9%
MF	27.0%	17.0%	25.0%	25.7%	23.6%
FM	23.0%	35.0%	22.2%	27.6%	27.3%
MM	25.0%	24.0%	27.8%	21.0%	24.1%
Observations	100	100	72	105	377

TABLE C.3– EFFECTS OF ULTIMATUMS ON RESPONDER'S PIE SHARE

	Successful Agreements	*All*
	(1)	(2)
Ultimatum	0.0308*	0.0897***
	(0.0176)	(0.0195)
Offer	0.00359	0.000326
	(0.0106)	(0.0124)
Ultimatum#Offer	-0.0694***	-0.161***
	(0.0201)	(0.0216)
Other Controls	YES	YES
Observations	626	748
Clusters Prop.	138	140
Clusters Resp.	137	139

Notes: All regressions control for *Pie Size* and include *Period* and *Treatment* Fixed Effects. *Offer* is a dummy variable that takes a value of 1 if the proposal is coming from the *Proposer* and 0 otherwise. *Other Controls* include the following for both proposer and responder: gender, risk and social preferences, self-assessed ability in the real effort task, and self-assessed ability in bargaining. Standard errors are clustered at the *Proposer* and *Responder* level using two-way clustering. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE C.4– THE EFFECTS OF EXPERIENCE IN GENDER AND GENDER INTERACTION EFFECTS: PROB ACCEPTING

Panel A: First Half (Periods 1–5)

	Symmetric			Empowerment			Entitlement			Information		
	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	0.0303	-0.151***	-0.0396	-0.0613	-0.00177	-0.0476	0.0218	0.149**	-0.0172	-0.0641	-0.00774	-0.0659
	(0.0942)	(0.0553)	(0.0799)	(0.0556)	(0.0503)	(0.0382)	(0.0709)	(0.0674)	(0.0402)	(0.0642)	(0.0333)	(0.0442)
β_2: Male Resp.	0.0464	0.0123	0.0427	-0.00630	-0.0584	-0.0935**	0.110	-0.00909	0.0710*	-0.0271	-0.119**	0.0879**
	(0.0701)	(0.0517)	(0.0892)	(0.0508)	(0.0424)	(0.0389)	(0.0681)	(0.0872)	(0.0418)	(0.0513)	(0.0500)	(0.0447)
β_3': Male#Male	0.149	0.149	0.202	0.0471	0.0981	0.0569	-0.338**	0.374***	0.0769	-0.0619	-0.0948	-0.0627
	(0.159)	(0.167)	(0.176)	(0.121)	(0.0924)	(0.0983)	(0.139)	(0.0927)	(0.0994)	(0.104)	(0.0648)	(0.0874)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	82	277	250	90	421	359	58	272	245	88	440	402
No. Bargains	82	53	48	90	81	74	58	41	38	88	78	72
No. Clusters	37	33	33	37	34	36	26	26	27	38	38	36

Panel B: Second Half (Periods 6–10)

	Symmetric			Empowerment			Entitlement			Information		
	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands	1^{st} Offer	Subsequent Offers	Demands
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	-0.114**	-0.0210	0.0152	0.0182	-0.0198	-0.0180	-0.0285	-0.0195	0.0357	0.0186	-0.0102	0.00880
	(0.0498)	(0.0336)	(0.0545)	(0.0728)	(0.0365)	(0.0278)	(0.0657)	(0.0578)	(0.0566)	(0.0424)	(0.0185)	(0.0230)
β_2: Male Resp.	-0.0322	0.0199	0.00729	-0.0360	-0.0342	0.0271	0.0195	0.146**	-0.0290	-0.0333	0.0323	0.0495***
	(0.0446)	(0.0587)	(0.0507)	(0.0572)	(0.0404)	(0.0479)	(0.0572)	(0.0583)	(0.0386)	(0.0282)	(0.0261)	(0.0178)
β_3': Male#Male	0.117	-0.120	-0.136	0.151	-0.130**	0.0666	0.217	-0.122	-0.00197	0.00771	-0.0619	-0.0689**
	(0.131)	(0.0749)	(0.158)	(0.138)	(0.0605)	(0.0786)	(0.187)	(0.131)	(0.0481)	(0.0984)	(0.0406)	(0.0339)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	82	387	358	80	564	518	64	434	399	82	693	652
No. Bargains	82	64	61	80	74	72	64	52	48	82	76	74
No. Clusters	37	33	33	37	34	36	26	26	27	38	38	36

Notes: GLS random-effects model for the probability of accepting opening offers (1^{st} *Offer*), *Subsequent Offers* and *Demands*. *Other Controls* include the round and the time remaining in seconds at the point when the offer (demand) is made, the current offer (demand), the previous offer, the previous demand, and the individual controls for the responder (proposer). Clustered standard errors at the responder level for the probability of accepting offers and at the proposer level for demands. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE C.5– ROBUSTNESS: GENDER AND GENDER INTERACTION EFFECTS BY EXPERIENCE

Panel A: First Half (Periods 1–3)

	Symmetric			Empowerment			Entitlement			Information		
	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	0.103	-35.41*	-0.0394**	-0.0262	-45.44***	-0.00414	-0.105	-0.592	-0.0325	0.125	11.02	-0.0816
	(0.103)	(21.31)	(0.0168)	(0.0422)	(10.97)	(0.0400)	(0.110)	(28.94)	(0.0284)	(0.0880)	(11.75)	(0.0500)
β_2: Male Resp.	-0.00500	14.61	0.0151	0.0346	-31.20**	-0.104***	-0.118	17.84	0.0972**	0.0343	-15.93	0.0580**
	(0.110)	(13.32)	(0.0337)	(0.0583)	(12.59)	(0.0306)	(0.0789)	(19.93)	(0.0406)	(0.0818)	(14.56)	(0.0266)
β_3': Male#Male	-0.0356	71.78**	0.00851	-0.0665	24.03	0.0161	0.0822	15.39	0.00411	-0.261	-31.70*	0.105***
	(0.215)	(31.20)	(0.0397)	(0.0987)	(20.23)	(0.0820)	(0.145)	(38.94)	(0.0599)	(0.161)	(17.83)	(0.0403)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	60	49	49	60	57	57	40	35	35	63	57	57
Clusters Prop	30	27	27	31	31	31	22	21	21	34	32	32
Clusters Resp	30	28	28	30	29	29	21	21	21	35	34	34

Panel B: Second Half (Periods 8–10)

	Symmetric			Empowerment			Entitlement			Information		
	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share	Success	Remaining Time	Responder's Pie Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
β_1: Male Prop.	0.0695	-17.74	0.0181	-0.161	-12.54	-0.0268	-0.0105	-12.97	0.0239*	-0.139	-36.66**	0.0270
	(0.125)	(14.30)	(0.0248)	(0.130)	(11.97)	(0.0336)	(0.0611)	(21.53)	(0.0143)	(0.0935)	(14.63)	(0.0332)
β_2: Male Resp.	-0.0257	19.71	-0.0305	-0.105	-24.79**	0.0181	-0.158	-10.19	-0.0275	-0.0130	5.972	-0.00607
	(0.0969)	(15.31)	(0.0294)	(0.0729)	(11.65)	(0.0260)	(0.110)	(17.64)	(0.0318)	(0.0933)	(12.47)	(0.0519)
β_3': Male#Male	0.0358	-53.29	0.0239	0.0708	16.01	0.00359	-0.00855	-58.15	0.00951	0.319*	-11.96	0.0373
	(0.219)	(32.97)	(0.0571)	(0.147)	(25.08)	(0.0469)	(0.268)	(44.16)	(0.0504)	(0.181)	(21.84)	(0.0824)
Other Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	60	50	50	60	48	48	39	34	34	63	49	49
Clusters Prop	28	24	24	28	26	26	22	22	22	33	28	28
Clusters Resp	29	27	27	27	25	25	21	21	21	33	32	32

Notes: All regressions control for *Pie Size* and include Period Fixed Effects. OLS for the three main outcome variables for each treatment. *Success* takes a value of 1 if the subjects reach a deal within the 3-minute limit and 0 otherwise. *Remaining Time* is the time left in seconds from the time when the agreement is reached to the end of the three-minute limit. *Responder's Pie Share* is the share that is obtained by the responder in each successful bargaining. *Entitlement* considers only those matchings in which an effective entitlement is implemented, so ties between subjects' productivities are disregarded. *Individual Controls* include subjects' risk and social preferences and self-assessed ability in the real effort task and in the bargaining ability, separately for *Proposers* and *Responders*. Standard errors are clustered at the proposer and responder level using two-way clustering.*** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE C.6– THE EFFECTS OF EXPERIENCE IN GENDER AND GENDER INTERACTION EFFECTS IN FRIENDLY AGREEMENTS

First Half (Periods 1–5)

	Symmetric (1)	Empowerment (2)	Entitlement (3)	Information (4)
β_1: Male Prop.	0.0356	-0.0287	-0.0829***	-0.0561
	(0.0278)	(0.0439)	(0.0320)	(0.0424)
β_2: Male Resp.	-0.000593	-0.0712***	0.0408*	0.0509**
	(0.0224)	(0.0245)	(0.0231)	(0.0259)
β_3': Male#Male	-0.0193	-0.0198	-0.0349	0.157***
	(0.0470)	(0.0633)	(0.0431)	(0.0443)
Other Controls	YES	YES	YES	YES
Observations	68	75	41	67
Clusters Prop.	32	31	21	32
Clusters Resp.	31	29	21	31

Second Half (Periods 6–10)

	Symmetric (1)	Empowerment (2)	Entitlement (3)	Information (4)
β_1: Male Prop.	0.00500	-0.0157	0.00821	0.00527
	(0.0147)	(0.0311)	(0.0217)	(0.0489)
β_2: Male Resp.	-0.0125	-0.00816	-0.0356	0.000500
	(0.0208)	(0.0299)	(0.0224)	(0.0453)
β_3': Male#Male	0.000992	-0.0612	0.0137	-0.0911
	(0.0438)	(0.0400)	(0.0470)	(0.0593)
Other Controls	YES	YES	YES	YES
Observations	66	54	44	45
Clusters Prop.	28	26	21	20
Clusters Resp.	30	28	21	25

Notes: All regressions control for *Pie Size* and for who makes the last offer (Proposer or Responder) and include *Period* Fixed Effects. *Other Controls* include risk and social preferences and self-assessment in the real effort task and in bargaining for both, proposers and responders. Standard errors are clustered at the *Proposer* and *Responder* level using two-way clustering. Clustered standard errors at the individual level using two-way clustering. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Bibliography

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *The American Economic Review*, 101(2), 470–492.

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions: With formulas, graphs, and mathematical tables*, volume 55. Courier Corporation.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and Identity. *The Quarterly Journal of Economics*, (pp. 715–753).

Andersen, S., Ertac, S., Gneezy, U., List, J. A., Maximiano, S., et al. (2013). On the cultural basis of gender differences in negotiation.

Ariely, D., Ockenfels, A., and Roth, A. E. (2005). An experimental analysis of ending rules in internet auctions. *RAND Journal of Economics*, 36(4), 891–908.

Arrow, K. (1973). The theory of discrimination. *Discrimination in labor markets*, 3(10).

Atwater, L. E., Brett, J. F., Waldman, D., DiMare, L., and Hayden, M. V. (2004). Men's and women's perceptions of the gender typing of management subroles. *Sex Roles*, 50(3-4), 191–199.

Ayres, I. (1991). Fair driving: Gender and race discrimination in retail car negotiations. *Harvard Law Review*, (pp. 817–872).

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

Azmat, G. and Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics*, 30, 32–40.

Babcock, L. and Laschever, S. (2009). *Women don't ask: Negotiation and the gender divide*. Princeton University Press.

Barbulescu, R. and Bidwell, M. (2013). Do women choose different jobs from men? Mechanisms of application segregation in the market for managerial workers. *Organization Science*, 24(3), 737–756.

Bayard, K., Hellerstein, J., Neumark, D., and Troske, K. (2003). New evidence on sex segregation and sex differences in wages from matched employee–employer data. *Journal of Labor Economics*, 21(4), 887–922.

Becker, G. (1971). The Economics of Discrimination. *University of Chicago Press*.

Bertrand, M. and Hallock, K. F. (2001). The gender gap in top corporate jobs. *Industrial & Labor Relations Review*, 55(1), 3–21.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59(5), 960–970.

Beyer, S. and Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, 23(2), 157–172.

Bian, L., Leslie, S.-J., and Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, 355(6323), 389–391.

Binmore, K., Shared, A., and Sutton, J. (1989). An outside option experiment. *The Quarterly Journal of Economics*, 104(4), 753–770.

Blau, F. D. and Kahn, L. M. (1997). Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of Labor Economics*, (pp. 1–42).

Blau, F. D. and Kahn, L. M. (2000). Gender differences in pay. *The Journal of Economic Perspectives*, 14(4), 75–99.

Booth, A. and Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81(2), 542–555.

Bordalo, P., Gennaioli, N., and Shleifer, A. (in press). Stereotypes. *The Quarterly Journal of Economics*.

Brandts, J., Groenert, V., and Rott, C. (2014). The impact of advice on women's and men's selection into competition. *Management Science*, 61(5), 1018–1035.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31–72.

Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review*, 89(1), 306–318.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2), 238–249.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, 131(2), 633–686.

Castillo, M., Petrie, R., Torero, M., and Vesterlund, L. (2013). Gender differences in bargaining outcomes: A field experiment on discrimination. *Journal of Public Economics*, 99, 35–48.

Cejka, M. A. and Eagly, A. H. (1999). Gender–stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, 25(4), 413–423.

Chatard, A., Guimond, S., and Selimbegovic, L. (2007). "how good are you in math?" The effect of gender stereotypes on students' recollection of their school marks. *Journal of Experimental Social Psychology*, 43(6), 1017–1024.

Clance, P. R. and Imes, S. A. (1978). The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychotherapy: Theory, Research & Practice*, 15(3), 241–247.

Coate, S. and Loury, G. C. (1993). Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 83(5), 1220–1240.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4), 1625–1660.

Cramton, P. C. and Tracy, J. S. (1992). Strikes and holdouts in wage bargaining: Theory and data. *The American Economic Review*, (pp. 100–121).

Crosetto, P., Weisel, O., and Winter, F. (2012). A flexible z-tree implementation of the social value orientation slider measure (murphy et al. 2011)–manual–. *Jena Economic Research Papers*, 2012, 062.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448–474.

Davies, P. G., Spencer, S. J., Quinn, D. M., and Gerhardstein, R. (2002). Consuming Images: How Television Commercials that Elicit Stereotype Threat Can Restrain Women Academically and Professionally. *Personality and Social Psychology Bulletin*, 28(12), 1615–1628.

Davies, P. G., Spencer, S. J., and Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, 88(2), 276–287.

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

Dohmen, T. and Falk, A. (2011). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *The American Economic Review*, 101(2), 556–590.

Dreber, A., von Essen, E., and Ranehill, E. (2014). Gender and competition in adolescence: task matters. *Experimental Economics*, 17(1), 154–172.

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

Eckel, C. C. and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4), 281–295.

Ehrlinger, J. and Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, 84(1), 5–17.

Eriksson, K. H. and Sandberg, A. (2012). Gender differences in initiation of negotiation: Does the gender of the negotiation counterpart matter? *Negotiation Journal*, 28(4), 407–428.

Ertac, S. and Szentes, B. (2011). : TÜSİAD-Koç University Economic Research Forum working paper series.

Ewers, M. (2012). : IZA Discussion Paper.

Exley, C. L., Niederle, M., and Vesterlund, L. (2016). *Knowing When to Ask: The Cost of Leaning In*. Technical report, National Bureau of Economic Research.

Fernandez, R. M. and Friedrich, C. (2011). Gender sorting at the application interface. *Industrial Relations: A Journal of Economy and Society*, 50(4), 591–609.

Fernandez-Mateo, I. and Fernandez, R. M. (2016). Bending the pipeline? Executive search and gender inequality in hiring for top management jobs. *Management Science*.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–178.

Frome, P. M. and Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology*, 74(2), 435–452.

Furnham, A. and Gasson, L. (1998). Sex differences in parental estimates of their children's intelligence. *Sex Roles*, 38(1-2), 151–162.

Furnham, A., Reeves, E., and Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *The Journal of Genetic Psychology*, 163(1), 24–39.

Gächter, S. and Riedl, A. (2005). Moral property rights in bargaining with infeasible claims. *Management Science*, 51(2), 249–263.

Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5), 1637–1664.

Gneezy, U., Niederle, M., Rustichini, A., et al. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3), 1049–1074.

Goenaga, J. M. and Garaño, J. (2014). *Loreak*. Irusoin / Moriarti Produkzioak.

Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114–125.

Grosse, N., R., G., and Dertwinkel-Kalt, M. (2014). Explaining gender differences in competitiveness: Testing a theory on gender-task stereotypes. *Working paper, University of Mannheim*.

Haveman, H. A. and Beresford, L. S. (2012). If you're so smart, why aren't you the boss? Explaining the persistent vertical gender gap in management. *The ANNALS of the American Academy of Political and Social Science*, 639(1), 114–130.

Hernandez-Arenaz, I. n. and Iriberri, N. (2016). Women ask for less (only from men): Evidence from alternating-offer bargaining in the field.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2013). The allocation of talent and U.S. economic growth. *National Bureau of Economic Research*.

Iriberri, N. and Rey-Biel, P. (in press). Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. *Journal of Economic Behavior & Organization*.

Joensen, J. S. and Nielsen, H. S. (2015). Mathematics and gender: Heterogeneity in causes and consequences. *The Economic Journal*.

Kamas, L. and Preston, A. (2012). The importance of being confident; gender, career choice, and willingness to compete. *Journal of Economic Behavior & Organization*, 83(1), 82–97.

Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.

Lazear, E. P. and Rosen, S. (1981). Rank-Order Tournaments as Optimum Labor Contracts. *The Journal of Political Economy*, 89(5), 841–864.

Leibbrandt, A. and List, J. A. (2014). Do women avoid salary negotiations? Evidence from a large-scale natural field experiment. *Management Science*, 61(9), 2016–2024.

Lenney, E. (1977). Women's self-confidence in achievement settings. *Psychological Bulletin*, 84(1), 1–13.

Leslie, S.-J., Cimpian, A., Meyer, M., and Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219), 262–265.

Lippmann, W. (1922). *Public Opinion*. New York, Harcourt, Brace.

List, J. A. (2006). Friend or foe? A natural experiment of the prisoner's dilemma. *The Review of Economics and Statistics*, 88(3), 463–471.

Ludsteck, J. (2014). The impact of segregation and sorting on the gender wage gap: Evidence from German linked longitudinal employer-employee data. *Industrial & Labor Relations Review*, 67(2), 362–394.

Macpherson, D. A. and Hirsch, B. T. (1995). Wages and gender composition: Why do women's jobs pay less? *Journal of Labor Economics*, 13(3), 426–471.

Mengel, F. (2015). Gender differences in networking.

Morgan, J., Sisak, D., and Vardy, F. (in press). The ponds dilemma. *The Economic Journal*.

Murphy, R. O., Ackermann, K. A., and Handgraaf, M. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771–781.

Nash Jr, J. F. (1950). The bargaining problem. *Econometrica: Journal of the Econometric Society*, (pp. 155–162).

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3), 1067–1101.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, and others (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106(26), 10593–10597.

O'Neill, J. (2003). The gender gap in wages, circa 2000. *The American Economic Review*, 93(2), 309–314.

Paglin, M. and Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics*, 8(1), 123–144.

Petrides, K. V., Furnham, A., and Martin, G. N. (2004). Estimates of emotional and psychometric intelligence: Evidence for gender-based stereotypes. *The Journal of Social Psychology*, 144(2), 149–162.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, 62(4), 659–661.

Pomerantz, E. M., Altermatt, E. R., and Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology*, 94(2), 396–404.

Post, T., Van den Assem, M. J., Baltussen, G., and Thaler, R. H. (2008). Deal or no deal? Decision making under risk in a large-payoff game show. *The American Economic Review*, 98(1), 38–71.

Reuben, E., Rey-Biel, P., Sapienza, P., and Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization*, 83(1), 111–117.

Rigdon, M. (2012). An experimental investigation of gender differences in wage negotiations.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33–38.

Roth, A. E., Murnighan, J. K., and Schoumaker, F. (1988). The deadline effect in bargaining: Some experimental evidence. *The American Economic Review*, 78(4), 806–823.

Roth, A. E., Ockenfels, A., et al. (2002). Last-minute bidding and the rules for ending second-price auctions: Evidence from ebay and amazon auctions on the internet. *American Economic Review*, 92(4), 1093–1103.

Säve-Söderbergh, J. (2007). Are women asking for low wages? Gender differences in wage bargaining strategies and ensuing bargaining success.

Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social Issues*, 57(4), 675–688.

Shurchkov, O. (2012). Under pressure: Gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, 10(5), 1189–1213.

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

Sterbenz, F. P. and Phillips, O. R. (2001). Bargaining experiments with deadlines and random delays. *Economic Inquiry*, 39(4), 616–626.

Sutter, M., Bosman, R., Kocher, M. G., and van Winden, F. (2009). Gender pairing and bargaining – Beware the same sex! *Experimental Economics*, 12(3), 318–331.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, 99(1), 1–10.

van Dolder, D., van den Assem, M. J., Camerer, C. F., and Thaler, R. H. (2015). Standing united or falling divided? High stakes bargaining in a TV game show. *The American Economic Review*, 105(5), 402–07.

Van Poucke, D. and Buelens, M. (2002). Predicting the outcome of a two-party price negotiation: Contribution of reservation price, aspiration price and opening offer. *Journal of Economic Psychology*, 23(1), 67–76.

Weber, E. U., Blais, A.-R., and Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15(4), 263–290.

Wozniak, D., Harbaugh, W. T., and Mayr, U. (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics*, 32(1), 161–198.