FAIRNESS, GENDER AND THEIR CONFOUNDERS

José Gabriel Castillo[1]

Abstract

Gender differences in behavior, both in economic and non-economic domains, have been observed consistently in experimental evidence. A general view derived from these efforts is that women are more altruistic and tend to show more pro-social behavior. By means of an Ultimatum Game, combined with other constructs to control for ability, preferences and personality traits, I present evidence of a laboratory experiment on senior high school students that suggests that gender is not a determinant factor on fairness behavior; in the sense that, once controlling for potential confounders, observed differences are negligible in statistical sense. I present results on two versions of the Ultimatum Game, the direct and strategy method, and find strong evidence of mean behavioral differences across methods but no gender differences within each approach. The document explores some potential routs of explanation.

Author for correspondence
Email:
[1] jcastil@espol.edu.ec (José Gabriel Castillo), Department of Social Sciences and Humanities, Escuela Superior Politécnica del Litoral, Ecuador.

EQUIDAD, GÉNERO Y SUS CONFUSORES

Resumen

Las diferencias de género en el comportamiento, tanto en los ámbitos económicos como no económicos, se han observado consistentemente en la evidencia experimental. Una visión derivada de estos esfuerzos es que las mujeres son más altruistas y tienden a mostrar un comportamiento más pro-social. Por medio de un Juego de Ultimátum, combinado con otros constructos para controlar las habilidades, las preferencias y los rasgos de personalidad; presento evidencia de un experimento de laboratorio en estudiantes de secundaria que sugiere que el género no es un factor determinante en el comportamiento de equidad; en el sentido de que, una vez que se controlan los posibles factores de confusión, las diferencias observadas son insignificantes en sentido estadístico. Presento los resultados en dos versiones del Juego de Ultimátum, el método directo y el método de estrategia, y encuentro una fuerte evidencia de las diferencias de comportamiento promedio entre los métodos, pero no diferencias de género dentro de cada enfoque. El documento explora algunas posibles rutas de explicación.

1. Introduction

Plenty of research in experimental sciences concentrates in the study of gender differences in behavior, both in economic and non-economic domains. Whether in experimental psychology, sociology or political science, gender differences arise in several aspects such as: moral behavior, giving behavior, impulsivity or self-control, social interaction, and responses to games (Eagly and Crowley, 1986; Gilligan, 1982; Uesugi and Vinacke, 1963; Chapple and Johnson, 2007), criminality and drug use (Cooperstock and Parnell, 1982; Gottfredson and Hirschi, 1990), political preferences and motivations (Moore, 1996; Goertzel, 1983; Christy, 1987). In economics, gender differences have been observed in different aspects of individual preferences, decision making and labor productivity. Women's behavior is generally related to lower levels of risk tolerance (Eckel and Grossman, 2008) and lower tendency towards competition involvement (Niederle and Vesterlund, 2007). Gender's reaction to competition environments that involve strategic interaction, has also been documented to be significant. Women tend to perform better in competitive environments where they compete against other women, and these differences appear in early ages and in tasks that are gender neutral, e.g. solving mazes (Gneezy et al., 2003; Gneezy and Rustichini, 2004).

Also, gender differences are observed in terms collective action, pro-social behavior, cooperation and coordination; although results are somewhat mixed. Bolton and Katok (1995) find no differences in the play of dictator's games[2] that involve choices restrictions. Eckel and Grossman (1998) on a double blind dictator's game environment, which avoids risk involvement in the decision, find that women donate twice as much as men to anonymous partners. A common perception that derives from this evidence is that women are more altruistic and tend to show more pro-social behavior. Explanations over these differences are deeply rooted in evolutionary biology and anthropology; hence, understanding the mechanisms through which these phenomena act is worth pursuing.

An important aspect to consider in the use of experimental measurements is the fact that traits in human behavior can seldom be studied on isolation. The decision making process comprises a complex interaction of different domains and circumstances difficult to isolate or even control for. Consequently, researchers should be careful in interpreting the experimental results, especially if one takes the lab to the field. The lab is not free of these problems, experimental subjects bring to the lab all institutional and cultural background, social norms in which they coexist, institutions that might shape not only beliefs but actual preferences. Experimental replication and cross cultural studies offer an opportunity to understand unexpected environmental influences (Eckel et al., 2015; Henrich et al., 2005). Schechter (2007) shows, on a lab-in-the-field experiments in rural Paraguay, that measurements of risk aversion are good predictors of behavior in trust games; more importantly for inference is the fact that risks measures are related to other controls, strongly affecting their levels and significance. As Eckel and Grossman (1998) suggested, not controlling for these potential confounders can strongly affect conclusions.

Another source of variation can come directly from the game environment implemented. Subtle changes can affect considerably the inference. Thus, for consistency, it is important to evaluate whether the setup has some unexpected influence on the results. Such is the case of the study of fairness by means of the canonical Ultimatum Game (UG, henceforth) (Güth et al., 1982; Güth and Tietz, 1990). Eliciting fairness through this environment has typically been

---

[2] A *dictator game* is a setup where two subjects are paired and receive a fixed endowment; however, only one will make a decision over how much of the pie to distribute, while the other has no decision to make and receives whatever is left by the other player.

implemented in two different setups, the "direct" and the "strategy" method. Both approaches have pros and cons, and standard theoretical game theory would conclude that both yield the same result. Nevertheless, previous experimental evidence shows that both methods can result in different responses, in particular, as suggested by Brandts and Charness (2011), the strategy method can be viewed as "too psychologically 'cold' to be realistic as an abstraction of the natural setting." Emotions and self-reflexion triggered on experimental environments can influence results in undetermined ways, including heterogeneous gender responses.

The purpose of this research is to analyze whether personal traits on risk tolerance and ability act as potential confounders of the influence of gender on fairness measures, and whether results hold for environments with different levels of emotion-involvement.

We collected data from an incentivized laboratory experiment on high school students during a career fair visit to the university. By means of the UG, as the standard environment for fairness assessment, I analyze whether two potential confounders of gender differences intervene. On one hand, it is embedded in the UG the uncertainty over the potential response of the second mover; thus, risk is directly related to behavior on offers observed. On the other, cognitive abilities and individual attitudes are directly related to the decision making process; as a result, subjects with better cognitive abilities are potentially better utility maximizers. High-ability individuals should recognize the incentives of the game and objectively decide over maximizing their payoff, as opposed to involve other emotions in assessing unfair offers. Previous research on neural activity using the UG has shown that areas of the brain involved in emotion formation highly interact with cognitive processing while accepting or rejecting "unfair offers," i.e. offers below the 50:50 threshold (Sanfey et al., 2003). I proxy the cognitive abilities using a common effort task based on precisely rewriting a not-so-random text where I collect the Levenshtein's distance as a measure for precision. To control for risk attitudes, the Boom Risk Elicitation Task-BRET is used (Crosetto and Filippin, 2013). Finally, I consider whether responses are conditional on the method applied for the UG: direct versus strategy method; which can be interpreted as differences related to emotion's involvement, preference revelation and optimization behavior.

Results show that failing to control for potential confounders, and in particular failing to control for subject's abilities, can bias estimates of the relationship between fairness and gender, in magnitude as well as significance level. I also report evidence over heterogeneous responses when using the UG on its direct approach versus its strategy version. Acceptance rates are greatly and significantly reduced in the strategy method. Gender plays a small role in the potential explanations; however, I find men underestimate their willingness to accept an offer by twice as much as women do, i.e. differences in acceptance rates between the direct and strategy method are twice as big for men than for women.

The rest of the document is organized as follows. The second section describes the experimental design, data and briefly discusses the empirical approach; the third section summarizes the main results of the analysis. The fourth section concludes.


2.   Experiment design and procedures

To analyze the relationship between fairness and gender I combine standard games that allow to evaluate the fairness attitude and perception of players as well as control for potential confounders. Finally I collect some additional information and demographics from participants. This section describes the experimental approach taken and the data collected for the analysis.

2.1. The experiment

Upon entering the laboratory, students sit at randomly assigned stations and after a quick welcome to the career fair and the lab, the experimental session starts. General instructions and all tasks in the experiment are provided through the "O-Tree" computer interface (Chen et al., 2016).

To tests for gender differences and fairness, conditional on individual characteristics, the experiment includes 3 tasks and a questionnaire for demographic information. The first task collects risk tolerance heterogeneity through the Bomb Risk Elicitation Task - BRET (Crosetto and Filippin, 2013) on its dynamic version. Students observe on the screen an 8×8 cells, i.e. 64 boxes, and the program collects randomly one box each $1\frac{1}{2}$ seconds. One randomly chosen box contains a bomb and its location is unknown for participants. Subjects decide when to stop the box collection process, after what the boxes' content is revealed. If the bomb is collected it explodes and no points are gathered; if not, each box collected add one point to the subjects account (see Figure 1a and 1b). They play the same setting 3 rounds and one is randomly chosen for payment. I explain incentives latter in the section. There are several advantages of this method, as opposed to more standard risk elicitation strategies, such as the Multiple Price List (Holt and Laury, 2002). The task is simple enough as for participants to concentrate in one decision only: stop the dynamic mechanism. Also, they have a clear notion over the probability they face for choosing the bomb ($p = 1/64$).

Figure 1: BRET (Crosetto and Filippin, 2013)-Own adaptation

(a) Box collection   (b) Bomb revelation
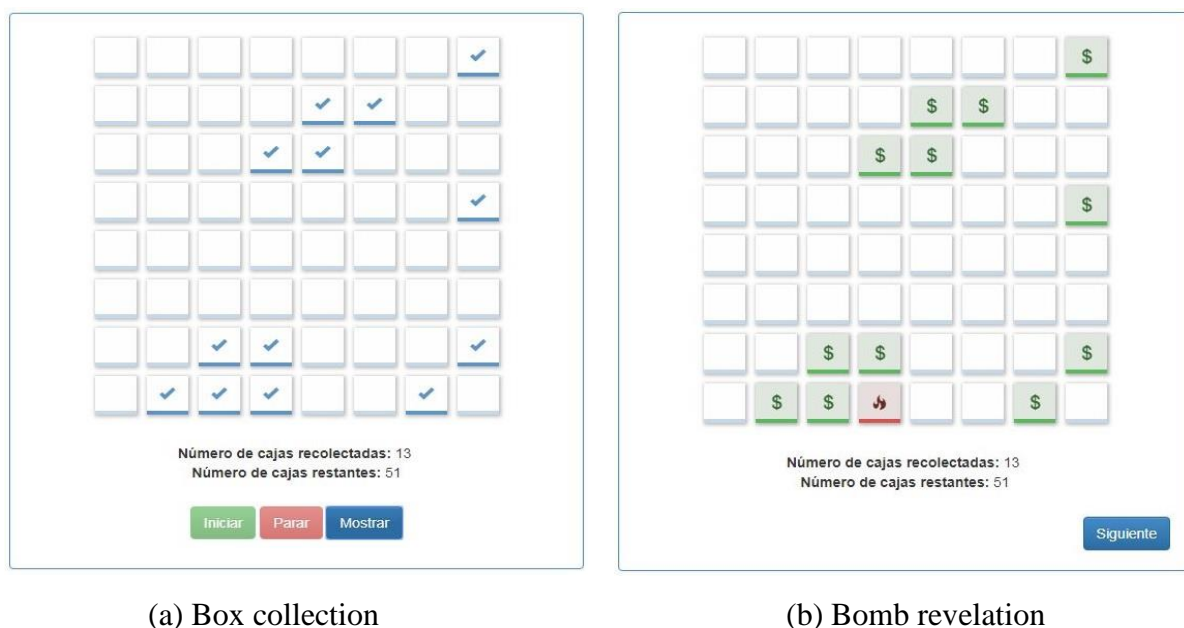
The second task comprises a simple real effort task; simple enough not to trigger gender differences. Subjects have to type precisely two paragraphs; one clear two-words paragraph for them to practice, and a two-lines paragraph composed by a not-so-random text.[3] The task is

---

[3] All participants faced the same text in both instances. Also, while the first paragraph was linguistically understandable, the second paragraph was a dummy text not-so-randomly generated, *Loren Ipsum* style (www.lipsum.com). I use this text to push their concentration on a text for which they are not familiar.

simple enough for them to understand how to work, and difficult enough to require their attention and precision on typing. Characters typed, including spaces, are counted; hence, a "Levenshtein's distance" information is collected as a measure for imprecision that comes from constantly editing the text until it is correct.[4] As shown in the summary table, there are no significant gender differences in terms of these measures. It provides information over the level of effort, concentration or skill of subjects over the task performed, hence offers some skill heterogeneity.

The third task is the core of the experiment and corresponds to the standard Ultimatum Game (Güth et al., 1982; Güth and Tietz, 1990). In this environment, subjects are randomly and anonymously paired within the experimental session. One player, the first mover, receives an endowment of 100 experimental units (EU) and makes the decision of how much of the endowment to share with the second player. The decision of the first mover is common on both versions of the game, the direct versus the strategy method; while the decision of the second player changes in each environment. On the direct method, the second player observes what the first mover assigned for him and decides whether to accept or reject the offer. If the offer is accepted, each player keeps the corresponding endowment distribution; if not, none of the players earns anything, i.e. each receives zero. The environment on the strategy method differs, second movers do not observe what has been assigned to them by the first mover, but decide over a list of ten contingent decisions they might face, from receiving 0 to receive the whole endowment, on a 10-points difference each step. In other words, they decide whether they would accept the offer if they receive an amount $x \in X = \{0, 10, 20, ..., 100\}$, without knowing how much has been assigned by the other player. Payments are realized after each individual has made a decision. Once these tasks are finished, subjects fill questionnaire for demographic information.

The experiment was monetarily incentivized by offering a total payment of $ 20.00 for the student that gained the most points during the experiment in each session, the three tasks added.[5]

2.2. Data

Six sessions were conducted in the Laboratory for Experimental Economics (L.E.E.) of ESPOL - Polytechnic University, in Guayaquil-Ecuador (http://lee.fcsh.espol.edu.ec) Subjects were recruited among senior high school students visiting the career fair at the university in October 2017. Out of 190 students, 114 (62.5%) were women. Importantly, students attended the laboratory by groups of each institution during the visiting schedule. Although within sessions subjects are randomly assigned to stations and paired for the UG; being from the same

---

[4] In information theory, the Levenshtein's distance, also called *edit distance*, is a measure of similarity between two strings. It captures the number of insertions, deletions and substitutions required to transform string one (source) to string 2 (target), hence it is a measure of imprecision since it collects the minimum number of edits to reach correctly the required text. There are several algorithms available, I used the default implemented by the computer interface "O-tree" (Chen et al., 2016).

[5] Admittedly, incentives do not meet typical payment for experimental standards. Due to logistic reasons this strategy had to be used and many aspects could affect the results. In general when the incentives were explained, students showed great motivation and interest, the amount is significant for a high school student in Ecuador, as a reference, the payment represents 5.3% of the basic monthly salary, approximately the equivalent of a day of work (8.5 hours) for a formal employee under basic salary. Results might be biased towards composition of the incentives; I doubt it since for an unexperienced subject a simple direction such as "obtain the most points" is equivalent to a suggestion over maximizing the payoff. Sympathetic to potential doubts and criticisms, I abstract from this discussion and concentrate on the results exposition.

institution one can assume they knew each other. To account for potential confounders derived from the institution we control for session fixed effects and adjust standard errors to session level. Finally, 88 subjects (46.3%) participated in a standard Ultimatum Game on its direct method; while the rest participated on the UG on its strategy version. Table 1 describes the sample and shows that sample is balanced on observable individual characteristics, among UG methods.

2.3. Empirical approach

To test gender differences in offering behavior and responses in the UG, I use two approaches for the data analysis. For the offers of first movers, I estimate an OLS regression of the form:

$$\text{Offer}_i = \alpha + \beta Gender_i + X_i'\gamma + \mu_i \tag{1}$$

where X_i is a vector of covariates for each individual i, that includes: the Levenshtein's distance, average boxes collected in the 3 rounds of the BRET, age, age squared, whether the family of the student owns the house they live in and a dummy for those that live in the city of Guayaquil.[6]

Regarding the acceptance rates of second movers, intuitively, we would like to know the actual subject's threshold of his true *willingness to accept* (WTA) an offer, which is a latent variable y^*. Instead, we observe only whether subject i accepted the offer or not, i.e. a dummy variable $y = 1$(accept the offer), which represents that the offer is higher than his minimum willingness to accept: $y = 1(y^* - \text{Offer} \leq 0)$. By allowing heterogeneous subject's sensitivity over the offer (λ), his acceptance threshold can be modeled as the acceptance probability in terms of observable characteristics and personality traits. I implement this analysis using a *probit* model that derives from a linear combination on the latent model as follows:

$$y_i^* = \delta \text{Gender}_i + X_i'\rho + \lambda \text{Offer}_i + \varepsilon_i$$

$$P(y_i = 1) = P(y_i^* \leq 0)$$
$$= P(\varepsilon_i \leq -\delta \text{Gender}_i - X_i'\rho - \lambda \text{Offer}_i)$$
$$= \Phi(-\delta \text{Gender}_i - X_i'\rho - \lambda \text{Offer}_i)$$

where $\Phi(.)$ represents a standard normal distribution, X_i is the same vector of covariates as before, and Offer$_i$ is the actual offer made by the first mover and observed by individual i. Results are shown in terms of the corresponding marginal effects.

3. Results

To analyze the relationship between fairness and fairness perception of subjects, conditional on gender, we need to analyze first whether the implemented method, direct or strategy, triggers different behavior, unconditional on gender. The first section of the results focus on this aspect while the next section summarizes the general results on gender differences.

---

[6] Some students commute from nearby cities.

3.1. Direct versus Strategy method in the Ultimatum Game

A first aspect that appears from direct observation of the data is that offers in the UG are not different between the direct and strategy method. On average, first movers offer roughly 35% of the endowment to the second movers on the direct method, while the proportion rises only around 3 percentage points on the strategy method. Differences are not significant under standard t-test ($t = -0.527$, and $p = 0.599$) or Wilcoxon-Mann-Whitney test ($z = -0.708$ and $p = 0.479$).

Table 1: Summary statistics

|	Direct	Strategic	Mean-Diff.	(p(T-test))
Ultimatum game				
Ultimatum offer	35.23	38.24	-3.01	0.458
Ultimatum response	0.86	0.54	0.32	0.000***
Individual Characteristics				
Woman	0.62	0.63	-0.01	0.851
Levenshtein distance	0.93	0.88	0.05	0.750
BRET-av	31.51	31.03	0.48	0.881
Age	16.85	16.89	-0.04	0.783
Age^2	284.61	286.52	-1.91	0.713
House owner	0.92	0.86	0.06	0.200
Guayaquil	0.90	0.91	-0.01	0.745
Observations	88	102	190	

While these results talk about the behavior of first movers, the response behavior is significantly different between methods. I find that experimental responses are on average 32 percentage points lower on the strategy method and differences are statistically significant at 1% significance level ($p(t) = 0.000$ for a t-test; $z = 3.308$ and $p(z) = 0.0009$ for a Wilcoxon-Mann-Whitney test).

From Figure 2 it is clear that most of the mass in the strategy method concentrates in offers lower than the 50% equality mark, i.e. offers can be considered more "unfair," whereas on the direct method the mass is more distributed across the support. Nevertheless, the distribution of offers is relatively equal in statistical terms (Kolmogorov-Smirnof tests for distribution equality, $p = 0.382$); thus, I cannot attribute the higher rejection rate in the strategy method to differences in the offers observed in the mechanism. On the contrary, the reason for these differences is not well defined and the experimental procedure cannot by itself disentangle the issue. This behavior might be well rooted on the influence of emotions. Brandts and Charness (2011) summarizes some previous results over differences in the direct and strategy method. Although there is no concluding evidence, it appears that punishment on unfair offers, which in the case of the UG translate on rejection rates, is lower in the strategy method.

On this study I find contradicting evidence, rejection rates are consistently higher in the strategy method (acceptance = 1- rejection). Students in the direct method seem to realize that rejecting whatever offer they receive, regardless of its fairness, will derive in lower points, hence a lower probability to win the price. Since the direct method offers a straight piece of information over the allocation offered by the first mover, the second mover's decision shows a good level of rationalization and rejection rates are of the order of 14%. Still, on the strategy

method, second movers ignore the actual decision of the allocation they face, they only reflect on contingent conditions ("what if" type). Instead of any animosity involved on the direct method, the strategy method triggers a different mechanism of decision making more related to their *willingness to accept*. In a way, the strategy method isolates the extensive form of the game and the actual realization is the result of non-binding decision.[7]

Figure 2: Histogram of contributions, by method

Graphs by Ultimatum method applied: direct or strategy
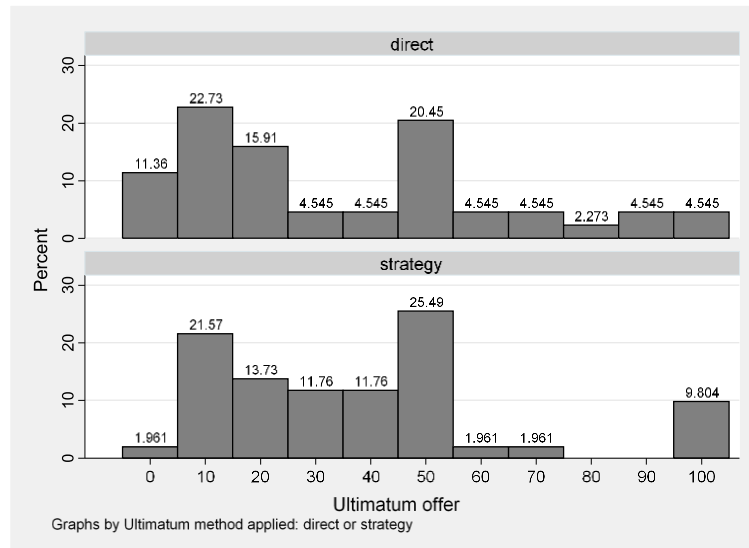
This is interesting evidence to understand self-reflexion mechanisms. When subjects are asked to decide on contingent options, they consistently underestimate what they would be willing to accept if receiving an actual offer. Mean offers are not different between methods, and random assignment within sessions (tested on observables) shields the argument over possible selection confounders. It appears subjects, are not able to reflect on their true acceptance threshold. It is likely that confusion occurs during the game, despite instructions were clear and the game fairly simple. They might try to somehow signal their valuation to first movers to push for a better distribution. Regardless, to my view, this rises some doubts over the research that relies solely on contingent behavior[8] and experimental work needs to consider this seriously. Some previous work suggests that; for example, in order to avoid *cognitive dissonance* participants should play both roles during a session; in the UG case this means playing as the one that makes the offer and later the one who accepts. In this way self-reflexion cools one's emotions and differences across methods reduce (Brandts and Charness, 2011).

It remains to see whether any of these arguments influence heterogeneous gender differences in behavior. I analyze this jointly with the fairness results in the next section.

---

[7] It is worth noting that I do observe some erratic responses as in gambling over contingent instances, i.e. subjects rejecting contingent higher values after accepting lower levels. Whether I extract those observations from the analysis do not affect the results, hence leave them in for the results exposition.

[8] A huge amount of empirical literature in environmental research, for example, fully relies on contingent valuation methods to elicit the *willingness to pay / willingness to accept* for particular services. Although, this approach is clearly needed, particularly where no market information is available, some precaution needs to be taken in making sure subjects understand the mechanisms, in order to adjust for their inability to self-reflect.

3.2. Fairness, gender differences and potential confounders

Figure 3 summarizes the results on the behavior in the UG, conditional on gender and the method used. As mentioned, there are no significant differences in the mean offers across methods; these results also extend to differences by gender between methods. Conditional on gender offers are no different from each other across methods ($z = -1.178$ and $p(z) = 0.2389$ for a Wilcoxon-Mann-Whitney test on men, $z = 0.370$ and $p(z) = 0.7114$ for women).

Figure 3: Percentage contribution and response, by gender and method
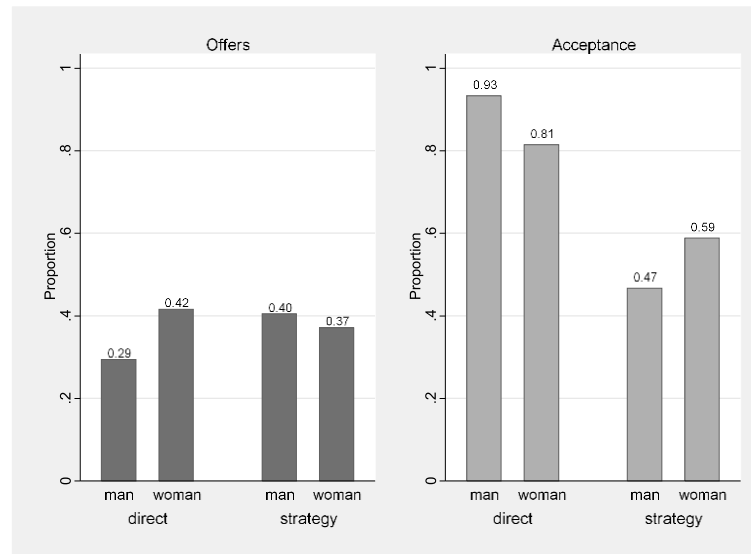
When it comes to the responses of second movers in the UG, mean differences are statistically significant across methods. This result extends to the gender domain and I find a weak significance level (10%) for men and women's responses across methods ($z = 2.742$, $p(z) = 0.0061$, for a Wilcoxon-Mann-Whitney test method differences for men; and, $z = 1.882$, $p(z) = 0.0598$, for the same tests on women). These differences, I argue, account for the underestimation of the willingness to accept when facing the actual offer (direct) rather than deciding over hypothetical potential scenarios or contingencies (strategy). Furthermore, this bias seems stronger in men which underestimate their acceptance threshold by around 46 percentage points, whereas women underestimate by a bit less than half, 22.6 percentage points difference on the acceptance rate, between methods.

It remains to test whether within each method, men and women differ in their fairness, measured by the offer on the UG; the response to the perception of fairness, and the punishment inflicted to unfair offers, measured by the acceptance/rejection rates. When analyzed on isolation, i.e. unconditional on other factors, gender differences in offers show that women offer on average 12.19 points more than men on the direct method and 3.33 less than men on the strategy method. However, variation is high on each method, and offers of men and women are no different from each other ($z = -1.142$, $p(z) = 0.2534$, for a Wilcoxon-Mann-Whitney test of gender differences in the direct method; and, $z = 0.585$, $p(z) = 0.5587$, for the same tests on the strategy method). Also, unconditional to other factors, gender differences in the acceptance rate are mixed; while on the direct method women accept on average at lower rates than men, in the strategy method results switch in favor of women; nevertheless, differences are not statistically

significant for a Wilcoxon-Mann-Whitney test ($z = 1.039$, $p(z) = 0.2987$ for the direct method; and, $z = -0.780$, $p(z) = 0.4351$ for the strategy method).[9]

Interestingly, once controlling for potential confounders and observables in a regression environment, i.e. risk tolerance and effort/skills; offering behavior becomes statistically irrelevant and smaller in magnitude. I fail to reject the hypothesis of no gender effect on offers, hence on fairness in the direct method. Standard errors are clustered at session level. Thus, correlation in the treatment has two sources, first by groups paired in the UG, and second, due to the setup of the experiment where students attended the lab per each institution. Correlation across observations cannot be isolated from the session level. Since behavior is analyzed conditional on the type of player, first movers versus second movers, I remain on the more conservative side and present clustering at session level only, as opposed to standard robust estimation.[10] Also, results are robust to various specifications including session fixed effects to account for differences that relate to the institution students belong to.

There are no such gender differences in the strategy method, whether controlling for additional covariates available or not.

One important aspect of the UG is the fact that it entails a risky decision and strategic environment in the sense that lower offers have a higher risk of being rejected. Men and women might react different to risky environments due to differences on risk tolerance (Eckel and Grossman, 2008).

---

[9] Same conclusions arise from the regression analysis without controls. See Table 2 and note that the first regression includes session fixed effects, hence the difference in the coefficient.

[10] Further results are available upon request.

Table 2: Summary of results in the Ultimatum Game (direct method)

	(1) Offer_nc	(2) Offer_c	(3) Accept_nc	(4) Accept_c
Woman	15.4406*	-3.5164	-0.2858	0.2978
	(6.1251)	(10.6807)	(0.5841)	(0.5895)
Levenshtein distance		-11.7617**		-0.4114
		(3.3450)		(0.4445)
BRET-av		0.1568		0.0082
		(0.1638)		(0.0257)
Age		64.0689		1.3241
		(145.7254)		(1.5081)
Age^2		-1.6729		0.0029
		(3.8621)		(0.0880)
House owner		46.7114*		
		(20.1170)		
Guayaquil		-29.4375**		1.6891**
		(8.9358)		(0.7210)
Ultimatum offer			0.0489**	0.0851***
			(0.0230)	(0.0248)
Session fixed effects	Yes	Yes	No	No
R-squared	0.0824	0.2792		
Observations	42	42	42	42

Notes: Standard errors clustered at session level, in parenthesis.
    Due to the sample size and collinearity, probit estimations did not allow for session fixed effects.
    For acceptance behavior, marginal changes in the probability are shown in the table.
*   Significant at 10 percent level.
**   Significant at 5 percent level.
***   Significant at 1 percent level.

Hence risk and fairness are potentially confounded in the offer levels since, if women are more risk averse, we should observe higher offers regardless of their level fairness. Eckel and Grossman (1998) control for this potential confounder avoiding completely the risk on the environment by means of a Dictator's Game where true fairness arises from the only decision to be made, how much of the pie to distribute. They reject the null hypothesis of no gender differences in mean donations, and find women are more generous, other things equal. By means of the BRET elicitation task, I do not find evidence of risk tolerance effects on the offer levels, although there are significant gender differences in the risk tolerance measure. Women

collected on average around 9.46 more boxes than men on the BRET, facing a higher probability of gaining zero. Hence in this experiment men appear as more risk averse.[11]

When looking at the factors that influence and reduced consistently the magnitude and significant level of gender on the direct method, i.e. covariates that are correlated with the gender regressor and affect offer levels, three covariates are statistically significant at standard inference levels. One is the effort task, measured by the Levenshtein's distance (significant at 5% level); in other words, the higher the distance or the more editing mistakes, which I interpret as lower skill, the lower the offer. The other two are the dummy for living in the city of Guayaquil (significant at 5% level) and a dummy variable for whether the family of the teenager owns a house (significant at 1% level), both can be considered as a summary statistic for living conditions, although rather contradicting; while as expected offers are significantly higher for students whose families are household owners, they are lower for those who live in the city.

Taken at face value, subjects that performed worst on the effort task and committed more mistakes during the editing process measured by the Levenshtein's distance, i.e. are either less skilled, less committed or less able to figure out the mistakes; the more unfair. Note that these differences appear only on the direct method and are not statistically significant on the strategy method. It is difficult to assert why such behavior is observed. On one hand, subjects in the direct method know the second mover will observe their offer and react upon it. It seems as if low-skilled individuals fail to internalize this feature and either are more risk loving, something that was discarded in terms of the risk tolerance construct, or they are simply more egotistic (unfair) and decide to take a chance as opposed to sharing a safe and fair amount. The strategy method avoids such influences by means of capturing the willingness to accept which is an independent declaration over contingent situations; hence, less convoluted with other potential animosity involved in the decision making process.

---

[11] Note that women are also over represented in the collected sample. The extent to which this might bias the results is uncertain, hence I refrain from this discussion and show the general results.

Table 3: Summary of results in the Ultimatum Game (Strategy method)

	(1) Offer_nc	(2) Offer_c	(3) Accept_nc	(4) Accept_c
Woman	-6.5898	-8.9315	0.1567	-0.2279
	(6.2122)	(10.1632)	(0.5275)	(0.4533)
Levenshtein distance		1.4757		0.3866
		(7.2693)		(0.3889)
BRET-av		0.2229		0.0021
		(0.2619)		(0.0108)
Age		-107.7090*		-4.1405
		(45.5198)		(4.3935)
Age^2		3.0013*		0.0898
		(1.1774)		(0.1151)
House owner		-1.5156		-0.3790
		(15.1312)		(0.7906)
Guayaquil		10.8557**		0.8459
		(8.7109)		(1.3677)
Ultimatum offer			0.0139*	0.0141
			(0.0078)	(0.0089)
Session fixed effects	Yes	Yes	Yes	Yes
R-squared	0.1250	0.2636		
Observations	49	49	49	49

Notes: Standard errors clustered at session level, in parenthesis.
     For acceptance behavior, marginal changes in the probability are shown in the table.
* Significant at 10 percent level.
** Significant at 5 percent level.
*** Significant at 1 percent level.

Importantly in terms of the results, it is the fact that once controlling for skills heterogeneity, the coefficient on gender (women) becomes insignificant; which suggested that there is omitted variables bias that needs to be accounted for, and unconditional results hinder this relationship. Gender differences on the effort task in the direct method are statistically relevant, and women show more skill on average.[12]

In terms of the responses in the UG, i.e. the punishment inflicted by the second mover due to the perception over the fairness/unfairness of the first mover's offer; on top of the previous result of higher acceptance rate on the direct method, again, acceptance rates are higher for women, although not statistically significant at standard levels either in the direct or strategy

---

[12] Significant 10% level, tested by means of a regression of the Levenshtein's distance measure and the gender regressor, controlling for session level and clustered standard error at session. Results not shown but available upon request.

method of the UG. It is worth noting that the actual offer *(Ultimatum offer)* on the direct method is strongly significant in terms of positively influencing acceptance rates whereas it is not the case of the strategy method; therefore, supporting the idea of emotions being triggered by facing the decision as opposed that failing to reflect on the acceptance threshold on contingent environments.

Table 4: Factors related to the minimum WTA (Strategy method)

	(1) Offer_nc	(2) Offer_c
Woman	-3.6998	-0.9536
	(5.5826)	(4.5534)
Levenshtein distance		-3.3647
		(3.3075)
BRET-av		-0.0964
		(0.1074)
Session fixed effects	Yes	Yes
R-squared	0.2694	0.4433
Observations	49	49

Notes: Standard errors clustered at session level, in parenthesis.
Other controls include: age, age squared, house owner and a dummy if he lives in Guayaquil.
* Significant at 10 percent level.
** Significant at 5 percent level.
*** Significant at 1 percent level.

Finally, returning to the discussion over heterogeneous reactions of men and women in both UG environments, the strategy method setup requires contingent decisions without knowing the actual decision of the first mover, hence collecting a form of *willingness to accept* (WTA) decision, over which subject underestimate his actual acceptance threshold. I then analyze whether information on the minimum acceptance amount declared by subjects on the strategy method sheds some light on the gender differences. I create a variable for the minimum acceptance rate that accounts for the minimum amount for which subjects switch their decision from "non-acceptance" to "acceptance," and then regress such variable on the gender regressor and other controls (see Table 4).

Although the magnitude of the differences on minimum acceptance rates favor women, differences are not statistically significant to standard significant levels, and this results holds whether we control for other personality traits and session level fixed effects. Figure 4 shows the distribution of the minimum amount for acceptance for both gender on the ten contingent alternatives. As expected, once the equal distribution of the endowment is reached, almost all subjects declare their willingness to accept the offers; however, half (50%) of women would accept offers as low as 10% of the endowment; on the same contingent decision only around 20% of men would accept it. When the minimum amount has reached 30 EU (30% of the endowment) around 63% of men would accept the offers while 66% of women would do it too. Despite these difference, distributions are not statistically different (Kolmogorov-Smirnof $p = 0.213$).

Taken together, evidence shows that, although women's tendency to accept lower offers is higher; overall, there are no significant differences on the willingness to accept the offer on the strategy method. This is enlightening since I showed previously that women underestimated their acceptance threshold for fair offers by half of men's deviation. Provided the willingness to accept, as a *cold* measure, is not to blame for this difference, it is intriguing what drives such huge underestimation in men. It is possible that men have a higher tendency to confuse the contingent environment as an opportunity to signal the first mover (or the experimenter). It is hard to distinguish the source of confusion, instructions are quite simple and there is no reason to think any of the versions of the UG environment are gender biased.

Figure 4: Histogram of minimum WTA in the strategy method, by gender
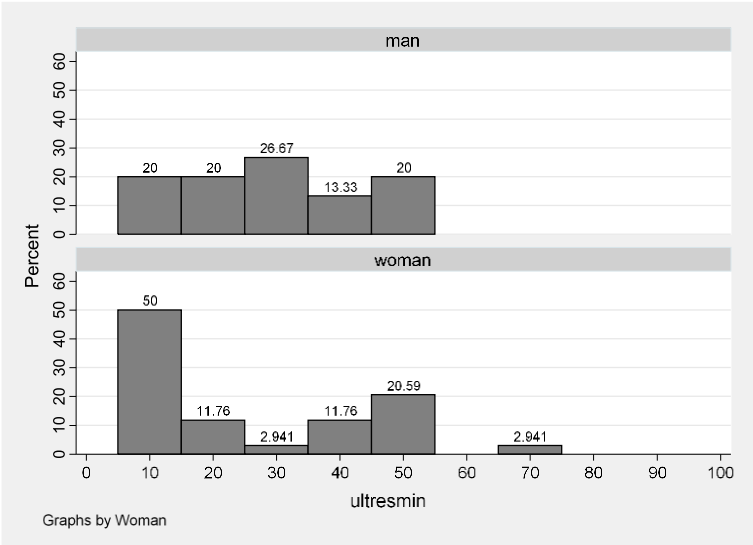
4. Conclusions

Gender differences in behavior, have been observed consistently in experimental evidence and a general conclusion that derives from this literature is that women are more altruistic and tend to show more pro-social behavior.

By means of an Ultimatum Game, I study whether gender differences in offers and responsiveness to fairness change depending on the decision's environment used: direct method or strategy method. Other elicitation mechanisms are implemented jointly to control for relevant personality traits that might confound the analysis, in particular: risk tolerance and subject's ability.

Overall, I present evidence that suggests that general ability is a relevant confounder in the analysis of fairness measures on the direct method in a UG. Once controlling for ability's heterogeneity, there are no gender differences in offer levels. Similarly, perceptions over fair distributions; thus, willingness to accept/reject offers, are no different across gender. Lastly, when using the strategy method in the UG, gender differences disappear either for offering behavior or acceptance rates. This is important because the strategy method constitutes an environment that isolates emotions from the decision making process, avoiding heterogeneous gender responses to fair/unfair offers.

A final observation is worth considering. Unconditional con gender, differences on responses to the offers are not only statistically significant but big in magnitude. Acceptance rates are around 30 percentage points different between the direct and strategy method. Taking into account that decision in the strategy method converge to a *willingness to accept*-WTA type of response, similar to what contingent valuation methodology does in several empirical applications in the field, it is clear that subjects consistently underestimate their true WTA when asked hypothetically, that when they know the decision of the first mover, as in the direct method, where the real offer is revealed before the second mover's decision. Further, it appears that there are important gender differences in the underestimation of the WTA with women converging better to their actual response when facing the offer. It is suggestive that such framing differences in the lab, would be amplified in the field; hence, researchers using such methods should strive to reduce the noise on empirical applications that comes from hypothetical environments or contingent decisions, in particular considering that there is hypothetical environments and decisions that might be gender biased.

Further research on gender differences on pro-social behavior should consider the influence of potential confounders that derive not only from personality traits but from experimental environments or games that affect self-reflexion mechanisms and trigger emotions on the decision making process.

Acknowledgments

References

I. Bolton, G. and E. Katok (1995). An experimental test for gender differences in beneficent behavior. *Economics Letters 48*(3-4), 287–292.

II. Brandts, J. and G. Charness (2011, Sep). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics 14*(3), 375–398.

III. Chapple, C. L. and K. A. Johnson (2007). Gender differences in impulsivity. *Youth Violence and Juvenile Justice 5*(3), 221–234.

IV. Chen, D. L., M. Schonger, and C. Wickens (2016). otree- an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*(Supplement C), 88 – 97.

V. Christy, C. A. (1987). *Sex differences in political participation: Processes of change in fourteen nations*. Praeger Publishers.

VI. Cooperstock, R. and P. Parnell (1982). Research on psychotropic drug use: A review of findings and methods. *Social Science & Medicine 16*(12), 1179–1196.

VII. Crosetto, P. and A. Filippin (2013, Aug). The "bomb" risk elicitation task. *Journal of Risk and Uncertainty 47*(1), 31–65.

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

IX. Eckel, C. and P. Grossman (1998, 05). Are women less selfish than men? evidence from dictator games. *108*, 726–35.

X. Eckel, C. and P. Grossman (2008, 01). Men, women and risk aversion: Experimental evidence. *1*.

XI. Eckel, C. C., H. Harwell, and J. G. Castillo G (2015). Four classic public goods experiments: A replication study. In *Replication in experimental economics*, pp. 13–40. Emerald Group Publishing Limited.

XII. Gilligan, C. (1982). *In a different voice*. Harvard University Press.

XIII. Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

XIV. Gneezy, U. and A. Rustichini (2004). Gender and competition at a young age. *The American Economic Review 94*(2), 377–381.

XV. Goertzel, T. (1983). The gender gap: Sex, family income and political opinions in the early 1980's. *JPMS: Journal of Political and Military Sociology 11*(2), 209.

XVI. Gottfredson, M. R. and T. Hirschi (1990). *A general theory of crime*. Stanford University Press.

XVII. Güth, W., R. Schmittberger, and B. Schwarze (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization 3*(4), 367 – 388.

XVIII. Güth, W. and R. Tietz (1990). Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology 11*(3), 417–449.

XIX. Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, and et al. (2005). "economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences 28*(6), 795–815.

XX. Holt, C. A. and S. K. Laury (2002, December). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

XXI. Moore, D. (1996). Clinton's increased support over past year: older independents, richest and poorest americans.'. *Gallup News Service 60*(48).

XXII.   Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

XXIII.   Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen (2003). The neural basis of economic decision-making in the ultimatum game. *Science 300*(5626), 1755–1758.

XXIV.   Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural paraguay. *Journal of Economic Behavior & Organization 62*(2), 272–292.

XXV.   Uesugi, T. K. and W. E. Vinacke (1963). Strategy in a feminine game. *Sociometry*, 75–88.