Jordi Brandts
Orsola Garofalo

Gender Pairings and Accountability Effect
November 2010
Gender pairings and accountability effects

Jordi Brandts†
Universitat Autònoma de Barcelona
Institut d’Anàlisi Econòmica, CSIC

Orsola Garofalo ‡
Universitat Autònoma de Barcelona

1 November 2010

Abstract

We conduct an experiment to investigate how the gender composition of an audience interacts with the gender of a player thereby shaping her/his degree of responsibility in decision-making. Together with measures of accountability based on decision theory, we employ two physiological measures, the blood pressure and heart rate, that allow us to disentangle the separate effects of stress and accountability in decision-making. Our results show that men are more sensitive to changes in the gender composition of the audience; specifically, men lower their accountability when paired with women. By contrast, women display a level of accountability that does not change with gender pairing.

**JEL classification:** J16, D71

**Key words:** accountability; gender differences; simple and compound events; physiological measurements.

---

We are grateful to Vicente Cuñat, Giorgio Coricelli, Christiane Schwieren, Magdalena Buckert, Mario Daniele Amore, Ludovico Minati, Christina Rott, Carles Solà as well as participants at the Summer Workshop 2010 in Neuroeconomics in Heidelberg, the LabSi Conference 2010 in Neuroeconomics and Decision Making in Siena, and the Alhambra Experimental Workshop 2010 in Granada for valuable comments and discussions. We also thank the Antoni Serra Ramoneda, UAB – Caixa Catalunya Research Chair for providing financial support. Finally, we are grateful to all the researchers that participated in the experiment as audience members. All errors remain our own.

† Universitat Autònoma de Barcelona - Department of Business Economics, and Istituto de Anàlisis Economica (CSIC). Address: Campus UAB, 08193 Bellaterra (Barcelona), Spain. Email: jordi.brandts@uab.es

‡ Universitat Autònoma de Barcelona - Department of Business Economics. Address: Campus UAB, 08193 Bellaterra (Barcelona), Spain. Email: orsola.garofalo@uab.es
1. Introduction

The role played by accountability\(^1\) in decision-making has been extensively investigated by psychology scholars. According to this strand of literature, accountable individuals exhibit greater coherence between gain and loss frames (Miller and Fagley, 1991; Takemura, 1993, 1994) and lower overconfidence (Arkes et al., 1987).

By contrast, economic scholars have only recently turned their attention to the concept of accountability, despite its numerous real-world applications to politics or corporate governance. Recent studies argue that a simple scenario in which a decision-maker expects to justify \textit{ex post} her/his choices in front of an audience might positively influence her/his ex-ante decisions (Lerner and Tetlock, 1999). Exploring this effect in an experimental setting, Vieider (2010) shows that indeed the presence of an audience increases the individual effort in solving a decision task and thus the choice of superior events.

This paper offers a novel examination of the audience-based accountability by testing whether the gender composition of the audience interacts with the gender of a decision-maker thereby shaping her/his degree of responsibility. In other words, we do not analyze single gender effects in accountability but rather we propose to investigate the effect induced by gender pairings between audience members and decision makers.

We focus on gender pairings because they are crucial to provide a comprehensive description of male and female behavior, as indicated by the extant evidence in evolutionary psychology (Buss, 1999) and recent economic works.\(^2\) Failing to account for gender pairing effects may lead to severely biased estimates of gender differences in several outcomes.

Quantifying accountability has proven to be challenging. In the spirit of Lerner and Tetlock (1999), Vieider (2010) adopts measures based on decision theory, such as the

---

\(^1\)According to the Cambridge Advanced Learner's Dictionary: "someone who is accountable is completely responsible for what they do and must be able to give a satisfactory reason for it".

\(^2\)For example, Gneezy and Rustichini (2004) found that in a single (non-competitive) race there is no difference between boys and girls at young age. However, a significant difference shows up when the children are paired with other children of the same or opposite sex: whereas boys competing against boys perform better than the non-competitive setting, girls paired with girls perform worse than when they run alone. In the mixed race, the performance improvement of boys is much larger than the one of girls.
number of correct answers and the time employed in solving a task. We claim that introducing an audience might induce not only an effect on individual accountability (i.e. by anticipating that the player has to explain her/his choices, she/he should make more responsible choices), but also an effect on psychological stress. In fact, explaining or making a presentation has been considered as valid behavioral stressor in the medical literature (Girdler et al., 1990). This confounding factor turns out to be particularly relevant for our gender context because women and men tend to react differently to behavioral stressors (see again the experimental evidence in Girdler et al., 1990 and Allen et al., 1993).

In order to mitigate this concern and provide an accurate description of how gender pairings influence accountability, we adopt two physiological measurements throughout the experiment: heart rate and blood pressure.

As discussed by prior medical research, such indicators represent reliable measures of an individual’s cardiovascular response to stress. Matthews and Stoney (1988) adopt similar measures to study gender differences in response to behavioral stressors. In their framework, the participants are asked to do serial subtractions and mirror image tracing, meanwhile the experimenter measures the heart activity and the blood pressure in specific moments of the experiment. Results show that gender is a significant determinant of the cardiovascular adjustments to stressors. In line with these results, Lawler et al. (1995) find that as response to three laboratory tasks (mental arithmetic, video-game, and anger recall interview) men have higher systolic blood pressure at all intervals and higher cardiac output during math and video-game tasks.

By combining decision-making scores and physiological indicators, we are able to examine a player’s accountability behavior from an economic viewpoint (i.e. effort and choices of the superior simple event) while monitoring physiological reactions during the
resolution of the task.

Our experimental treatments involve one player for each session and three persons forming the audience. The player is asked to make some decisions and afterwards to justify her/his choices in front of the audience; meanwhile, her/his hear rate is measured through a chest belt that she/he wears from the beginning of the experiment. In addition, we measure the blood pressure in three topical moments throughout the experiment.

In the empirical analysis, we test whether gender pairings induce significant differences in the accountability indicators while controlling for distinct patterns in the physiological measures employed. Results from the decision tasks show that men choices entail a different level of accountability depending on the gender of the audience: they make decisions with more responsibility when paired with the same sex. By contrast, women behave equally, regardless of the audience's gender. These results are consistent with the patterns highlighted by the analysis of physiological measurements.

By finding that women's level of accountability is not affected by the gender of the audience, our results are in line with the notion that women are more responsible and take more contemplated choices. Recent studies suggest that women tend to have a better behavior in monitoring (Adams et al., 2009), they are more stakeholders oriented (Matsa and Miller, 2010), and less corrupt than men (Rivas, 2008).

Our paper contributes to different strands of research. First, by measuring the accountability of male and female depending on the gender pairing, we provide new insights into the literature about gender differences, which thus far has mainly focused on competition (Niederle and Vesterlund, 2007; Gneezy and Rustichini, 2004; Gneezy et al., 2003; Larson, 2005), cooperation (Charness and Rustichini, 2009; Simpson, 2003) and coordination behavior (Cadsby and Maynes, 1998; Cadsby et al. 2007; Holm, 2000; Croson...
Second, our experimental design allows us to disentangle the accountability effect induced by the audience from the stress that the participant is subject to during the experiment on the responsibility during the task resolution, thereby adding to the research on accountability measurements.

Although our experiment does not have any specific characterization, it is far from being a merely abstract scenario; by contrast, it displays numerous potential real-world applications. For example, it can resemble a corporate situation in which a CEO has to make decisions and justify them in front of the board of directors. Managerial decisions may be influenced by the gender composition of the board, and lead a manager to make decisions that entail a different degree of accountability. Our setting may also relate to a political context in which a mayor is the ultimate responsible for governing a municipality while the counselors provide monitoring activities. Gagliarducci and Paserman (2010) document that in municipalities headed by female mayors the probability of an early termination of the legislature is higher. In line with the notion that gender pairings matter, they also argue that “the likelihood that a female mayor survives until the end of her term is lowest when the council is entirely composed by male, and in regions with less favorable attitude towards working women”.

The paper is organized as follows. In Section 2, we propose a background and advance possible relationships between gender pairings and accountability. In Section 3, we illustrate the experimental design. In Section 4, we introduce the physiological measurements adopted throughout the experiments. In Section 5, we illustrate and interpret our empirical findings. In Section 6, we conclude.

---

3See Croson and Gneezy (2009) for a review of literature on gender differences.
2. Background

According to Lerner and Tetlock (1999), individuals living in social groups do not make decisions in isolation; rather, they are often called onto defend their conclusions and their reasoning to peers, subordinates or superiors.

Following this notion, accountability can be thought as the social pressure to justify one’s views to others. More refined definitions consider accountability as a multiple phenomenon grounded on the implicit or explicit expectation that one may be called on to justify her beliefs, feelings and actions to others (Scott and Lyman, 1968; Semin and Manstead, 1983; Tetlock, 1992). Several context-based types of accountability can be discerned: (a) *mere presence of another* (individuals expect that another will observe their performance); (b) *identifiability* (individuals expect that what they say or do in a study will be linked to them personally); (c) *evaluation* (individuals expect that their performance will be assessed by others according to some normative ground rules and with some implied consequences); (d) *reason-giving* (individuals except that they must give reasons for what they say or do).

Although the results may vary depending on the definition adopted, accountability-enhancing contexts have repeatedly found to influence an individual’s engagement in high-effort elaborations, and her coherence among choices (Miller and Fagley, 1991; Takemura, 1993; 1994). Examining the roots of such impact, Tetlock (1983) and Boettger and Tetlock (1989) present evidence that accountability is ultimately shaped by what people think (i.e. preferences, cultural values and beliefs).

To motivate the importance of looking at gender pairings in accountability, we posit that a player's attitudes and beliefs toward the opposite gender might either foster or diminish the decision-maker's accountability.
Social values and gender prejudices are receiving a great deal of attention by scholars because, in addition to biological considerations, they contribute to explain variations in gender outcomes. An example is the study by Gneezy at al. (2009), which finds that in a patriarchal society men are more competitive than women, while in a matrilineal society the opposite is true. Thus Gneezy at al.’s results seem to indicate that the way in which people look at the opposite gender shapes distinct individual behaviors.

We draw on these arguments and extend them to our context: since accountability might embody values, prejudices and beliefs toward the opposite gender, gender pairings can unveil important differences in accountability behavior.

To test whether gender pairings influence accountability, we adopt the audience context such as in Vieider (2010) based upon the conceptual framework in Lerner and Tetlock (1999). However, measuring gender differences in accountability using audience effects while neglecting biological factors might be misleading. In fact, the extant evidence shows that women express a wide variety of emotions more intensively than do men. Certainly, biological factors, including genetic, hormonal, and neuropsychological variables, as well as social factors, including different gender roles, and status differences between the two genders, contribute to determine gender differences in emotions (Ablon et al., 1993). Both biological and social factors affect parents and peer socialization processes, which differ powerfully for each sex and which create different affective environments for boys versus girls (Brody, 1985; Hall, 1987; Boker and Maltz, 1982).

Whereas we measure accountability using the performance in a decision task, we attempt at measuring different emotional reactions that are related to a behavioral stressor such as the presence of an audience. Emotions can be expressed in four basic ways: verbally, behaviorally, through non-verbal facial expressions, and through physiological arousal.
increased heart rate, galvanic skin response, respiration, temperature) (Adelmann and Zajonic, 1989; Brody and Hall, 2008). Buck et al. (1974) suggest that men tend to internalize emotions; they manifest emotion in their levels of physiological arousal and not in their facial expressions. In contrast, women tend to externalize emotions; they manifest emotion in facial expressions and not in levels of physiological arousal. The medical literature on gender and cardiovascular activities reports that typically men exhibit larger blood pressure (in particular systolic values) change than women as a response to psychological challenges such as mental arithmetic, mirror image tracing, speech-making (Johnson et al., 1988). By contrast, women respond to stress by increasing the heart rate (Fineberg et al., 1990). To detect potential gender differences in stress during the resolution of the task, we track blood pressure and heart rate for each participant involved in the experiment.

3. Experimental design

The participants are recruited by email from a database (ORSEE) of students at Universitat Autònoma de Barcelona, who voluntarily have registered in the database to participate in prior experiments, or using flyers. The total number of participants is 61, of whom 33 girls (16 of them participate at the experiment with a female audience and 17 with a male audience) and 28 boys (15 of them with a female audience and 13 with a male audience). Each subject can participate only in one session. As reported in Table 1, Panel A, the average age is around 24, without significant differences between women and men. The audience is composed by three components (in turn, three women or three men) recruited from Phd students at Universitat Autònoma de Barcelona. In the instructions we specify that the audience is composed by three experts in that specific kind of task, so each participant is
aware of it before starting the resolution of the task. The average age of the audience is about 33 and it is statistically different from the participant's average age. The differences in participants' age by audience's gender (untabulated) are small and statistically not significant.

To mitigate the impact of confounding factors when using physiological measures, we collected through a questionnaire information on health characteristics that are typically associated with blood pressure or heart rate activities, such as the weight, the height, problems of high blood pressure, smoking status and so on (see Table 1, Panel B). The average weight for male subjects is 73 Kg and the average height is 1.76 cm, while for the female it is 60 Kg and 1.64 cm, respectively. None of the participants have problems of hypertension and only the 15% of the participants are smokers (Table 1, Panel B).

Each session takes about 45 minutes, it involves only one participant and the audience, and they do not know each other. All subjects (both players and audience members) are paid a flat fee of 9€ for their participation. We pay a flat fee to avoid that the participant's effort in the resolution of the task is incentivized by the compensation variable.

The pairings between gender of the subject and the audience change depending on the session. The experiment involves the analysis of a treatment represented by the gender of the audience paired with the opposite gender of the participant. Overall, we have four pairs: Female Participant-Male Audience; Female Participant-Female Audience; Male Participant-Female Audience; Male Participant-Male Audience.

The sessions take place in a semi-empty room containing a big table in which the participant sits on one side and the audience on the opposite side. The session is divided into several parts. In the first part, the experimenter explains a part of the rules and makes the participant ready for the blood pressure and heart rate measurements. In particular, the first
part of the instructions only specifies that the experiment consists in measuring the blood pressure and the heart rate, and in solving a decision task (See Appendix 2). The experimenter helps the participant in putting on the chest belt to measure the heartbeat throughout the experiment and then, in order to take time for the stabilization of the blood pressure, she gives her/him a document to fill in with personal information and a questionnaire (Multidimensional Mood State Questionnaire, Steyer et al., 1997) containing questions about personal emotional status at that moment⁴. Afterwards, the experimenter measures the blood pressure of the subject. Up to now, the subject does not know the main structure of the experiment, thus we can use the first measure of blood pressure as a baseline.

In the second part, the experimenter hands to the participant the second part of the instructions and then reads them aloud to be sure that the participant understood the structure of the experiment. This part of the experiment consists in solving the task, and in justifying the options chosen in front of an audience. Before leaving the participant alone in the room for the resolution of the task, the experimenter tells to the participant that the audience will enter the room to fill in a document, to deliver the task sheet to her/him and then to explain how the resolution of the task works.

The audience enters the room, sits down and fills the document in for a few seconds, whereas the experimenter goes out of the room. While the audience fills in the document, the participant can observe the audience members for the first time, and thereby she/he creates her/his own idea about them. Afterwards, one person of the audience explains to the participant the rules to solve the task and gives her/him the sheet containing the exercises. Then the audience leaves the room and the participant remains alone while solving the task.

⁴ It is a list of expressions that characterize different moods. For each word, the participant has to mark a number (ranged from 1 to 5) that best represents the actual intensity of her/his mood status.
As soon as the participant has solved all six problems, she/he has to ring a small bell and then the experimenter enters in the room and takes a second blood pressure measurement. Afterwards, the audience enters the room for the second time and the last part of the experiment starts: the participant justifies the decisions taken. At the end of the explanation, the audience leaves the room and the experimenter enters the room to take the third blood pressure measurement. To conclude the experiment, the participant fills in a slightly different version of the same questionnaire she/he filled in at the beginning of the experiment and signs a receipt for the payment.

In the experiment the audience plays a neutral role, it has only to pay attention to the subject during the explanation. We ask to the components of the audience to be serious during the first meeting and during the explanation but they are not allowed to directly stress or pressure the participant. The main role of the audience is to increase the commitment in the resolution of the task.

The task consists of six problems in which the participant has to choose between simple and compound prospects (as in Bar-Hillel, 1973). Bar-Hillel (1973) presents a case in which subjects were given the opportunity to choose between one of two events. Three types of events were used: (1) simple events, such as drawing a red ball from an urn containing 50 red balls and 50 black balls; (2) conjunctive events, such as drawing a red ball seven times in succession, with replacement, from a bag containing 90 red balls and 10 black balls; and (3) disjunctive events, such as drawing a red ball at least once in seven successive tries, with replacement, from a bag containing 10 red balls and 90 black balls. A significant majority of subjects chose the conjunctive event (probability of 0.48) rather than the simple event (probability of 0.50); they also preferred the simple event rather than the disjunctive event (probability of 0.52). These biases can be explained as effects of an anchoring and
adjustment process (Holtgraves and Skeel, 1992; Tversky and Kahneman, 1974; Kruglanski and Freund, 1983). The stated probability of the elementary event (successful at any one stage) provides a natural starting point for the estimation of the probabilities of both conjunctive and disjunctive events. Since the adjustment from the starting point is typically insufficient, the final estimates remain too close to the probability of the elementary events in both cases (Tversky and Kahneman, 1974).

The problems are all of the same type but with different level of analytic difficulty. Lower number of extractions in the compound event makes the calculation easier and should thus increase the choice of the superior event. In the problems chosen, the compound probability to draw the winning balls is always lower than the simple prospect probability (See Appendix 2).

4. Data and measurements

In this section, we describe in detail the physiological measurements adopted in the experiment. The heart rate was measured noninvasively using the Polar Heart Rate Monitor RS400, consisting in a belt and a wear-link transmitter that the participant wears on the chest from the beginning until the end of the experiment. The heartbeats are recorded in a watch and then downloaded on a computer using an infrared USB.

The software records the heart rate expressed in number of heartbeats per minute and it provides several indicators that describe the heart rate variability (HRV), such as standard deviation, mean, maximum, RMSSD\textsuperscript{5}, pNN50\textsuperscript{6}, Low Frequency (LF), High Frequency (HF) and LF/HF ratio expressed in milliseconds.

\textsuperscript{5}Square root of the mean squared difference of successive Normal-to-Normal intervals (NNs)

\textsuperscript{6}Proportion of NN50 divided by total number of NNs.
The HRV represents the variability of the cardiac frequency as response to respiratory rhythm, emotional stress, anger and relaxation. In a healthy heart, the cardiac frequency reacts quickly to these factors in order to better adapt the body to the external environment. The HRV is correlated with the interaction between the Sympathetic and Parasympathetic Nervous System. The Sympathetic Nervous System is responsible for acceleration of the heart rate, bronchial dilatation, increase of the blood pressure, pupillary dilatation, etc. By contrast, the Parasympathetic Nervous System produces a decrease of the cardiac rhythm, increase of the bronchial muscular tone, dilatation of the blood vessels, decrease of the blood pressure, slowdown of the breath, increase of the muscular relaxation etc. The ability of our body to tend to one or the other nervous systems represents the possession of a dynamic equilibrium from a physiological and psychological point of view.

The HRV indicators that best measure the activity of the Parasympathetic and Sympathetic Systems are the LF and HF components. While the LF component expresses the Sympathetic modulation, the HF assesses the Parasympathetic activity. As already mentioned above, our instruments provide precise measurements of all these relevant indicators. Moreover, the beginning and the end of each session are marked in the sequential measurement of the heart activity to isolate the heart activity during specific moments of the task resolution.

Figure 1 contains an example of heart activity together with three markers. The first marker is taken at the eighth minute of the experiment that corresponds with the entrance of the audience in the room for the first time; the second one is at the nineteenth minute of the experiment and it coincides with the end of the task and the beginning of the explanation; the third one is marked just after the explanation that was at the twenty-fourth minute.

To further quantify the level of stress during the task resolution, we also measure the
blood pressure, measured noninvasively using an automated blood pressure meter. Such an instrument provides the two main indicators of blood pressure: systolic and diastolic levels, respectively the maximum and minimum values of blood pressure during each heartbeat.

5. Results

Following Vieider (2010), we assume that being held accountable by an audience leads to a more thorough assessment of probabilities and thereby to more frequent choices of the superior simple event by focusing the attention on the probabilities involved. In this section, we test how the accountability measured in such a way is influenced by gender pairings between decision-makers and audience.

The choices made by a participant are encoded as dummy variables, with 1 indicating the choice of the normatively superior event (simple event), and with 0 the compound event. These dummies are summed for all six choice pairs to obtain a general index ranging from 0 to 6. Figure 2 shows the average index values in the unpaired cases, first by male and female audience, and second by male and female participants. Panel (a) shows that the mean number of superior simple event choices is in the whole sample 2.7 with male audience and 1.94 with female audience. Panel (b) reports that, independently of the audience's gender, men choose on average 3 simple event choices, and women about 2 out of six. In all cases, we do not find any statistically significant difference.

Figure 3 shows the average index values once we take into account gender pairings. The left part of the graph reports the values referred to male participants when they justify their choices in front of male or female audience. Male participants in the male audience treatment choose, on average, 3.8 superior choices out of 6; yet, with a female audience the number of superior choices goes down drastically to 1.7 out of 6. The difference between the
two averages is statistically significant at the conventional level ($p$-value=0.0064). Focusing the attention on female participants, we do not find any significant difference between the two audience treatments ($p$-value=0.682). This result suggests that women put the same level of effort in solving the task regardless the gender of the audience.

Panel A in Table 2 shows the results of OLS regressions, separately for men (Column 1 and 2) and women (Columns 3 and 4), in which the dependent variable is the sum of the superior simple event choices. The explanatory variables are the dummy treatment, equal to one when the gender of the audience and participant is the opposite, the player's age and her/his field of study (three categories are considered: social science, science and humanities; humanities is used as reference group). Our results confirm that being paired with the opposite gender has a negative and statistically significant effect on male's accountability. By contrast, the effect is quantitatively small and statistically insignificant for women.

We test the robustness of the findings above in several ways (results underported to save space, available upon request). Given the structure of our dependent variable, we estimate an ordered probit model (See Panel B in Table 2) rather than an OLS. Alternatively, we adopt a poisson regression model. As a next step, in the OLS specification we control for the difference between audience and participant's age (coefficient=-2.510; $p$-value=0.01). We explore the validity of our results within a smaller range of the dependent variable, excluding participants that choose all right or all wrong answers. Finally, we propose two alternative ways to compute standard errors; we cluster standard errors by the hour of the experiment, and we compute bootstrapped standard errors (using 500 replications) to deal with the limited sample size. Our results are in all cases qualitatively unchanged. Men's level of accountability is strongly affected by the gender of the audience; on the other hand the
level of accountability in the decision process of the female participants is not affected by the gender of the audience.

To test the theory that irrationalities in choices may due to anchoring at the single-extraction probability and insufficient adjustment away from the probability (Vieider, 2010), we examine how choices are influenced by the differences in probability of extracting a winning ball in a single-extraction. We create a panel dataset with the choices in the six tasks entered as consecutive decisions, and run a random-effects probit analysis using as dependent variable a dummy equal to one for choices of the superior simple prospect event and zero otherwise. The results show that in the group of male participants the interaction between female audience treatment and the difference in probability of extracting a winning ball in a single-extraction increases significantly the probability of choosing the inferior compound event ($Z=-2.57; \ p-value=0.01$). By contrast, we do not find any significant difference on the sample of female players ($Z=1.44; \ p-value=0.168$). Overall, this result provides further support to our previous finding that men paired with a female audience display a lower accountability.

We turn now our attention to the physiological measurements. Figure 4 presents the average values of the systolic blood pressure taken in three different moments during the experiment: the first value is measured at the beginning of the experiment and it is considered as baseline value; the second is referred to the moment just after the task and before the explanation; the third one is taken after the explanation. In panel (a) and (b) are reported, respectively, the average values for male and female subjects paired with male or female audience. Among male subjects we find a significant difference on the second systolic blood pressure across variations in the gender of the audience: the average value with the male audience is 122.08 mmHg but it increases to 130.4 mmHg with the female
Controlling for physical characteristics such as weight and height, this difference becomes slightly larger and remain statistically significant at 10% level.

Since the second measurement is taken just few seconds before the explanation, it should indicate if the participant feels stressed or not. Male subjects feel more stressed when they are paired with an audience of the opposite gender. This result seems to be in line with Van der Meij’s et al. (2008) study which shows that a short presence of a woman induces an hormonal reaction, such as the salivary testosterone (T), in young men. Findings in animals studies (Fischer and Swain, 1977; Granger et al., 1998) show that to an increasing level of T corresponds an increasing in the blood pressure values. On the other hand, the effect of a change in the T level on the arterial pressure in individuals has not been well detected, prior medical research only shows that men with low levels of T tend to have high blood pressure (Barrett-Connor al., 2004).

From medical and psychological researches we know that stress harms performance and this finding is consistent with our previous result that men perform worse when they are paired with a female audience. As in the case of task performance, we do not find that the gender of the audience matters for female blood pressure.

To identify heart rate variations in how the gender of the audience influences the reaction of male and female participants to stress (i.e. providing an explanation in our context), we compare the LF/HF Ratio around the three above-mentioned markers. In unreported analyses, we do not find any significant impact of gender pairings on heart rate activity, neither male nor female participants vary the LF/HF ratio as the gender of the audience changes. This lack of significance is not surprising though in light of the results in the decision task. In fact, as already discussed above, the medical literature indicates that

---

7 In the blood pressure graphs, we only report the systolic values because, as reported in the medical literature (see e.g. Musini and Wright, 2009), the diastolic values typically display little variations.
male responsiveness to a stressful situation mostly shows up in the blood pressure values and not in the heart rate; because we have demonstrated that gender pairings matter for men only, once we examine physiological activities we find a significant effect of gender pairings for men only on blood pressure. Also, consistent with the finding that women behave in accountable manner regardless of the gender of the audience, in the female sample physiological measures of stress do not vary depending on the gender pairing with the audience.

6. Concluding remarks

The accountability of decision makers is important for the functioning of modern institutions such as corporations and political organizations. In this paper, we have examined for the first time whether male and female shape their degree of responsibility depending on the gender of an audience that calls for an explanation of each choice made.

The experimental results obtained are twofold. First, we found that male participants are strongly influenced by the gender of the audience: when men are paired with an audience of the opposite sex, their level of accountability goes down drastically compared with the pairing with a male audience. Second, our evidence suggests that women's decisions are not affected by the gender of the audience: we find that the level of accountability does not differ between female or male audience. Moreover, women's accountability is similar to men's level in front of a female audience. For men paired with a male audience the accountability level is significantly higher. A possible explanation is that male decision-makers may feel competitive, even though the situation does not involve real competition.

The physiological indicators, which capture the stress during the resolution of the task, corroborate our previous findings: we obtained a significant difference in blood
pressure values for men paired with a female audience, and insignificant differences in the female sample. Because stress harms performance, this result is in line with the evidence that men perform worse when they are paired with women.

Overall, an interesting result of our paper is that significant differences in accountability show up only when participants are paired with people of the same or opposite gender. This finding does support the idea that examining gender differences in isolation does not offer an adequate description of the existing differences in male and female outcomes; rather, it is necessary to analyze how individuals behave once they are paired with the opposite gender. This is in line with previous works by Bosman et al. (2009) and Bolton and Katok (1995).
References


Learned from two Systematic Reviews of Randomized Controlled Trials. PloS One 22.


APPENDIX 1. FIGURES AND TABLES

Figure 1.
An example of heart rate measurement

The figure represents the measurement of the heart rate during the experiment. On the time bar, the blue part in the lower part of the graph, there are three markers: the first, approximately at the eighth minute, marks the moment in which the participant sees the audience for the first time, the second is at the nineteenth minute and it is marked just before the explanation starts, the last one is at the twenty-fourth minute and it is marked at the end of the explanation.
Figure 2.
Mean number of choices of simple prospect event

The graphs report the mean number of choices of the superior simple event. The choices are encoded as 1 if it is simple prospect and 0 if compound prospect. In panel (a) are reported the choices of the whole sample but distinguishing for the gender of the audience; panel (b) reports the results of female and male participants without a distinction of the gender of the audience.

(a) by audience’s gender  
(b) by participant’s gender

Figure 3.
Mean number of choices of simple prospect event, by gender pairings

The left part of the graph reports the values referred to male participants when they justify their choices in front of male or female audience. The right part shows the results about female participants when paired with a male or female audience.
**Figure 4a.**
Mean values of systolic blood pressure for male participants

The graph reports three values of systolic blood pressure of the male participants when paired with male or female audience. The first value is taken at the beginning of the experiment, the second at the end of the task and before the explanation starts; the last after the explanation.

**Figure 4b.**
Mean values of systolic blood pressure for female participants

The graph reports three values of systolic blood pressure of the female participants when paired with male or female audience. The first value is taken at the beginning of the experiment, the second at the end of the task and before the explanation starts; the last after the explanation.
Table 1.
Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Individual characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Participant</td>
<td>23.10</td>
<td>25.24</td>
</tr>
<tr>
<td>Age Audience</td>
<td>33.50</td>
<td>32.44</td>
</tr>
<tr>
<td>Degree/Master in social science (%)</td>
<td>67.86</td>
<td>60.61</td>
</tr>
<tr>
<td>Degree/Master in science (%)</td>
<td>7.14</td>
<td>30.30</td>
</tr>
<tr>
<td>Degree/Master in humanities (%)</td>
<td>25</td>
<td>9.09</td>
</tr>
<tr>
<td><strong>Panel B.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Physiological conditions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-smoker (%)</td>
<td>89.29</td>
<td>81.82</td>
</tr>
<tr>
<td>Any problem with high pressure (%)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Physical activity before the experiment (%)</td>
<td>3.57</td>
<td>3.03</td>
</tr>
<tr>
<td>Alcohol consumption before the experiment (%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Weight (Kg)</td>
<td>72.57</td>
<td>60.15</td>
</tr>
<tr>
<td>Height (Cm)</td>
<td>1.76</td>
<td>1.64</td>
</tr>
</tbody>
</table>
Table 2. Regression results

Panel A and B report respectively the results of the OLS regression and the Ordered Probit. In both Panels, columns (1) and (2) report the results using the sample of male participants; columns (3) and (4) report the results using the sample of female participants. The dependent variable in Columns (1)-(4) is an index built as a sum of the number of the superior simple prospect choices; the index range from 0 to 6 (each simple event chose is counted as 1 and each participant answers to six questions). The treatment represents the group in which the gender of the participant is the opposite of the one of the audience. Columns (1) and (3) report the regressions using only as explanatory variable the gender of the audience; Column (2) and (4) show the results of the regressions using in addition two control variables, age of the participant and education. Robust standard errors are reported in parenthesis. *, ** and *** denote significance at 10%, 5% and 1% respectively.

**Table 2.** Regression results

<table>
<thead>
<tr>
<th>Dependent variable: Sum of simple superior Prospect choices</th>
<th>Male participants</th>
<th>Female participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PANEL A.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>OLS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>-2.113*** (0.707)</td>
<td>-2.600*** (0.745)</td>
</tr>
<tr>
<td>Age Participant</td>
<td>-0.077 (0.073)</td>
<td>-0.030 (0.070)</td>
</tr>
<tr>
<td>Science degree</td>
<td>-1.223** (0.459)</td>
<td>2.472*** (0.871)</td>
</tr>
<tr>
<td>Social science degree</td>
<td>1.483 (1.026)</td>
<td>0.083 (1.078)</td>
</tr>
<tr>
<td><strong>PANEL B.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ORDERED PROBIT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>-1.121** (0.475)</td>
<td>-1.543** (0.609)</td>
</tr>
<tr>
<td>Age Participant</td>
<td>-0.046 (0.043)</td>
<td>-0.018 (0.042)</td>
</tr>
<tr>
<td>Science degree</td>
<td>-1.080* (0.588)</td>
<td>1.365*** (0.537)</td>
</tr>
<tr>
<td>Social science degree</td>
<td>0.958* (0.568)</td>
<td>0.090 (0.671)</td>
</tr>
<tr>
<td>Number of participants</td>
<td>28</td>
<td>28</td>
</tr>
</tbody>
</table>
APPENDIX 2. INSTRUCTION AND TASKS

INSTRUCTIONS PART 1

We thank you for coming to the experiment. The experiment involves decision-making and measurements of your blood pressure and of your heartbeat. The purpose of the experiment is to study how people make decisions in a particular situation.

At the end of the experiment we will pay you 9€ for your participation.

We first will put on the chest-belt for measuring your heartbeat. Afterward, we will measure your blood pressure. While we prepare for the measurement of blood pressure, please sit still and fill in these two questionnaires.

INSTRUCTIONS PART 2

Now we will give you 4 sheets of paper with 6 decision problems. You have to make a decision for each of these problems. You can take your time in making the decisions, there are no time limits.

After you have made the decision you will have to explain your decision during 10 minutes to three juries who are experts on these decision tasks.

When you are finished with your decision please ring this bell and then the jury will come in to the room.
Choice between Simple and Compound Prospects

In the following pages you find 6 hypothetical problems. In each problem you are asked to choose between two options. The first option allows you to extract only one ball from a bag: the second option one allows a "multiple extraction" from a different bag. In the multiple extraction option, the ball you have extracted will be placed back in the bag and the balls in the bag will be mixed before you extract again, so as to keep the composition of the bag constant.

These rules are valid for all the problems below. Please pay attention to the composition of the bags and the number of extractions because both vary across problems.

Please, mark one of the two options for each problem. Choose the option you prefer in this way:

___ X ___ Option # (extract ### time from a bag with ### red and ### green balls, win if red)

_____ Option # (extract ### time from a bag with ### red and ### green balls, win if red)

or

_____ Option # (extract ### time from a bag with ### red and ### green balls, win if red)

___ X ___ Option # (extract ### time from a bag with ### red and ### green balls, win if red)

Problem 1

Imagine you were given a choice between two options to win €20.

Option A allows you to extract one ball from a bag containing 10 red and 10 green balls. If you extract a red ball, you win €20; if you extract a green ball, you win nothing.

Option B allows you to extract 7 balls in sequence with replacement from a bag containing 18 red balls and 2 green balls. If all 7 balls extracted are red you win €20; if one or more of the balls extracted are green, you win nothing.

What would you choose?

___ Option A (extract 1 time from a bag with 10 red and 10 green balls, win if red)

_____ Option B (extract 7 times from a bag with 18 red and 2 green balls, win if 7 times red)
Problem 2

Imagine you were given a choice between two options to win €20.
Option A allows you to extract one ball from a bag containing 5 red and 15 green balls. If you extract a red ball, you win €20; if you extract a green ball, you win nothing.
Option B allows you to extract 5 balls in sequence with replacement from a bag containing 15 red balls and 5 green balls. If all 5 balls extracted are red you win €20; if one or more of the balls extracted are green, you win nothing.

What would you choose?

- Option A (extract 1 time from a bag with 5 red and 15 green balls, win if red)
- Option B (extract 5 times from a bag with 15 red and 5 green balls, win if 5 times red)

Problem 3

Imagine you were given a choice between two options to win €20.
Option A allows you to extract one ball from a bag containing 5 red and 15 green balls. If you extract a red ball you win €20; if you extract a green ball, you win nothing.
Option B allows you to extract 7 balls in sequence with replacement from a bag containing 16 red balls and 4 green balls. If all 7 balls extracted are red you win €20; if one or more of the balls extracted are green, you win nothing.

What would you choose?

- Option A (extract 1 time from a bag with 5 red and 15 green balls, win if red)
- Option B (extract 7 times from a bag with 16 red and 4 green balls, win if 7 times red)

Problem 4

Imagine you were given a choice between two options to win €20.
Option A allows you to extract one ball from a bag containing 2 red and 18 green balls. If you extract a red ball, you win €20; if you extract a green ball, you win nothing.
Option B allows you to extract 4 balls in sequence with replacement from a bag containing 10 red balls and 10 green balls. If all 4 balls extracted are red you win €20; if one or more of the balls extracted are green, you win nothing.

What would you choose?

- Option A (extract 1 time from a bag with 2 red and 18 green balls, win if red)
- Option B (extract 4 times from a bag with 10 red and 10 green balls, win if 4 times red)
Problem 5

Imagine you were given a choice between two options to win €20.
Option A allows you to extract one ball from a bag containing 4 red and 16 green balls. If you extract a red ball, you win €20; if you extract a green ball, you win nothing.
Option B allows you to extract 6 balls in sequence with replacement from a bag containing 15 red balls and 5 green balls. If all 6 balls extracted are red you win €20; if one or more of the balls extracted are green, you win nothing.

What would you choose?

_____ Option A (extract 1 time from a bag with 4 red and 16 green balls, win if red)
_____ Option B (extract 6 times from a bag with 15 red and 5 green balls, win if 6 times red)

Problem 6

Imagine you were given a choice between two options to win €20.
Option A allows you to extract one ball from a bag containing 6 red and 14 green balls. If you extract a red ball, you win €20; if you extract a green ball, you win nothing.
Option B allows you to extract 2 balls in sequence with replacement from a bag containing 10 red balls and 10 green balls. If all 2 balls extracted are red you win €20; if one or more of the balls extracted are green, you win nothing.

What would you choose?

_____ Option A (extract 1 time from a bag with 6 red and 14 green balls, win if red)
_____ Option B (extract 2 times from a bag with 10 red and 10 green balls, win if 2 times red)
<table>
<thead>
<tr>
<th>Issue</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>n. 3/2005</td>
<td>Luigi Luini Patrizia Sbriglia</td>
<td>Social Learning in Market Games (June 2005)</td>
</tr>
<tr>
<td>n. 4/2005</td>
<td>Roberto Ricciuti</td>
<td>Bringing Macroeconomics into the Lab (December 2005)</td>
</tr>
<tr>
<td>n. 5/2006</td>
<td>Alessandro Innocenti Maria Grazia Pazienza</td>
<td>Altruism and Gender in the Trust Game (February 2006)</td>
</tr>
<tr>
<td>n. 6/2006</td>
<td>Brice Corgnet Angela Sutan Arvind Ashta</td>
<td>The power of words in financial markets: soft versus hard communication, a strategy method experiment (April 2006)</td>
</tr>
<tr>
<td>n. 7/2006</td>
<td>Brian Kluger Daniel Friedman</td>
<td>Financial Engineering and Rationality: Experimental Evidence Based on the Monty Hall Problem (April 2006)</td>
</tr>
<tr>
<td>n. 8/2006</td>
<td>Gunduz Caginalp Vladimira Ilieva</td>
<td>The dynamics of trader motivations in asset bubbles (April 2006)</td>
</tr>
<tr>
<td>n. 9/2006</td>
<td>Gerlinde Fellner Erik Theissen</td>
<td>Short Sale Constraints, Divergence of Opinion and Asset Values: Evidence from the Laboratory (April 2006)</td>
</tr>
<tr>
<td>n. 12/2007</td>
<td>Drift and Equilibrium Selection with Human and Virtual Players (April 2007)</td>
<td></td>
</tr>
<tr>
<td>Volume Number</td>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>15/2007</td>
<td>Nicolao Bonini, Ilana Ritov, Michele Graffeo</td>
<td>When does a referent problem affect willingness to pay for a public good? (September 2007)</td>
</tr>
<tr>
<td>16/2007</td>
<td>Jaromir Kovarik</td>
<td>Belief Formation and Evolution in Public Good Games (September 2007)</td>
</tr>
<tr>
<td>17/2007</td>
<td>Vivian Lei, Steven Tucker, Filip Vesely</td>
<td>Forgive or Buy Back: An Experimental Study of Debt Relief (September 2007)</td>
</tr>
<tr>
<td>18/2007</td>
<td>Joana Pais, Ágnes Pintér</td>
<td>School Choice and Information. An Experimental Study on Matching Mechanisms (September 2007)</td>
</tr>
<tr>
<td>19/2007</td>
<td>Antonio Cabrales, Rosemarie Nagel, José V. Rodríguez Mora</td>
<td>It is Hobbes not Rousseau: An Experiment on Social Insurance (September 2007)</td>
</tr>
<tr>
<td>20/2008</td>
<td>Carla Marchese, Marcello Montefiori</td>
<td>Voting the public expenditure: an experiment (May 2008)</td>
</tr>
<tr>
<td>21/2008</td>
<td>Francesco Farina, Niall O’Higgins, Patrizia Sbriglia</td>
<td>Eliciting motives for trust and reciprocity by attitudinal and behavioural measures (June 2008)</td>
</tr>
<tr>
<td>22/2008</td>
<td>Alessandro Innocenti, Alessandra Rufa, Jacopo Semmoloni</td>
<td>Cognitive Biases and Gaze Direction: An Experimental Study (June 2008)</td>
</tr>
<tr>
<td>23/2008</td>
<td>Astri Hole Drange</td>
<td>How do economists differ from others in distributive situations? (September 2008)</td>
</tr>
<tr>
<td>25/2009</td>
<td>Annamaria Nese, Patrizia Sbriglia</td>
<td>Individuals’ Voting Choice and Cooperation in Repeated Social Dilemma Games (February 2009)</td>
</tr>
<tr>
<td>26/2009</td>
<td>Alessandro Innocenti, Antonio Nicita</td>
<td>Virtual vs. Standard Strike: An Experiment (June 2009)</td>
</tr>
<tr>
<td>27/2009</td>
<td>Alessandro Innocenti, Patrizia Lattarulo, Maria Grazia Pazzienza</td>
<td>Heuristics and Biases in Travel Mode Choice (December 2009)</td>
</tr>
<tr>
<td>29/2010</td>
<td>Valeria Faralla, Francesca Benuzzi, Paolo Nichelli, Nicola Dimitri</td>
<td>Gains and Losses in Intertemporal Preferences: A Behavioural Study (June 2010)</td>
</tr>
<tr>
<td>No.</td>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>---------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>30/2010</td>
<td>Angela Dalton, Alan Brothers, Stephen Walsh, Paul Whitney</td>
<td>Expert Elicitation Method Selection Process and Method Comparison (September 2010)</td>
</tr>
<tr>
<td>31/2010</td>
<td>Giuseppe Attanasi, Aldo Montesano</td>
<td>The Price for Information about Probabilities and its Relation with Capacities (September 2010)</td>
</tr>
<tr>
<td>33/2010</td>
<td>Valeria Faralla, Francesca Benuzzi, Fausta Lui, Patrizia Baraldi, Paolo Nichelli, Nicola Dimitri</td>
<td>Gains and Losses: A Common Neural Network for Economic Behaviour (September 2010)</td>
</tr>
<tr>
<td>34/2010</td>
<td>Jordi Brandts, Orsola Garofalo</td>
<td>Gender Pairings and Accountability Effect (November 2010)</td>
</tr>
</tbody>
</table>