

Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability

By LINDA BABCOCK, MARIA P. RECALDE, LISE VESTERLUND, AND LAURIE WEINGART*

Gender differences in task allocations may sustain vertical gender segregation in labor markets. We examine the allocation of a task that everyone prefers be completed by someone else (writing a report, serving on a committee, etc.) and find evidence that women more than men volunteer, are asked to volunteer, and accept requests to volunteer for such tasks. Beliefs that women, more than men, say yes to tasks with low promotability appear as an important driver of these differences. If women hold tasks that are less promotable than those held by men, then women will progress more slowly in organizations.

* Babcock: Heinz College and Department of Social and Decision Sciences, Carnegie Mellon University, 5000 Forbes Avenue, BP 208, Pittsburgh, PA 15213 (email: lb2k@andrew.cmu.edu); Recalde: International Food Policy Research Institute Markets, Trade, and Institutions Division, 2033 K St NW, Washington, DC 20006 (email: m.recalde@cgiar.org); Vesterlund: Department of Economics, University of Pittsburgh, 4928 WW Posvar. Hall, Pittsburgh, PA 15260 and NBER; Weingart: Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213 (email: weingart@cmu.edu). The first three authors contributed equally to this project. We thank David Tannenbaum, Craig Fox, Noah Goldstein, and Jason Doctor for giving us permission to use their data. We thank David Klinowski, Rachel Landsman, Conor Lennon, and Amanda Weirup for superb research assistance. Participants at seminars and conferences at the University of Zurich, Stanford, Harvard, UCSB, UCLA, CMU, Monash University, University of Adelaide, University of Bonn, University of Pittsburgh, University of Pennsylvania, and Stockholm School of Economics are thanked for their helpful comments. We thank the Carnegie Bosch Institute and the NSF (SES-1330470) for generous financial support. Finally, we are forever grateful to MJ Tocci and Brenda Peyser for inspiring our research on low-promotability tasks. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Despite significant female educational advances, we continue to see gender differences in labor market outcomes (Goldin, Katz and Kuziemko, 2006; Bertrand, Goldin, and Katz, 2010). Particularly striking is the persistent vertical gender segregation (Altonji and Blank, 1999; Bertrand and Hallock, 2001). To better understand the process by which men and women advance in the workplace, researchers have begun to examine whether the tasks that they perform at work vary, and whether such differences contribute to differences in advancement.

Of particular interest is whether relative to men, women spend less time on tasks that are likely to influence their performance evaluations (high-promotability tasks) and more time on tasks that, while benefitting the organization, are less likely to affect their evaluation and career advancement (low-promotability tasks). For example, in industry, revenue-generating tasks may be seen as more

promotable than non-revenue-generating tasks. In research-oriented universities, research-related tasks may be considered more promotable than service-related tasks. We find in a survey of 48 Carnegie Mellon faculty that there is broad agreement that promotion is more likely when more time is spent on research and less time is spent on service. Faculty in our survey were presented with four tasks (working on a research paper, presenting research talks at conferences, serving on an undergraduate curriculum revision committee, and serving on the faculty senate) and were asked to rank them by how an assistant professor should best spend 50 additional hours over a semester to increase the likelihood of promotion. 89.6 percent of those surveyed ranked research paper and conferences as more important than curriculum committee and faculty senate. While the university benefits from service on faculty senate and on curriculum review, individual benefits are more limited. The assignments take time away from research and play a smaller role than research at time of promotion. Faculty who spend more time on service are disadvantaged relative to those who spend less time.

Our faculty task-ranking study suggests that there is consensus about what tasks are more and less promotable, and indeed recent survey evidence from academia show gender differences in the allocation of time on tasks that differ in promotability. In a survey of 349 faculty at the University of Amherst, Misra, Hickes, Lundquist and Templer (2012) find that relative to male faculty, female faculty spent 2.45 fewer hours per week on research. Mitchell and Hesli (2013) find, in a sample of 1,399 political science faculty in the US, that women advised more undergraduate students and participated in more department and college-level committees than men. They conclude that women more than men provide service that while helping the organization, may not help them to advance in their careers. Finally, Porter (2007) finds in the National Survey of Postsecondary Faculty (NSOPF) that female faculty spend 15 percent more hours on committee work than do men.

Task assignments outside of academia are also shown to differ by gender. Benschop and Doorewaard (1998) find that women employees of a large bank performed fewer developmental tasks than men, and Ohlott, Ruderman and McCauley (1994) find, in a sample of professionals, supervisors, middle- and upper-level managers, that women had fewer challenging and developmental opportunities (high stakes, managing diversity and external pressure). In a study of

mid-level jobs De Pater, Van Vianen, and Bechtoldt (2010) show that men, more than women, evaluate their individual task assignments as challenging, and find that these differences partially result from managers being more likely to assign challenging tasks to their male rather than to their female subordinates.

The standard explanations for gender differences in labor market experiences can help explain differences in task allocations. Differences in ability and preferences as well as discrimination may cause men and women to hold a different portfolio of tasks in the work place (e.g., Polachek, 1981; Goldin and Rouse, 2000; Black and Strahan, 2001). Other explanations for gender differences in the portfolio of tasks may be that women are more reluctant than men to negotiate (Babcock and Laschever, 2003) and to compete (e.g., Gneezy, Niederle and Rustichini, 2003; Niederle and Vesterlund, 2007) and thus fail to ‘lean-in’ for tasks with high promotability (Sandberg, 2013).

While recent work has emphasized factors that can distort the allocation of tasks with high promotability, the objective of the present paper is to examine the allocation of tasks with low promotability. Rather than treating low-promotability tasks as the residual of the allocation of more desirable tasks, we examine directly whether men and women differ in their response to requests to perform tasks with low promotability, whether there are differences in the frequency by which men and women face such requests, and what factors likely contribute to such potential differences.

Our interest is in low-promotability tasks for which the worker has some discretion and can decide whether or not to perform the task. Consider, for example, an untenured assistant professor at a research university. She knows that in terms of promotion, the best use of her time is to focus on her research. What will she do when asked by her Dean to serve on a Faculty Senate committee? She knows that this will take a lot of her time, take her away from her research, and is less likely to produce rewards relative to spending time on research. Yet, important faculty matters are debated in the faculty senate and the institution is important to the well-functioning of the University. Will her response differ from that of a male colleague with comparable credentials? Is she more likely than a comparable male colleague to receive such requests?

Gender differences in the frequency of requests and in the acceptance of requests for less-promotable tasks may help explain why women advance at a slower rate than men in the work place.¹ Unless women spend more time at work than do men, working on less-promotable tasks means that they spend less time on more-promotable tasks. The career consequences of accepting a discretionary low-promotability task may, however, extend beyond the opportunity costs of the assignment itself. In particular such assignments may generate lower job satisfaction and in turn reduce the worker's commitment and investment in her job.²

In understanding what may give rise to gender differences in the allocation of low-promotability tasks we examine tasks that individuals prefer be completed yet prefer be completed by others. Avoiding the task leads to a relative advancement. We examine first whether men and women differ in their 'supply' of such tasks and second whether the 'demand' for such tasks differs by gender.³ Examining the response to requests to perform low-promotability tasks we report on field evidence that suggest that women more than men accept such requests. Since this finding may result from men and women having different preferences for the tasks in question, we next explore this gender difference in a controlled laboratory setting. Specifically, we conduct multiple experiments where participants in a group are presented with a task that only one person can undertake. The return from performing the task is such that the individual will only undertake it if no one else is willing to do it. Our design captures the incentives members of a group face when asked to volunteer for a task that each member prefers that another member of the group undertakes (such as writing a report, serving on a committee, organizing an event, etc.)—all group members

¹ As tasks with low promotability are important to the organization one may ask why the completion of such tasks does not warrant greater compensation. Central to the framework we have in mind is that independent of worker skills it is a dominant strategy for workers to ensure that low-promotability tasks are performed. As an example, a volunteer for the IRB will always be found among scholars who conduct experiments, because absent such a volunteer no-one will be able to conduct such research. While the performance on low-promotability tasks may vary, the return to improvements in performance is likely smaller than for high-promotability tasks. Hence with the low-promotability tasks always being provided, in recruiting, retention, compensation, and promotion, greater emphasis will be placed on performance of tasks with high promotability (e.g., publications).

² An example of this phenomenon is provided by Chan and Anteby (2016). They find in a study of TSA employees that, relative to men, women perform more undesirable tasks, have lower levels of job satisfaction, and develop more narrow skills sets. These gender differences contributed to lower rates of promotion, lower pay, and higher rates of turnover for women than for men. Thus the finding that career interruptions and differences in weekly hours explain differences in salaries for male and female MBAs (Bertrand, Goldin, and Katz, 2010) need not imply gender equality if differences in labor market attachment result from differential task allocations.

³ In discussing both the demand and supply of less-promotable tasks we consider tasks for which there is a request and some discretion over the acceptance of the request. The tasks examined fall between those considered in the psychology literature on "organizational citizenship behaviors" (OCBs), where individuals on their own initiate tasks that benefit the organization (e.g., Organ, 1988), and those considered in the organizational psychology literature's examination of task allocation, where the employee must accept an assigned task (e.g. De Pater, Van Vianen, and Bechtoldt, 2010).

want the task to be completed yet the person who undertakes the task is put at a relative disadvantage.

In our first experiment men and women are anonymously paired in groups. All members are treated equally and face the same incentives to perform a task. Despite facing the same incentives we find that women volunteer 50 percent more than men, and we do not find evidence that the differential is explained by individual characteristics such as risk and altruism. To examine the key driver of the response we use an experimental manipulation to simultaneously assess the role of preferences and beliefs. Specifically, in a second experiment we manipulate the gender composition of groups. The study deviates from the first only by having all participants in the lab be of the same sex and thus securing that participants know that they are grouped only with members of their own sex. Results from this same-sex experiment reveal that men and women are equally likely to volunteer. This response to the gender composition of the group shows that the willingness to volunteer is not fixed, and it suggests that the documented gender gap in volunteering likely results from the belief that women are more likely than men to volunteer. Our third experiment further explores the role of beliefs and examines whether women more frequently are asked to volunteer. Specifically, we add an outside requestor to our initial design and charge this requestor with the task of asking one group member to volunteer. Consistent with the belief that women are more likely than men to accept requests, we find that requestors more frequently ask female rather than male group members to volunteer. Confirming this belief, women more than men agree to volunteer when asked to do so. Using two additional experiments we find, as further evidence of the role played by beliefs, that third parties asked to predict behavior in our first experiment anticipate a higher rate of volunteering for women than for men, and that the participants' altruistic preferences are such that they cannot generate the observed difference in volunteering.

The documented gender differences in volunteering and in requests to volunteer for low-promotability tasks likely contribute to gender differences in task allocations, and these differences may create barriers to the advancement of women in organizations and in society as a whole.

I. Response to Volunteer Requests

We begin by examining whether women more than men volunteer to perform tasks with low promotability. Finding field evidence that men and women differ in the frequency by which they volunteer for such tasks, we proceed to examine volunteering behavior in the laboratory where we can control and manipulate the incentives individuals face for volunteering.

A. Field Evidence

Each year a large public University sends an email from the Chair of the Faculty Senate to all faculty members asking them to volunteer to serve on a Faculty Senate committee. As seen in our survey, faculty view service on faculty senate as less promotable than research-related tasks. The responses to these emails are therefore useful in understanding responses to requests to perform low-promotability tasks. Manipulating the email requests, Professors Tannenbaum, Fox, Goldstein and Doctor conducted an experiment to determine how the language in the email affects the probability that a faculty member agrees to serve on a committee. These researchers kindly gave us access to their data for the 2012-2013 academic year, consisting of email requests to a total of 3,271 faculty members, 24.7 percent of whom were female. Faculty responded to the email in one of three ways: did not respond; declined the request; or volunteered to join a committee. As these data contain both the faculty member's response to the email and their demographic characteristics, we can determine whether, when presented with a request to do the same low-promotability task, men and women differ in their likelihood of accepting such requests.

Consistent with the view that service on faculty senate is a low-promotability task, we see across all faculty that only 3.7 percent volunteered to serve, 4.3 percent responded to the email but indicated that they did not wish to serve, and 92 percent ignored the email. There are, however, gender differences in the response. Female faculty are significantly more likely than male faculty to volunteer to be on a committee (7.0 percent versus 2.6 percent, a Fisher's exact test yields $p < 0.001$).⁴ Looking at the results of a probit model for the probability of volunteering, we see in

⁴ Failure to respond is treated as declining the request. For non-volunteers we find that women more than men politely and directly declined the request by return email: 6.1 percent of female faculty and 4 percent of male faculty directly declined the request, $p < 0.05$.

Table 1 that this gender difference is robust to controlling for faculty rank (assistant professor is the excluded category), as well as to controlling for being in the medical school and to being in a STEM related field.

TABLE 1—PROBABILITY OF VOLUNTEERING TO JOIN A COMMITTEE (PROBIT)

	(1)	(2)
Female	0.034 (0.000)	0.034 (0.000)
Associate Professor	-0.005 (0.612)	-0.004 (0.603)
Full Professor	-0.016 (0.063)	-0.015 (0.064)
Emeritus Professor	-0.033 (0.000)	-0.030 (0.000)
Other Rank	-0.016 (0.373)	-0.014 (0.442)
Medical School		0.040 (0.001)
STEM		-0.024 (0.018)

Notes: Dependent variable: Individual decision to volunteer (1-volunteer, 0-don't volunteer). The table presents marginal effects. Assistant professor is the excluded category. Faculty at the medical school are also in a STEM related field. P-values are reported in parenthesis. 3,271 participants.

These gender differences in the likelihood of saying yes to a request to serve on a faculty senate committee translate into higher representation of women on these committees. For the 2012-2013 academic year we find that although women constituted 24.7 percent of faculty they accounted for 37.5 percent of faculty senate committee members at the university.⁵

What is not clear from these field data is why women are more likely than men to accept such requests. One explanation may be that men and women differ in their preferences for performing such low-promotability tasks. Women may simply have a stronger preference for service work such as serving on a faculty senate committee. To understand the differential response to requests we next move to the laboratory where we can better control and manipulate the incentives

⁵ We collected statistics about the gender composition of faculty senate committees from the University's website. The faculty gender composition numbers are those in the data from Profs. Tannenbaum Fox, Goldstein and Doctor. As another example of gender differences in willingness to volunteer to perform less-promotable tasks Weingart, Babcock, Vesterlund, and Weirup (2014) find gender differences in the response to perform a 1-week diary pilot study. Participation can be seen as a less-promotable task. With the endorsement of the Pittsburgh Human Resources Association 539 of their members were emailed a request to participate in the study. 10.3 percent of men and 18.7 percent of women agreed to participate (Fisher's exact test $p < 0.01$).

associated with volunteering and thus can begin to understand why men and women differ in their response to such requests.

B. Are Women More Likely than Men to Volunteer (Experiment 1)?

To study differences in the propensity by which men and women accept requests to perform low-promotability tasks, we conduct a laboratory experiment mirroring the incentives that a small group faces when it is asked to find a volunteer for a task that everyone is reluctant to undertake (writing a report, serving on a committee, planning a holiday party, etc.). The setting we have in mind is one where every member of a committee or group prefers that the task be undertaken, yet everyone prefers that it be undertaken by someone other than themselves. An individual is relatively better off if the task is done by someone else because it allows that individual to spend more time on more-promotable tasks.⁶ With the request for a volunteer being made to the group, every member waits for a volunteer to step forward, fully aware that an excessive delay increases the likelihood that an inferior outcome will result (such as the task not being completed in time or not completed at all). As no explicit request is made of any one individual, the request is implicit and arises through time pressure.⁷

Design

Capturing the incentives described above, our experimental design is as follows. In each of ten rounds participants are randomly and anonymously assigned to groups of three. Members of the group are then given 2 minutes to make an investment (volunteering) decision. Individual earnings are \$1 in the event that no one invests before the end of the 2 minutes. If one group member makes the investment, the round ends, and the individual making the investment secures a payment of \$1.25, while the other two group members each receive \$2. The investor is randomly determined in the event that multiple parties simultaneously invest.

⁶ There are likely also low-promotability tasks that are not individually rational to undertake.

⁷ While delay does not carry a monetary cost it may carry a psychological cost. Bliss and Nalebuff (1984) develop a model with costly delay where individuals decide whether to secure the provision of a binary public good.

With no cost of waiting, investments will be made in the last second of the round and the game reduces to one of simultaneous moves. Accounting for the possibility of ties, the game gives rise to three types of equilibria: A pure strategy asymmetric Nash equilibria where one individual invests and the others do not; a mixed strategy symmetric equilibrium where each player invests 23.2 percent of the time; and a mixed strategy asymmetric equilibrium where one person does not invest and the two others invest 40 percent of the time. Depending on the equilibrium selected the probability that an investment occurs is 100 percent, 54 percent or 64 percent, respectively.⁸

Participants and Procedures

This and other laboratory studies in this paper used a computerized interface (z-Tree, Fischbacher, 2007) and were conducted at the Pittsburgh Experimental Economics Laboratory (PEEL) at the University of Pittsburgh. Participants were recruited from introductory economics classes and were only informed that they would participate in a study on decision making. None of the participants had prior experience with studies at PEEL. The experiment lasted slightly less than an hour. Average earnings from the ten decision rounds were \$16.32.⁹ Nine sessions were conducted, with between 12 and 21 participants per session, for a total of 150 participants (82 males and 68 females). Sessions were roughly gender balanced with the share of women participating in a session ranging between 33 percent and 53 percent.¹⁰ The population was rather homogeneous. The average age was 18.9 years, with 18 and 19 year olds accounting for 77 percent of the participants, 74 percent were Caucasian, 87 percent were born in the US, and 83 percent were either freshmen or sophomores. None of these characteristics differed significantly by gender.¹¹

⁸ The payoff structure corresponds to that of a 3-player game of chicken, Hawk-Dove game (Maynard Smith and Price, 1973), Dragon-Slayer Game (Bliss and Nalebuff, 1984), or Volunteer's Dilemma (Diekmann 1985). Note that any participant's decision to cooperate immediately solves the coordination problem. Existing experimental work of the static volunteer's dilemma focuses on the theoretical predictions that group size decreases volunteering and the probability that no member of the group volunteers (Diekmann, 1993; Franzen, 1995; Goeree, Holt, and Moore, 2005; Healy and Pate, 2009; Murnighan, Kim, and Metzger, 1993). Recent work has also investigated the distribution of player types (Bergstrom, Garrat, and Leo 2015) and mechanisms to allow flexible turn taking (Leo 2014).

⁹ Including a \$6 show-up fee secured average earnings of \$22.32.

¹⁰ The share of females in each session was: 33.3 percent in 2 sessions, 44.4 percent in 2 sessions, 47.6 percent in 1 session, 50 percent in 3 sessions, and 53.3 percent in 1 session. The likelihood that individuals invest is not affected by this degree of variation in gender composition. Clustering on the individual and controlling for round a probit of the individual's propensity to invest on the share of women in the session reveals a marginal effect of -0.031 ($p=0.921$). Furthermore, the marginal effects reported in our central results are not affected by controlling for the share of women in the session (the coefficient on female is 0.113 rather than the 0.111 seen in Table 2, column 1), nor is the coefficient on the session share of women statistically significant (coefficient -0.148, $p=0.615$).

¹¹ The mean age of men and women is not significantly different (18.99 vs. 18.79, two-sided t-test $p=0.319$). Similarly using a Fisher's exact test there is no significant gender difference in the distribution of age ($p=0.716$), race ($p=0.681$), number of years in the US ($p=0.587$), years in college ($p=0.292$) or in choice of major ($p=0.681$).

Upon entering the lab, participants were seated in a pre-marked cubicle, asked to provide informed consent, and given instructions. These instructions were also read aloud and explained all procedures of the study, the payoff structure, the random matching protocol, and what information participants would receive during the study.¹² We then began the ten round decision phase of the experiment. In each round participants were anonymously matched in groups of three, with the stipulation that no one could be paired with the same person twice in a row. Each group member was shown an individual computer screen that displayed the seconds remaining in the round and a button that could be clicked if the individual wanted to invest. The round ended the instant someone in the group clicked the investment button. Participants waited until all groups had either made an investment decision or the two minutes passed without an investment being made. At the end of the ten rounds, participants answered a number of questions to assess individual preferences and personality attributes. A description of the measures and procedures used to elicit them is provided in Appendix C. A demographic questionnaire elicited gender, age, nationality, year in college, and college major. Gender was not mentioned until the very end of the experiment.

Results

To characterize behavior in the experiment we first determine whether groups succeeded in making investments and what the timing was of such investments. We then ask whether men and women were equally likely to make the investment.

Over the course of the ten rounds groups succeed in investing 84.2 percent of the time (std. error =1.63). With two thirds of these investments being made within the last two seconds of a round, participants largely treated the environment as one of a war of attrition.¹³ The likelihood that a group succeeded in making an investment decreased from 88.4 percent during the first five rounds to 80 percent during the last 5 rounds. The per round decrease in the investment rate is significant.¹⁴

¹² See Appendix A for instructions and sample decision screens of all experiments.

¹³ The share of investments made in the last two seconds of a round increases from 57.9 percent during the first half of the experiment (Rounds 1-5) to 83 percent during the second half (Rounds 6-10). Appendix Table B1 shows the distribution of decision times in Experiment 1.

¹⁴ Treating the group as the unit of observation, a probit regression of the probability that a group invests on round number yields a marginal effect of -0.019, $p=0.001$.

With four out of five groups securing an investment, the group's rate of success exceeds that predicted by the symmetric mixed strategy Nash equilibrium success rate of 54 percent. This raises the question of whether asymmetric play better characterizes behavior and whether certain members of the group more frequently make the mutually beneficial investment. Of particular interest is whether the likelihood that the individual invests differs by gender. Figure 1 reveals that women are systematically more likely than men to undertake the investment.¹⁵ Starting in round one the investment rate by women surpasses that of men. With a sustained differential over the ten rounds, this results in a substantial difference in the total number of times men and women invest over the ten rounds. Over the ten rounds women on average invest 3.4 times, whereas men invest 2.3 times. This 48 percent difference in total investment is statistically significant (two-sided t-test $p=0.003$). Figure 2 shows the distribution of total investments aggregated over the ten rounds by gender. The distribution for women first order stochastically dominates that for men, and the difference between the two is significant (a Kolmogorov-Smirnov test yields $p=0.020$).¹⁶ While 61 percent of men invest two or fewer times, only 40 percent of women fall in this lower investment range.

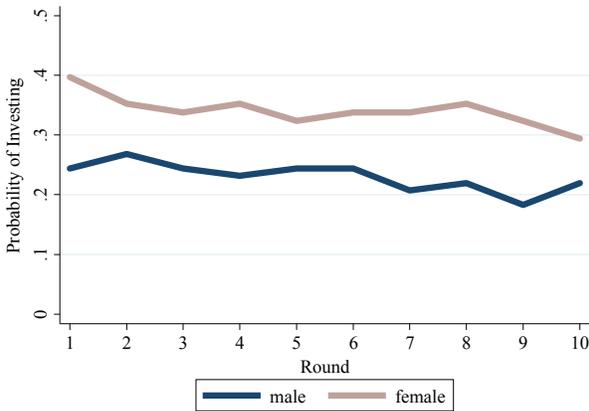


FIGURE 1: PROBABILITY OF INVESTING (EXP 1)

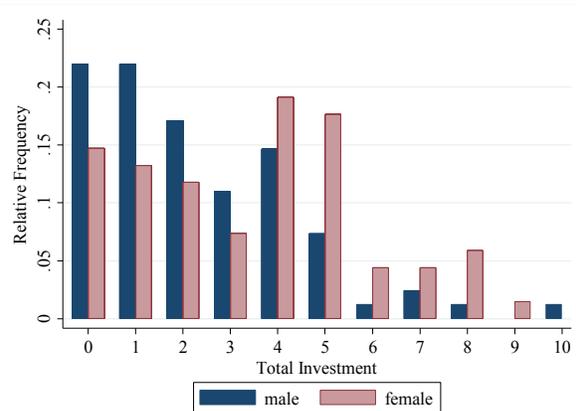


FIGURE 2: DISTRIBUTION OF TOTAL INVESTMENT (EXP 1)

¹⁵ As we only see the individual making the investment we cannot compare the individual investment rate with the equilibrium prediction.

¹⁶ A Fisher's exact test yields $p=0.159$. The median contribution is 2 for men and 4 for women (Wilcoxon Mann-Whitney rank-sum $z=-2.929$, $p=0.003$).

In Table 2 we estimate the probability that men and women invest in a given round. Standard errors are clustered on the individual to account for the repeated nature of decisions. The reported marginal effects confirm the insights from Figure 1 and 2. Pooling the data from all ten rounds we see in column 1 that participants become less likely to invest over the course of the experiment and that women are significantly more likely to invest than are men. The average investment rate for men is 23 percent and that for women is 11 percentage points higher. Columns 2 and 3 confirm that these results hold both for the first and second half of the experiment.¹⁷

TABLE 2 — PROBABILITY OF INVESTING (PROBIT), EXPERIMENT 1

	All rounds (1)	Rounds 1-5 (2)	Rounds 6-10 (3)	All rounds (4)
Female	0.111 (0.003)	0.107 (0.005)	0.115 (0.013)	0.090 (0.036)
Round	-0.006 (0.052)	-0.009 (0.407)	-0.009 (0.350)	-0.006 (0.050)
Non-conformity				-0.021 (0.399)
Risk-seeking				-0.018 (0.354)
Altruism				0.023 (0.377)
Agreeable				-0.014 (0.661)
N	1500	750	750	1500

Notes: Dependent variable: Individual investment decision (1-invest, 0-don't invest). The table presents marginal effects. Standard errors are clustered on the individual. P-values are reported in parentheses. 150 participants.

C. Why Does the Rate of Investment Differ by Gender?

There are two potential explanations for the gender difference in the investment rate. First, gender may be a proxy for individual preferences and these differences may cause women to invest more than men. For example, women may be more likely than men to agree to requests to perform non-promotable tasks if they are more other-regarding and more concerned for the welfare of others

¹⁷ Including a (female x round) interaction reveals a small and insignificant decrease in the gender gap over the course of the experiment. The marginal effect of the interaction term, corrected to account for the non-linear nature of the estimation, is -0.0005 (p=0.944) in the column-1 specification. Appendix Table B2 shows similar results when including session dummies. We do not lead with these specifications because session dummies cannot be included for our single-sex sessions of Experiment 2 given that they perfectly correlate (by design) with gender. We do not cluster standard errors at the session level either, because the number of clusters (sessions) is small. Score boot-strapped tests (Kline and Santos 2012) conducted for all models presented in the paper that cluster standard errors at the session level show similar results. Note also that results are similar when including round dummies instead of having round enter linearly in the model, and when analyzing only behavior in round 1.

(e.g., Eckel and Grossman, 1998; Andreoni and Vesterlund, 2001), if they are more agreeable and have a greater desire to be liked by the requestor (Braiker, 2002), if they have a greater desire to conform to a norm of accepting such requests (e.g., Santee and Jackson, 1982; Eagly, Wood, and Fishbaugh 1981), if they are more risk averse (e.g., Eckel and Grossman, 2008) and more concerned about the consequences from declining the request (Heilman and Chen, 2005).¹⁸ Second, differences in beliefs about whether others will invest can cause women to invest more than men. For example, both men and women may believe that women are more likely to invest than men.

Gender Differences in Preferences

We first explore whether there is direct evidence that preferences help to explain the gender gap in investment rates. After the experiment, we asked participants a series of questions to elicit measures of risk aversion, altruism, agreeableness, and non-conformity (see Appendix C for the full set of items used to measure these constructs and for the correlation between these variables and the outcome of interest). While finding significant gender differences in some of these measures, we see in Table 2 (column 4) that these additional controls do not eliminate the gender difference in investment rates. The coefficient on female remains large and statistically different from zero.¹⁹

Of course, the results in column 4 do not conclusively rule out that gender is standing in for individual differences in preferences and attitudes. For example, our included variables could be measured imprecisely or we might have omitted important factors in our empirical model.²⁰ To examine what gives rise to the gender gap in investments we therefore rely on an experimental manipulation for identification. By slightly changing the experiment we can simultaneously assess whether gender differences in investment is explained by gender differences in preferences and/or by beliefs.

¹⁸ See Croson and Gneezy (2009) and Niederle (2016) for reviews on gender differences.

¹⁹ Including controls for participant age, race, year in school and whether they were born in the U.S. leaves the coefficient on female virtually unchanged, the marginal effect is 0.099 ($p=0.020$).

²⁰ Our elicitation of risk and altruism was improved in Experiments 3 and 4 to mirror the incentives of investing. These improved measures are predictive of the decision to invest when gender is controlled for, but do not explain the gender gap (see Appendix Table B9, which shows the results of a 5-round replication of Experiment 1).

Beliefs: Does the Response to Requests Depend on the Gender Composition of the Group (Experiment 2)?

To manipulate beliefs we conducted a single-sex version of Experiment 1.²¹ That is, instead of inviting close to equal proportions of men and women to our laboratory for each session, we conducted sessions where only men or only women participated. The purpose of the single-sex experiment was to determine whether gender differences in investments are robust to the group's gender composition or if they are instead influenced by beliefs. If gender differences in investing are caused by women being more conforming, more altruistic and more risk averse, then we would expect to see higher investment rates in all-female session than in all-male sessions.²² However, if behavior is influenced by a belief that women more frequently invest than men, then we would expect the individual rate of investments to change in single-sex groups. In particular, when a man moves from what is likely a mixed-sex trio to knowing that his fellow group members are all men, he will see his decision to invest as more critical and will increase the probability that he invests. By contrast when a woman moves from what is likely a mixed-sex trio to knowing that her fellow group members are all female, she will see her decision to invest as less critical and will decrease the probability that she invests.

The recruitment method, instructions, and procedures of these single-sex sessions were identical to those of the mixed-sex sessions of Experiment 1. 117 undergraduate students were recruited from introductory economics classes at the University of Pittsburgh and the characteristics of this pool of participants were similar to those of Experiment 1. Participants were on average 18.7 years old, with 18 and 19 year olds accounting for 83 percent. 74 percent were Caucasian, 91 percent were born in the US, and 85 percent were either freshmen or sophomores. We conducted three sessions with all women (n=66) and three sessions with all men (n=51). Sessions consisted of

²¹ Examining coordination in the battle-of-the-sexes game, Holm (2000) points to the role of a gender-based focal point. He finds that when males and females are paired with a female, they are more likely to select the action associated with their preferred equilibrium. This 'hawkish' behavior implied that coordination more frequently was achieved in mixed-sex pairing – thus securing greater efficiency. In our setting the investment can only be made by one individual, and coordination issues are resolved when investing. Thus the decision not to invest cannot be justified by it reducing the chance of mis-coordination, and we do not find greater efficiency in mixed-sex groups.

²² In our first experiment participants did not know the gender composition of the group they were in. However, with women being more likely to invest than men we find an insignificantly larger chance that investments are secured in all-female groups than in groups with at least one male (All-female mean =0.935, Not-all-female mean =0.836, two-sided Fisher's exact test p=0.202).

between 15 and 24 participants. With ten rounds we have a total of 390 group decisions.²³ Following the procedures of Experiment 1 gender was not mentioned until the survey at the end of the experiment. Although gender was not mentioned, by looking around the room participants could see the population from which members of their anonymously drawn groups would be drawn. Hence, the gender composition of the room could influence the participants' beliefs about the likelihood that other members of their group would invest.

The results are shown below. Pooling first the data from the all-male and all-female sessions we find aggregate behavior and investment times very similar to those of Experiment 1. Compared to the 84.2 percent investment rate in the mixed-sex sessions of Experiment 1, we find an investment rate of 80.8 percent in the single-sex sessions of Experiment 2. The timing of the investments is also similar (see Appendix Table B3). Investments were primarily made within the last two seconds of a round, with this share increasing from 69.3 percent during the first half of the experiment (Rounds 1-5) to 82.7 percent during the second half (Rounds 6-10). Furthermore, over the course of the experiment the likelihood that a group invests decreases from 90.3 percent in the first half to 71.3 percent in the second half. A probit regression reveals that this decrease in investments over time is significant.²⁴

Next we ask whether these aggregate results mask differences between the all-female and all-male groups. Of interest is whether the gender gap in the probability of investing is sensitive to the single-sex setting. Figure 3 displays by gender the probability that an individual invests in a given round. In sharp contrast to our results from Experiment 1 we do not find that women are more likely to invest than men. The average number of investments over the ten rounds does not differ by gender (mean men = 2.67, mean women = 2.71, two-sided subject-level t-test $p=0.918$).²⁵ When the gender composition of the group is known to the participants there is no evidence that all-female groups fare better than all-male groups. The success rate is 81 percent for female groups

²³ Average individual earnings from the ten decision rounds were \$16.06. Including a \$6 show up fee secured average earnings of \$22.06. The mean age of men and women is not significantly different (18.75 vs. 18.68, two-sided t-test $p=0.761$). Similarly, using a Fisher's exact test there is no significant gender difference in the distribution of age ($p=0.951$), race ($p=0.879$), number of years in the US ($p=0.641$), years in college ($p=0.878$), or in choice of major ($p=0.112$).

²⁴ Treating the group as the unit of observation a probit of the likelihood that a group invests on round yields a marginal effect of -0.038 ($p<0.001$).

²⁵ The median contribution is 2 for men and 3 for women (Wilcoxon Mann-Whitney rank-sum $z=-0.198$, $p=0.843$).

and 80 percent for male groups. Figure 4 shows the distributions of total investments aggregated over the ten rounds by gender.²⁶

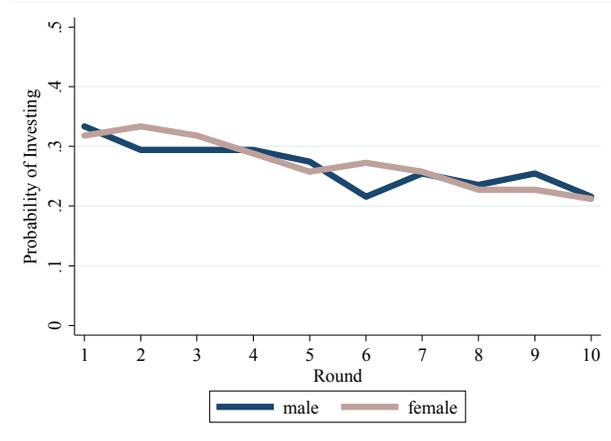


FIGURE 3: PROBABILITY OF INVESTING (EXP. 2)

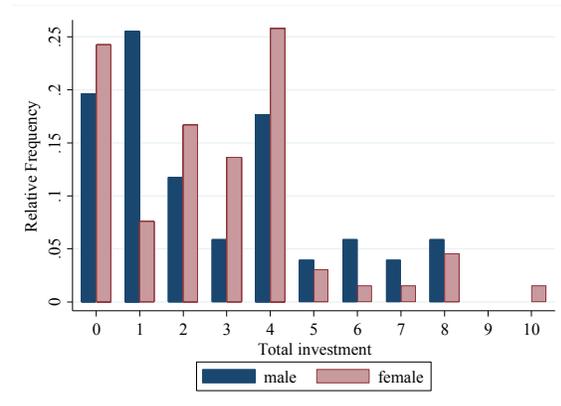


FIGURE 4: DISTRIBUTION OF TOTAL INVESTMENT (EXP. 2)

Verifying the results of Figure 3 we see from the probit models in Table 3 that in single-sex sessions the decision to invest does not depend on the sex of the participant. Across all ten rounds, the first five rounds, and the last five rounds, we find that men and women are equally likely to invest.²⁷ The coefficient on female is small in magnitude and not significantly different from zero. When the four measures of individual characteristics are included in the regression (risk aversion, altruism, agreeableness, and non-conformity) the coefficient on female remains insignificant (column 4).²⁸

²⁶ A Fisher's exact test for equality of distributions yields $p=0.158$. A two-sided variance test provides $p=0.687$.

²⁷ Treating the group decision as the unit of observation we see in Experiment 2 that controlling for round a probit of group investment on all female generates a marginal effect of 0.016 ($p=0.680$).

²⁸ Similar results are obtained when demographic variables such as race, year in school, and whether the participant was born in the US are included in the regressions. The marginal effect of gender is 0.002 ($p=0.968$). The results are also similar when analyzing only behavior in round 1.

TABLE 3—PROBABILITY OF INVESTING (PROBIT), EXPERIMENT 2

	All rounds (1)	Rounds 1-5 (2)	Rounds 6-10 (3)	All rounds (4)
Female	0.004 (0.920)	0.005 (0.916)	0.004 (0.935)	-0.002 (0.962)
Round	-0.012 (0.000)	-0.015 (0.248)	-0.009 (0.389)	-0.013 (0.000)
Non-conformity				0.038 (0.192)
Risk-seeking				-0.050 (0.041)
Altruism				0.035 (0.288)
Agreeable				0.020 (0.614)
N	1170	585	585	1170

Notes: Dependent variable: Individual investment decision (1-invest, 0-don't invest). The table presents marginal effects. Standard errors are clustered on the individual. P-values are reported in parentheses. 117 participants.

The finding that there is no gender difference in the probability of investing when participants make decisions in single-sex groups helps explain why differences in individual characteristics did not explain the gender difference in investing seen in the mixed-sex experiment. The changes in investment rates between Experiment 1 and 2 suggest that the individual's behavior is not caused by fixed preferences but instead depends upon the population from which group members are drawn. This response to the gender of the other group members is consistent with beliefs about the investment rates of men versus women playing a central role when deciding whether to volunteer. Figure 5 summarizes the results of the two group compositions in Experiments 1 and 2. The group's rate of success is independent of the group being drawn from a single- or mixed-sex population, and in the single-sex sessions it is independent of whether the group is all-male or all-female.²⁹ As demonstrated in Figure 6 the individual's propensity to invest is sensitive to the group's gender composition. The dashed-lines, *single_m* and *single_f*, refer to the all-male and all-female sessions (Experiment 2) respectively. The solid-lines, *mixed_m* and *mixed_f*, refer to males and females in the mixed-sex sessions (Experiment 1) respectively. Relative to the mixed-sex

²⁹ The marginal effect of a probit of group investment on single-sex that controls for round provides a coefficient on single sex of -0.031 (p=0.217). The similar provision rates in the mixed- and single-sex sessions suggest that the completion of low-promotability tasks is unlikely to contribute to performance differences that may arise as a result of gender diversity. For examples of work on gender diversity and performance see Ali, Kulik, and Metz (2011), Bear and Woolley (2011), and Woolley and Malone (2011).

results, the investment rate decreases for women and increases for men in single-sex sessions. While women are significantly more likely than men to invest in mixed-sex sessions, this gender gap is eliminated in single-sex sessions.³⁰

Our results are consistent with the predictions of evolutionary game theory, as it predicts in a hawk-dove game that a symmetric equilibrium will be selected within a single population, whereas an asymmetric equilibrium is predicted when the players are drawn from different populations (Maynard Smith, 1982). Oprea, Henwood, and Friedman (2011) confirm this prediction.³¹ While the group investment rate observed in our study does not correspond to that predicted in either of the three possible equilibria, it is easily reconciled with a model where individuals have other regarding preferences.³²

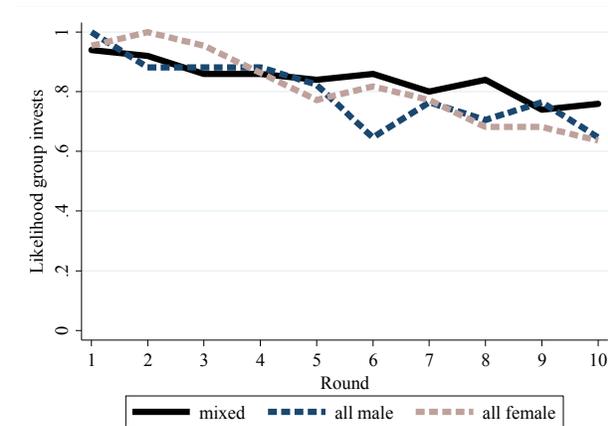


FIGURE 5: PROBABILITY GROUP INVESTS

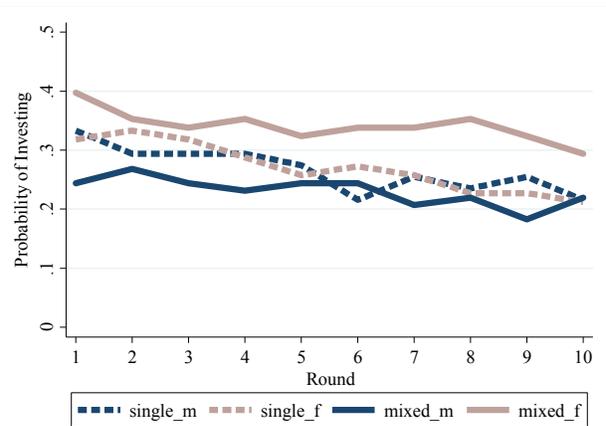


FIGURE 6: INDIVIDUAL PROPENSITY TO INVEST

³⁰ Clustering standard errors on the individual and controlling for round, a probit of investing on dummies for female and single-sex treatment and their interactions secures marginal effect on female of 0.109 ($p=0.003$), on single-sex of 0.038 ($p=0.375$), on the female-single-sex interaction of -0.106 ($p=0.066$), and on round of -0.009 ($p=0.000$). The coefficient and standard error on the interaction is corrected to account for the nonlinear nature of the estimation. In another set of analysis, there is no significant gender difference in mean decision time. Clustering the standard errors on the individual an OLS regression of remaining decision time on female and round reveals a coefficient on female of -5.266 ($p=0.532$) in Experiment 1 and of -6.671 ($p=0.272$) in Experiment 2.

³¹ Oprea, Henwood, and Friedman (2011) examine investments in a two-person Hawk-Dove game and find that play converge to the symmetric mixed Nash equilibrium under a one-population matching protocol, while it moves toward an asymmetric and inequitable pure strategy Nash equilibrium when the participants are assigned either to be row or column players and interact in a two-population environment. As we do not elicit individual strategies we are not able to determine whether play converges to a particular equilibrium.

³² Letting π_s , π_1 and π_2 , denote payoff to self and to the two other group members respectively, assuming preferences of the form $\pi_s + 0.148(\pi_1 + \pi_2)$ predicts a group investment rate of 80 percent in a symmetric mixed strategy equilibrium, and of 86 percent in an asymmetric mixed strategy equilibrium, where only two of the three group members invest. The literature on public good provision suggests that concern for payoffs to others is easily this large.

The results regarding gender differences in investing in mixed- and single-sex groups are intriguing. They document that men and women differ in the propensity by which they agree to implicit requests in mixed-sex groups, and provide insights into what drives these differences. When moving from a mixed-sex environment to a single-sex one, women see a decrease in the need to volunteer, while men see an increase. This change in behavior suggests that beliefs rather than preferences drive the gender difference documented in Experiment 1.

II. Are Women Asked to Volunteer More Than Men (Experiment 3)?

The studies of how faculty spend their time indicates that women are spending more time on service tasks than men. While it is possible that Deans and Department Heads believe women to be more skilled at these tasks than men, an alternative explanation, suggested by our two first experiments, is instead that they may believe that women are more likely than men to accept requests to do these assignments. In a laboratory experiment we can examine whether differences in requests arise when there are no gender differences in ability. This examination of the demand side allows us to further assess whether beliefs are likely to drive the gender difference in volunteering. We conduct a third experiment where we extend our design to include an outside requestor, who after seeing pictures of the three group members must ask one of them to invest. This outside requestor has the incentive to ask the person he or she believes is most likely to accept the request. Based on our findings in the first two experiments, we expect that participants will be more likely to ask a woman than a man to volunteer.

A. Design

We modify our experimental framework to study the demand side of task allocation. We add a fourth group member, a requestor, who prior to the two minute investment round is charged with asking one of the three members of the investment group to invest. While unable to personally invest, the requestor benefits from the group's investment. The requestor receives \$1 if no one invests and receives \$2 if any member of the investment group invests before the end of the two minute round. The choices and payoffs of the three members of the investment group are, as in our

first two experiments, \$1 if no one invests and if one person invests then the investor receives \$1.25 while non-investors receive \$2.

B. Participants and Procedures

We conducted four sessions of Experiment 3. With 20 participants per session we have choices from a total of 80 participants (37 males and 43 females). Sessions were roughly gender balanced with the share of women participating in a session ranging between 50 and 60 percent. Participants were recruited from introductory classes in the social sciences (economics, political science, and anthropology) and none of the participants had prior experience with studies at PEEL.³³ The average age was 19.4 years, with 18 and 19 year olds accounting for 68 percent of participants, 74 percent were Caucasian, 94 percent were born in the US, and 76 percent were either freshmen or sophomores. With the exception of age none of these characteristics differed significantly by gender.³⁴

To test whether there are differences in the likelihood of asking a female versus a male participant to invest, photos were secured of each participant. As participants entered the lab they were asked to provide informed consent, photos were then taken one by one on the count of three and participants were reseated at pre-marked cubicles. We then proceeded with instructions.

Participants were informed that in each round, one person of the 4-person group would be designated the role of “red” player while three individuals would be designated the role of “green” players and would form an investment group. The red player was unable to make the investment but was charged with asking one member of the investment group to invest. Requests by red players were solicited using the strategy method. After assignment to a group of four, and before learning who was assigned the role of red or green player, each participant was shown the photos of the potential investment group and was asked which member they would ask to invest in the

³³ Note that Experiment 3 is an examination of the demand side and an additional test of the role of beliefs. Since the design of Experiment 3 differs from that of Experiments 1 and 2, we never compare behavior between the two types of experiments. Small differences in participant characteristics between the two sets of experiments also caution against comparing behavior in the two sets of studies.

³⁴ The mean age of men is significantly larger than for women (19.89 vs. 18.95, two-sided t-test $p=0.005$). There are minor, but not significant differences in other characteristics. Fisher’s exact tests reveal no significant gender difference in the distribution of age ($p=0.168$), race ($p=0.148$), number of years in the US ($p=0.343$), years in college ($p=0.131$) or in choice of major ($p=0.936$). Average earnings from the ten decision rounds were \$17.60. Including a \$6 show up fee and two incentivized preference elicitation tasks secured average earnings of \$26.67.

event that they were assigned the role of a red player. Each member simultaneously saw the photos of the potential investment group members, with the order of photos being different for each participant. Once all four members of the group submitted their requests, photos of the four group members appeared on the screen indicating the selected role and associated request. The photo of the red player appeared at the top of the screen, photos of the three green players appeared in a row at the bottom of the screen clearly indicating which green player was asked by the red player to invest. The green player who was asked to invest saw a message below his or her photo stating “The red player asked you to invest,” while the two other green players saw the photos with the message “The red player asked this group member to invest” below the photo of the green player who was asked. After all participants acknowledged the role assignment and request, the two-minute investment period began and any green player could invest—both the green player who was asked and the other green players who were not asked to invest. Mirroring Experiments 1 and 2, participants learned if an investment was secured, but did not learn who invested.

For each of the ten rounds, participants were matched in groups of four, using an algorithm to ensure that participants were not matched with the same individuals during the first five rounds. While participants for rounds 6-10 can be characterized as ‘non-strangers’ it is important to keep in mind that they never learn the investment decisions of any other participants.

Following the ten rounds of decision making individual characteristics were elicited. We changed the incentivized risk elicitation task to closely parallel the type of risk participants experience in the investment task. We also elicited an incentivized measure of altruism that asks participants to make 6 binary decisions between two payoff distributions for group members that reflect the types of payoffs that participants experience in the investment task (see Appendix C for details). Finally individuals were asked to indicate whether they knew any of the other participants in the lab. Seeing photos of all the participants in the session they first had to click on photos of the people they knew or had seen before, then from this subset indicate whom they had communicated with, and from this remaining subset whom they were friends with.³⁵

³⁵ 52.5 percent of subjects knew at least one other participant in the session. The mean number of participants individuals report knowing, having communicated with, and being friends with is 1.34, 0.13, and 0.11 respectively.

C. Results

Investments occurred rather quickly in Experiment 3: 33.2 percent of investments were made in the first ten seconds and only 40.1 percent were made in the last two seconds of the round. The share of early investments is not sensitive to whether group members are ‘strangers’ (rounds 1-5 where participants had not yet been paired with each other) or ‘non-strangers’ (rounds 6-10 where participants had been paired with each other in previous rounds). With 93.5 of groups succeeding in investing in Experiment 3 we find a high level of coordination.³⁶

The central question of Experiment 3 is whether men and women are equally likely to receive requests from a requestor. That is, when a potential requestor sees photos of the three potential investment group members, does a group member’s gender help predict whether they are asked?

In using the strategy method to elicit requests, over the ten decision rounds an individual can at most receive 30 strategy-method requests. Our data reveal substantial heterogeneity in the number of requests participants receive, ranging between 1 to 16 requests. Consistent with participants holding the belief that women are more likely to invest than men, an individual’s gender is predictive of the number of requests. Figure 7 shows how the distribution of the total number of strategy-method requests for women first order stochastically dominates that for men (Kolmogorov-Smirnov $p=0.029$). The mean and median number of requests for men are 8.7 and 9 respectively, while for women they are 11.1 and 12, respectively (two-sided t-test $p<0.01$; Wilcoxon Mann Whitney Rank-sum test $p<0.01$).

³⁶ Appendix Table B4 shows the distribution of investment times in Experiment 3.

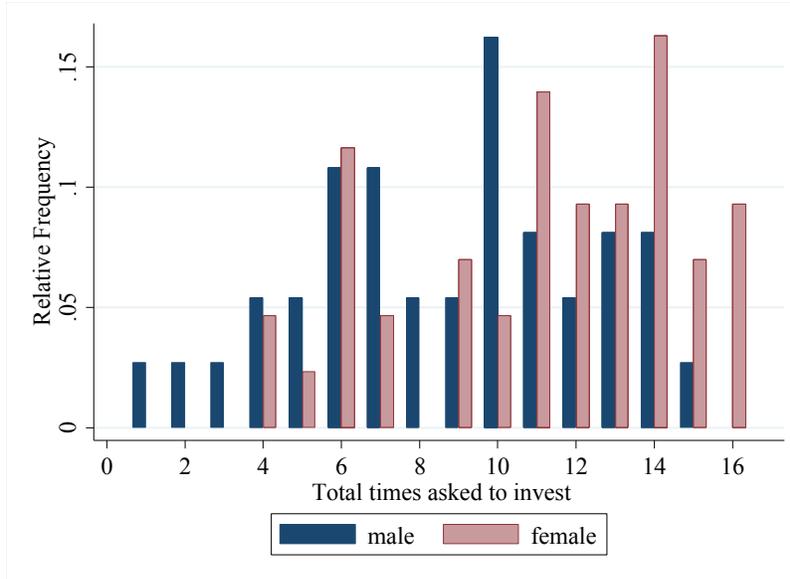


FIGURE 7: DISTRIBUTION OF STRATEGY-METHOD REQUESTS PER INDIVIDUAL (EXP. 3)

The differences in the number of requests received by men and women are confirmed in Table 4 when we control for other observable characteristics. Across all rounds gender is predictive of the number of requests an individual receives. Controlling for observable characteristics such as being Caucasian and being someone whom participants in the lab are familiar with, we find that on average females receive 2.5 more requests than males over the ten rounds of the experiment.³⁷ This difference in requests increases over the course of the experiment. Females receive on average 0.9 more requests than males during the stranger rounds (rounds 1-5), while they receive 1.6 more requests during the non-stranger rounds (rounds 6-10).

³⁷ The results are similar when analyzing only behavior in round 1. Given the small age variation and participant’s likely limited ability to distinguish participants by age we do not control for age. Adding an age control does not significantly affect the coefficient on the female dummy. Results also do not change when we include individual measures of preferences and personality attributes as controls (see Appendix Table B5). The effect of gender is similar if we instead control for the number of people who reported that they know or are friends with the participant. Results from multinomial probits that simultaneously control for the characteristics of all group members who could be asked to invest are provided in Appendix Table B6.

TABLE 4—REQUESTS RECEIVED VIA THE STRATEGY METHOD (OLS), EXPERIMENT 3

	Rounds 1-10		Rounds 1-5		Rounds 6-10	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	2.476 (0.070)	2.521 (0.070)	0.863 (0.214)	0.889 (0.214)	1.613 (0.070)	1.631 (0.070)
Non-Caucasian	-1.307 (0.114)	-1.196 (0.222)	-0.677 (0.114)	-0.610 (0.114)	-0.630 (0.222)	-0.585 (0.222)
N communicate with in session		2.880 (0.000)		1.709 (0.122)		1.171 (0.000)
Constant	9.012 (0.000)	8.599 (0.000)	4.714 (0.000)	4.469 (0.000)	4.298 (0.000)	4.130 (0.000)
N	80	80	80	80	80	80

Notes: Dependent Variable: Total requests received. N communicate with refers to the number of subjects who reported the subject as someone they communicate with. P-values are in parentheses. Standard errors are clustered at the session level using wild bootstrapping procedures that test the null hypothesis that the coefficient on female equals zero. 80 participants.

Another way of summarizing the differences in requests is to examine mixed-sex groups where both men and women could be asked to invest. If men and women are equally likely to be asked then they should be asked a third of the time in mixed-sex groups. Instead we see a 39 percent chance that a woman is asked and a 27 percent chance that a man is asked. This 12 percentage point difference is statistically significant across all rounds and for the first and second half of the experiment (stranger and non-stranger rounds).³⁸

The finding that women are asked to invest more than men is consistent with the belief that they are more likely to accept such requests. As evidence that such beliefs are commonly held we find that both male and female requestors are more likely to ask female rather than male group members to invest. When a male requestor has the option of asking either a man or woman to invest, a woman is asked 39 percent of the time and a man is asked 29 percent. The choice made by a female requestor is similar as she asks a woman 39 percent of the time and a man 26 percent of the time.³⁹

Requests have a substantial impact on the individual's decision to invest. Participants who are asked to invest have an investment rate of 65.5 percent, while the investment rate is only 14 percent for those who are not asked to invest. Thus a requestor can, by asking, significantly increase the

³⁸ Two-sided session-level paired t-test $p=0.014$ for rounds 1-10. Differentiating between stranger and non-stranger rounds gives $p=0.109$ and 0.070 , respectively.

³⁹ Controlling for round and clustering standard errors at the individual level a probit of asking a female to invest on the gender of requestor provides a marginal effect on female requestor of 0.024 ($p=0.627$).

likelihood that an individual invests.⁴⁰ Absent a request, the investment rate does not differ by gender (both men and women invest 14 percent of the time). The response to a request, however, does differ by gender, as the investment rate becomes 51 percent for men and 76 percent for women. The probit regression in Table 5 confirms that the response to requests is significant and that the likelihood of agreeing to the request is significantly greater for women than for men. This result holds both for the first and second half of the experiment.⁴¹ Using the improved measures for risk aversion and altruism, we find that both measures predict the likelihood of investing, but do not reduce the gender difference in investing (see columns 3 and 6 of Appendix Table B7). The net result of women more frequently being asked to invest and more frequently accepting such requests is of course that the investment rate for women exceeds that for men. The aggregate investment rate is 37 percent for women and 25 percent for men.⁴²

TABLE 5— PROBABILITY OF INVESTING (PROBIT): ALL GREEN PLAYERS, EXPERIMENT 3

	All rounds (1)	Rounds 1-5 (2)	Rounds 6-10 (3)
Asked to invest	0.382 (0.000)	0.402 (0.000)	0.368 (0.000)
Female	-0.010 (0.882)	0.082 (0.283)	-0.108 (0.226)
Female X asked to invest	0.258 (0.005)	0.245 (0.043)	0.270 (0.007)
Round	-0.0032 (0.643)	-0.006 (0.769)	-0.010 (0.572)
N	600	300	300

Notes: Dependent variable: Individual investment decision (1-invest, 0-don't invest). Marginal effects presented in the table. Standard errors clustered on the individual. P-values reported in parentheses. The coefficient and standard error on the interaction terms is corrected to account for the nonlinear nature of the estimation. 80 participants.

⁴⁰ A probit of investment that controls for round and clusters standard errors at the individual level indicates that being asked to invest increases the probability of investment by 0.515 ($p < 0.001$).

⁴¹ The results are similar when analyzing only behavior in round 1. With only 20 realized requests per round we are however underpowered, and while the coefficient on the female-being asked interaction is the same the coefficient is not significant (coefficient 0.260, $p = 0.280$). Specifications that interact female with all covariates, include session dummies, and control for individual preferences and personality attributes are presented in Appendix Table B7.

⁴² A probit regression of investment on round and gender with standard errors clustered at the subject level provides a marginal effect on female of 0.116 ($p = 0.022$).

III. Explaining the Differential Response by Women and Men: Altruism and Beliefs

Our results demonstrate that, in a mixed-sex group, women more than men perform tasks that, while benefitting the collective, place them at a relative disadvantage. The findings in our three experiments are consistent with this gender difference resulting from a commonly held belief that women more than men will agree to perform such tasks. Such a belief is consistent with the higher investment rate for women than men in the mixed-sex groups (Experiment 1); with women investing less and men investing more when moving from mixed-sex to single-sex groups (Experiments 1 and 2); and with women more than men receiving requests to invest and more frequently accepting such requests (Experiment 3).

While our results align well with the belief that women more than men will invest, we do not directly demonstrate that beliefs differ, nor do we fully rule out that the gender differences in behavior result from altruistic preferences.⁴³ To explore the role of beliefs and altruism we conduct two additional experiments (Experiments 4 and 5). Recruitment and general procedures mirrored those of our initial experiments. 189 undergraduates were recruited from introductory economics classes at the University of Pittsburgh. They were on average 18.5 years old, 90 percent were 18 and 19 years old, 75 percent were Caucasian, 86 percent were born in the US, and 93 percent were either freshmen or sophomores.

A. Experiment 4: Is the Investment Rate for Women Believed to be Higher Than it is for Men?

Experiment 4 was used to explore the role of beliefs. To confirm that beliefs drive investment choices in Experiment 1 we would ideally want to elicit participants' beliefs as they make decisions in the same environment as in Experiment 1. However, as investment decisions are made in real time with the round ending and all uncertainty being resolved once an investment is made, a strategy method would be needed to simultaneously elicit beliefs and behavior. Such a modification to our design is likely to affect behavior (see e.g., Croson, 2000). We instead asked third-party participants to predict behavior in a representative session from Experiment 1.

⁴³ While the results from Experiments 1 and 2 are not consistent with altruism varying between men and women, they are consistent with men and women being more altruistic toward men than women (see Section III.B).

We conducted 7 sessions of Experiment 4 with 21 participants in each session. Participants were asked to perform two different tasks, with the instructions for each task only being given immediately preceding the task. Task 1 was a five-round version of Experiment 1 and aimed to familiarize the participants with the strategic environment. Task 2 asked participants to predict the outcomes in a previous session of Experiment 1. Participants were informed that another session of 21 individuals had made decisions in a 10-round version of the preceding Task 1, and they were asked to predict, for ten groups of three, how likely it was that each member of the group invested. After participants completed these two tasks, we collected measures of individual characteristics, including improved measures of risk aversion and altruism that mirror the incentives associated with investing (see Appendix C).

With the exception of the smaller number of rounds the procedures for Task 1 were identical to that for Experiment 1, and thus should generate similar results. This is precisely what we find. On average 92.7 percent of Task-1 groups succeed in investing, and 63.4 percent of these investments occur in the last 2 seconds.⁴⁴ Importantly, the individual investment rates differ by gender. Consistent with our Experiment 1 results, the median number of investments is 1 for men and 2 for women, and the investment rates for women and men are 35 and 27 percent respectively (two-sided t-test $p=0.047$).⁴⁵ A probit regression of the probability of investment on female and round that clusters standard errors at the participant level shows a positive and significant coefficient on female (marginal effect of 0.081, $p=0.045$) and a negative but insignificant coefficient on round (marginal effect of -0.010, $p=0.319$). In Appendix Table B9 we present specifications that include the improved measures of risk and altruism preferences. Closely mirroring the incentives at stake we see that being more risk averse and more altruistic helps predict investment decisions. However, consistent with our earlier results these improved controls for preferences do not help explain the gender gap in the probability of investing.

⁴⁴ Appendix Table B8 shows the distribution of investment times in Task 1 of Experiment 4.

⁴⁵ The corresponding numbers for the first-five rounds of Experiment 1 were a group investment rate of 88.4 percent; 57.9 percent of investments occur in the last two seconds; and investment rates of 35 percent for women and 25 percent for men. Similar to Experiment 1 the share of females in a session ranged from 0.38 to 0.57. The share of females in each session was: 38.1 percent in 1 session, 42.9 percent in 1 session, 47.6 percent in 3 sessions, 52.4 percent in 1 session, and 57.1 percent in 1 session.

Having completed Task 1 participants proceeded to Task 2 where over ten rounds they predict how individuals behaved in a previous session of Experiment 1.⁴⁶ In each round, they were shown individual characteristics of three individuals who formed a group in the corresponding round of the Experiment 1 session. For each group, participants were informed of each of the three group member's age, sex, whether the participant was born in U.S., his or her year in school (Freshman, Sophomore, Junior, Senior), and choice of major (Business, Social Science, or Other major). Participants indicated for each group of three players the probability that four possible events occurred: group member 1 invested; group member 2 invested; group member 3 invested; and no one invested. The sum of probabilities had to sum to one.

For robustness we used two different procedures to elicit beliefs over the four events. The first procedure asked for a probabilistic assessment of how likely the individual thought each event was and incentivized the elicitation using a binarized quadratic scoring rule (BQSR). The second procedure asked participants to first rank the events from most to least likely to have occurred and then assess how likely they thought each event was to have occurred, only the rank was incentivized.⁴⁷ For each of 10 rounds participants were shown a new group of three players for which they were asked to guess behavior. Between rounds, participants were only informed whether an investment was made in the group, which corresponded to the feedback given between rounds in Experiment 1. The specific outcome of who invested in each round was only learned at the end of the ten rounds, at which point we calculated the participant's earnings as a result of their guess and the event that actually occurred.

While the quadratic scoring rule (Brier, 1950) is an incentive compatible elicitation for risk neutral individuals, risk averse individuals have an incentive to report less dispersed beliefs. To address this concern Hossain and Okui (2013), build on the insights of Smith (1961) and Roth and Malouf (1979), and propose instead a binarized-scoring rule. Rather than having the quadratic score secure

⁴⁶ As a representative session we asked participants to predict the behavior of 21 participants in Session 8 of Experiment 1. Participant characteristics and behavior in Session 8 mirrored those of Experiment 1 overall. 48 percent of participants in that session were female, the average age was 18.29 years old, 76 percent were Caucasian, 100 percent were either freshman or sophomores, and 90 percent were born in the US. Across Session 8 the group investment rate was 79 percent and the investment rates for women and men were 32 and 21 percent respectively.

⁴⁷ We conducted four sessions (n=84) using the BQSR elicitation and three sessions (n=63) where unincentivized beliefs were elicited after participants provided an incentivized measure of belief rank.

a payoff, the quadratic score generates a chance of winning a high rather than a low fixed prize.⁴⁸ Depending on the accuracy of their guesses, participants in Experiment 4 received either \$1 or \$2 for the round. Specifically, for the event that occurred in Experiment 1 the quadratic score was used to calculate the individual's chance to win (CTW) \$2 rather than \$1. The chance-to-win function was:

$$(I) \quad CTW = 100 - 50 \sum_{i=1}^4 (1_i - p_i)^2$$

where 1_i is an indicator function that acquires the value of 1 when event i occurs and p_i is the probability a participant places on event i . Participants were given examples of how this scoring rule worked and were provided a calculator button on the decision screen that allowed them to compute the chance-to-win for any probability distribution they entered before finalizing a decision. They were also told that they had the highest chance of winning \$2 if they honestly reported their best guess. At the end of the ten rounds we calculated the chance-to-win for the round given the event that actually occurred and the participant's belief submitted for that event. By randomly drawing a number between 1 and 100 for each round we then determined whether the participant won \$2. The individual won \$2 if the drawn number was less than or equal to the individual's chance-to-win for the event that occurred. We did this for each of the ten rounds.

The elicited beliefs under BQSR correspond to the observed group investment rate and its sensitivity to the number of women in the group. While informed that participants in the original experiment did not know who they were grouped with, individuals nonetheless predict that the groups' investment rate increases monotonically with the number of women in the group. The elicited investment rate starts at 0.87 in all-male groups and increases to 0.94 in all-female groups.⁴⁹ The sensitivity to the number of women in a group reflects that participants expect women to invest more than men. In mixed-sex groups participants tended to attach the greatest

⁴⁸ In a set of experimental studies Hossain and Okui (2013) show that the binarized scoring rule secures reported beliefs that are closer to the true probability than those seen under the standard quadratic scoring rule. A similar result is provided by Harrison, Martínez-Correa, and Swarthout (2013, 2014), who show that the binarized quadratic scoring rule induces a shift towards risk neutrality in objective and subjective probabilities. Harrison, Martínez-Correa, Swarthout, and Ulm (2015), however, show no difference in reported probabilities when eliciting subjective probability distributions over continuous events that are discretized into 10 possible outcomes.

⁴⁹ This response to gender composition is significant. An OLS regression of the guessed probability of group investment rate on the number of females in a group that clusters standard errors at the subject guessing level provides a coefficient on the number of females in a group of 0.021 (std. error 0.006, $p=0.001$).

chance of investing to a female participant. Eliminating observations where a man and a woman are tied for being most likely to invest, we find in groups with one female and two males (1F2M) that women were thought most likely to invest 47.1 percent of the time, in groups with two females and one male (2F1M) the corresponding number was 75.3 percent. If men and women were thought to be equally likely to invest, those numbers would be 33.3 and 66.6 percent respectively. The observed beliefs were significantly higher in both cases (subject-level two-sided t-test $p=0.008$ for 1F2M, and $p=0.004$ for 2F1M).⁵⁰

The elicited point estimates of the third-party beliefs reveal that women are thought to be more likely to invest than men. The expected probability of investment is 0.287 (std. error 0.004) for men and 0.308 (std. error 0.004) for women.⁵¹ As seen in Table 6 a regression of the probability of investment that clusters standard errors at the subject guessing level reveals that women are thought to be 2.3 percentage points more likely to invest than a man with the same demographic characteristics other than gender. Columns 3 and 4 report the corresponding probabilities from Experiment 1 and reveal that while the average investment rate mirrors that observed, under BQSR the anticipated gender gap in investment rates is smaller than the 11.2 percentage point differential observed in Experiment 1.

⁵⁰ Similar results are obtained when using a multinomial probit model. See Appendix Table B10.

⁵¹ Removing cases where participants report three-way ties (13 percent) for the three group members, the reported chance that a man and woman invests changes to 0.292 (std. error 0.004) and 0.314 (std. error 0.004) respectively.

TABLE 6—LINEAR PROBABILITY OF INVESTING ON GROUP MEMBER CHARACTERISTICS (OLS):
ALL GROUPS⁵²

	BQSR-belief		Experiment 1-behavior		Rank-belief	
	(1)	(2)	(3)	(4)	(5)	(6)
Group member female	0.022 (0.018)	0.023 (0.025)	0.111 (0.004)	0.112 (0.005)	0.015 (0.034)	0.019 (0.015)
Group member age 19+		-0.027 (0.071)		-0.030 (0.591)		0.009 (0.627)
Group member US born		-0.039 (0.022)		-0.021 (0.678)		-0.017 (0.337)
Group member sophomore+		0.015 (0.260)		0.034 (0.553)		-0.016 (0.334)
Group member business major		-0.002 (0.832)		-0.059 (0.298)		-0.000 (0.976)
Group member other major		0.002 (0.821)		-0.043 (0.383)		0.009 (0.231)
Round	-0.000 (0.738)	-0.000 (0.740)	-0.006 (0.054)	-0.006 (0.054)	-0.000 (0.774)	-0.000 (0.776)
Constant	0.288 (0.000)	0.324 (0.000)	0.265 (0.000)	0.322 (0.000)	0.274 (0.000)	0.285 (0.000)
N	2520	2520	1500	1500	1890	1890

Notes: Dependent variable: Individual investment decision (1=invest, 0=don't invest). Standard errors are clustered at the subject level. P-values are reported in parentheses.

While third parties anticipate the direction of the gender gap in investment rates, they fail to predict the magnitude of the difference. There are several potential explanations for the quantitative difference. First it is not easy to calculate an individual's observed investment rate. Imagine for example that a participant expects that each member invests 33 percent of the time. As only one participant can invest at a time this belief corresponds to an observed individual investment rate of 23.5 percent.⁵³ Second while the BQSR elicitation is incentive compatible, the chance-to-win function is very flat at high beliefs. Thus individuals may prefer to report probabilities that are less dispersed than their true belief, because such a report substantially increases the chance-to-win on events that are seen as less likely.⁵⁴

⁵² The gender gap in beliefs presented in Table 6 is robust to including round fixed effects in the regressions.

⁵³ Since a participant cannot invest if another participant has already chosen to invest, the observed individual investment rate is $(1-(1-0.33)^3)/3$.

⁵⁴ For example, submitting a belief $(p_1, p_2, p_3, p_4) = (1.0, 0, 0, 0)$ would create a chance to win of $(1.0, 0, 0, 0)$ in the case that each event occurred. However, a less extreme belief of $(p_1, p_2, p_3, p_4) = (.60, .20, .20, 0)$ would secure corresponding chances to win of $(.88, .48, .48, .28)$. This less extreme belief only slightly reduces the chance to win in the event that group member 1 invests, while increasing substantially the chance to win in the event that one of the three other events occurred. The flatness of the incentives under BQSR implies that an individual, who correctly believes that there is a 12 percentage point gender gap in the investment frequency, in expectation would lose 0.3 cents by falsely reporting probabilities consistent with the elicited 2 percentage point gender gap (assuming equal investment rates within gender and a 10 percent chance that no one invests).

Concerned that the incentives presented under BQSR caused participants to report probabilities that were less dispersed than their actual beliefs we also ran three sessions using a dual elicitation procedure. Participants ranked the four possible events from most to least likely and then reported how likely they thought each event was to have occurred. Only the elicitation on rank was incentivized.⁵⁵ The probabilistic assessments under this dual procedure also secured a statistically significant gender gap in predicted investment rates. The investment rate is on average believed to be 0.273 (std. error 0.003) for men and 0.288 (std. error 0.004) for women.⁵⁶ As seen in Table 6 columns 5 and 6 a regression of the elicited probability of investing that clusters standard errors at the subject guessing level reveals that women are thought to be 1.9 percentage points more likely to invest than a man with the same demographic characteristics. A gender difference is also found when examining the elicited ranks. Looking at strict rankings and thus removing observations where there are ties in the probabilities we find that the mean rank assigned to men is 2.241 (std. error 0.048) and that to women is 2.116 (std. error. 0.049).⁵⁷ These elicitations are further evidence that participants believe women to be more likely to invest than men.

Across two different elicitation procedures we find that participants believe that women are more likely than men to invest. While the elicited beliefs show a gap in investment rates that is smaller than that observed in Experiment 1, such differences in expectations may nonetheless perturb a greater than expected response in behavior. Furthermore, the elicited beliefs may not fully capture the beliefs participants held in the experiment as elicited beliefs frequently differ from the beliefs consistent with observed behavior. For example Costa-Gomes and Weizsäcker (2008) find inconsistencies between the actions individuals take and their belief statements, and argue that “we need to be cautious when evaluating elicited beliefs to understand action choices” (p. 731).

⁵⁵ Participants were paid \$2 if the reported rank 1 event occurred, \$1.5 if the rank 2 event occurred, \$1 if the rank 3 event occurred, and 50 cents if the rank 4 event occurred. Participants were told that earnings were maximized by choosing a ranking that reflected the events they thought most likely occurred. The unincentivized probabilities elicited under this procedure do not show greater variation in the reported beliefs. The frequency of three-way ties in the probability that group members invest increases from 13 percent under BQSR to 30 percent under Rank.

⁵⁶ This difference is significant in a two-sided t-test by subject ($p=0.025$). Removing the 30 percent of cases where participants report three-way ties between the three group members investing, the reported chances that a man and woman invests changes to 0.276 (std. error 0.004) and 0.297 (std. error 0.006).

⁵⁷ Two-sided subject-level t-test $p=0.046$. Restricting attention to mixed-gender groups $p=0.043$ (mean men=2.234, std. error=0.054, mean women=2.100, std. error=0.053). Our focus on strict rankings considerably reduces our sample size, but restricts attention to cases where the rankings have meaning. Using the elicited beliefs to construct comparable ranks from the BQSR reported beliefs, we find a mean rank for men of 2.240 (std. error 0.044) and for women of 2.041 (std. error 0.043). Restricting attention to mixed gender groups, the mean rank for men is 2.261 (std. error 0.052) and for women is 2.036 (std. error 0.046).

Similarly, in discussing the role of subjective survey data, Bertrand and Mullianathan (2001) argue that people may not be good at forecasting their behavior or understanding why they did what they did. While the magnitude of the differences in the probability that a man and a woman invests is smaller than that observed in Experiment 1, the finding that men and women are believed to have different investment rates may be sufficient to secure that it influences equilibrium selection and in turn cause women to more frequently invest than men.

B. *Altruism (Experiment 5)*

Much economic research has been done to investigate whether men and women differ in their altruistic preferences. Although the empirical evidence is somewhat mixed (see e.g., Andreoni and Vesterlund, 2001), women are commonly thought to be more altruistic than men. Such a conviction is consistent with the behavior seen in Experiments 1 and 3, and with the elicited beliefs in Experiment 4. However, it is not consistent with the behavior seen in Experiment 2. If women are more altruistic than men then the aggregate investment rate should be higher in the all-female sessions than in the all-male sessions of Experiment 2.⁵⁸ The differential individual response in the mixed- versus single-sex session may, however, result if altruism varies by the beneficiary's gender. If participants are more generous toward men than women, then women would give less in all-female groups compared to mixed-sex group, and men would give more in all-male groups compared to mixed-sex groups. While such male-recipient favoritism can explain our results, existing empirical evidence is not consistent with such a preference. Boschini, Muren and Persson (2012) conduct dictator games in both mixed- and single-sex groups and find that the transfers by both men and women are insensitive to the sex of the recipient.⁵⁹ As our setting differs from that

⁵⁸ The aggregate investment rate in Experiment 2 was 80 percent in all-male sessions and 81.4 percent in all-female sessions (two-sided Fisher's exact test $p=0.796$). As an additional assessment of altruism we also examine the improved altruism measure included at the end of Experiment 3. While eliciting altruism over the domain of interest, this measure did not provide evidence that women are more altruistically inclined than men (mean altruism men=3.432, std. error 0.231; mean altruism women=3.209, std. error 0.168; two-sided t-test $p=0.429$). A description of this incentivized altruism measure is provided in Appendix C.

⁵⁹ Evidence that participants are not more altruistic toward men than women can also be seen in the investment decisions of Experiment 3. A probit model of the investment decision shows that the gender of the requestor has no effect on the likelihood that individuals invest using the same specification as column 1 in Table 6 (the marginal effect of the coefficient on female requestor is -0.018 $p=0.660$). The insensitivity to the requestor's gender is also seen in single-sex groups of Experiment 3. In all-male groups a probit of investing on a dummy for female requestor that clusters standard errors at the subject level secures a marginal effect of 0.100 ($p=0.697$) for those asked to invest and of -0.050 ($p=0.764$) for those not asked to invest. The results are similar for the all-female groups where the marginal effect of the dummy for female requestor is -0.079 ($p=0.633$) for those asked to invest and of -0.032 ($p=0.699$) for those not asked to invest. Further evidence that participants are not more altruistic toward men than women is seen from the gender of the person who is asked to invest not influencing the decision to invest early by other group members. That is, when a man rather than a woman is asked to invest the remaining participants are no more likely invest within the first 10 seconds of a round (a

of the standard dictator game, we conduct an additional experiment (Experiment 5) to determine whether in a setting similar to that of our investment game we find that generosity varies with the gender of the decision maker and of the other members of the group.

To assess the impact of beneficiary’s gender on altruistic transfers, 42 undergraduates participated in Experiment 5. At the beginning of the experiment, each participant provided their demographic information. Mirroring the procedures employed in Experiment 4, participants were in each of ten rounds paired in groups of three and were provided demographic information (age, sex, born in the US, year in school and major) of the other group members. They knew that they would never be paired with the same participant twice in a row. For each of ten rounds participants were asked to make six decisions that result in payoffs to themselves and to the other two group members. Specifically, they were asked whether they preferred the payoffs associated with an option A or B, with the set of possible payoffs shown in Table 7.

TABLE 7—PAYOFFS ASSOCIATED WITH EACH DECISION, EXPERIMENT 5

Decision	Option A Payoffs			Option B Payoffs		
	You	Group member 1	Group member 2	You	Group member 1	Group member 2
1	\$1.25	\$2.00	\$2.00	\$1.00	\$1.00	\$1.00
2	\$1.25	\$2.00	\$2.00	\$1.20	\$1.05	\$1.20
3	\$1.25	\$2.00	\$2.00	\$1.40	\$1.10	\$1.40
4	\$1.25	\$2.00	\$2.00	\$1.60	\$1.15	\$1.60
5	\$1.25	\$2.00	\$2.00	\$1.80	\$1.20	\$1.80
6	\$1.25	\$2.00	\$2.00	\$2.00	\$1.25	\$2.00

Option A corresponds to the payoffs that result from the participant making the generous investment decision in the original investment game (Experiments 1 and 2): \$1.25 for themselves and \$2.00 for each of the two other group members. Option B by contrast corresponds to the payoffs that result from the participant not investing and instead taking the gamble that Group member 1 invests with a certain probability. Moving from decision 1 through 6 the probability of Group member 1 investing increases in 20 percentage point increments from 0 for decision 1 to 100 percent for decision 6. A more generous participant will delay the point at which he or she

probit of investing that clusters standard errors at the individual level provides a marginal effect on female of -0.008, p=0.578) or within the first 20 seconds of a round (the marginal effect on female in the same model is -0.001, p=0.943).

switches from choosing option A to B.⁶⁰ We measure the “switch point” for preferring payoffs for other group members versus oneself, where later switching points reflect greater altruism toward other group members, especially toward Group member 1. Participants knew that for each round and for each group, one group-member’s decision would be randomly chosen and implemented.

The key results of interest are shown in Table 8. Focusing on the last point at which men and women switch from Option A to B we see that the mean switch point is approximately Decision 4. That is, the average participant is willing to take the gamble of another participant investing when the chance that Group member 1 invests is perceived to be at least 60 percent. Looking at the differences in switch points we find, consistent with Experiment 2, that there is no evidence that women are more likely to choose a later and more altruistic switch point than men. Whether looking at the decision associated with the first or last switch, men chose to switch at a later decision and are thus making decisions that are more generous toward the other group members (two-sided subject-level t-test $p=0.016$ for the first switch point, $p=0.472$ for the last switch point).

Next we ask if there is evidence of greater generosity toward male group members than female group members. We see first that decisions are no more generous when an individual is faced with two male group members rather than with two female group members (two-sided subject-level paired t-tests $p=0.822$ for the first switch point, $p=0.895$ for the last). Similarly, we do not find that participants are more generous when the individual who is more influenced by the decision maker’s selfish choice is male. That is, when Group member 1 is male we do not find that our participants select a later switch point than when Group member 1 is female (two-sided subject-level paired t-test is $p=0.795$ for the first switch point, $p=0.632$ for the last).

In summary, the literature on other regarding behavior and the results from Experiments 3 and 5 suggest that the gender difference in investing seen in Experiments 1 and 2 is not driven by women being more altruistic than men, nor by participants being more generous toward men than women.

⁶⁰ Note that individuals who are more concerned about efficiency also will delay the switch point. Research shows that men are relatively more concerned about efficiency (Andreoni and Vesterlund, 2001), which may result in men selecting a later switch point. Greater concerns for efficiency would predict greater investments in Experiments 1-4. Our new altruism measure thus encompasses both elements (generosity and efficiency).

TABLE 8— MEAN DECISION WHERE PARTICIPANT SWITCHES FROM OPTION A TO B

	Full Sample		Subsample with one Switch Point
	First Switch Point	Last Switch Point	
<i>Decision Maker</i>			
Male	3.542 (0.100)	4.185 (0.097)	3.898 (0.100)
Female	2.638 (0.104)	3.925 (0.138)	3.099 (0.114)
<i>Group members</i>			
2 Males	2.988 (0.120)	4.137 (0.129)	3.518 (0.126)
2 Females	3.148 (0.197)	4.066 (0.215)	3.542 (0.206)
<i>Group member 1</i>			
Male	3.081 (0.098)	4.073 (0.104)	3.549 (0.104)
Female	3.388 (0.120)	4.106 (0.124)	3.742 (0.121)
N obs.	420	420	327
N participants	42	42	27

Notes: Later switch points suggest greater altruism towards other group members, especially Group member 1. Standard errors are in parentheses.

IV. Conclusion

Much research has been conducted to understand gender differences in the labor market. We add to this literature an understanding of differences in the allocation of tasks that while helping the group, place the individual performing the task at a relative disadvantage (low-promotability tasks). In a series of studies we find consistent evidence that women, more than men, perform such tasks. The difference arises both from demand and supply as women more than men are asked to perform such tasks and more frequently accept requests to perform such tasks.

On the supply side, we report on field and experimental studies. The field data show that relative to men, women are significantly more likely to respond favorably to requests to undertake a task with low promotability. While this field evidence helps motivate our research question, it does not help determine why a gender gap in willingness to engage in low-promotability tasks exists. Turning to the laboratory we conduct five experiments to determine whether gender differences in the propensity to volunteer arise when incentives are better controlled, and to determine what drives such differences.

Our experimental design mirrors the incentives that group members face when asked to find a volunteer for a task that they prefer another group member undertakes. In our first two experiments all group members face the same implicit request to volunteer. In Experiment 1, when both men and women are present in the lab, we find that women are 50 percent more likely to volunteer. While there is no evidence that this gap is driven by differences in preferences, we use an experimental manipulation to simultaneously evaluate whether it results from preferences or beliefs. Our Experiment 2 uses a single-sex rather than a mixed-sex subject pool. Interestingly, the gender gap in volunteering is eliminated when participants know that they are paired only with members of their own sex. Thus, an individual's willingness to volunteer is not fixed and responds to the gender composition of the group.

We interpret the differential response to the single-sex environment as evidence that the gender gap documented in Experiment 1 is not driven by preferences but rather by the belief that women are more likely than men to volunteer. Both men and women are less likely to volunteer when there are more women in the group. This response to gender composition is consistent with the belief that women more than men will volunteer.⁶¹

By directly examining the demand side of the problem our Experiment 3 further investigates the role of beliefs. We examine whether women more frequently are asked to volunteer than men. Modifying our design to add a requestor who can ask one other participant to volunteer, we show that women are asked to volunteer more than men, and that women are asked more by both men and women. The difference in requests is consistent with the belief that women more than men accept such requests. Confirming the role of beliefs we find that women respond more favorably to these directed requests than do men. While requests increase volunteering for both men and women, the increase for women is substantially larger.

Finally, Experiments 4 and 5 shed additional light on the extent to which beliefs and/or altruism drive the gender difference in volunteering. Replicating our initial result we find in Experiment 4 direct evidence that women are expected to volunteer more than men. Along with the findings in

⁶¹ In addition to shedding light on why men and women differ in their response to the mixed-sex sessions, the results from Experiment 1 and 2 point to the environment in which we may expect differences in willingness to volunteer to be severe. In particular, women may be more likely than men to volunteer in male-dominated environments.

Experiment 5 we argue that the gender difference in volunteering is driven not by preferences but rather by the belief that women more than men will volunteer.

In examining the allocation of tasks with low promotability, our analysis is focused on a simple laboratory setting where decisions are anonymous, group size is small, and behavior can largely be characterized as being one-shot. While such a setting mirrors the allocation of some low-promotability tasks, it is of interest in future work to extend the analysis to non-anonymous settings that allow for the possibility of reputation building. The field evidence from industry and academia suggest that the differences documented here may extend to such environments, and the evidence from Experiment 3 shows that when individuals are not anonymous women are more likely to be asked to invest and to accept such requests. Initial gender differences may become self-reinforcing in repeated interactions if they strengthen the belief that women more than men will agree to perform less-promotable tasks. Furthermore, initial gender differences in task allocations may perpetuate themselves if the resulting productivity differences that arise when women spend less time than men on more-promotable tasks may make it appear to be more efficient to allocate less-promotable tasks to the seemingly ‘less-productive’ women. Finally, future research could examine how our results extend to group sizes that are larger than those examined here. For some tasks, groups sizes could be quite small (e.g., needing a full professor in an academic department who conducts lab experiments to be on the IRB) whereas for other tasks it might be quite large (e.g., needing a representative for a faculty senate committee). It is possible that increases in group size would decrease volunteering overall (as found in prior research), though gender differences may remain, as seen in the field study on faculty senate committees.

In our research we find that, relative to men, women are more likely to volunteer, more likely to be asked to volunteer, and more likely to accept direct requests to volunteer. These results suggest that the allocation of tasks with low promotability may differ even when there are no gender differences in ability and preferences. The resulting differences in task allocations can create barriers to the advancement of women in organizations and in society as a whole.

While the gender difference in volunteering and in requests to volunteer is disturbing, it is important to note that the decision to perform such tasks is not made in error. If no-one else will

volunteer it is, in our setting, individually rational to do so even if it places the individual at a relative disadvantage. From the organization's perspective it may, however, be an error to let the acceptance of low-promotability tasks be discretionary. If performance on high-promotability tasks is more important than performance on low-promotability tasks then organizations should prefer to have the latter performed by those who are least able to perform tasks with high promotability. In improving the allocation of tasks it is promising that differential request and acceptance rates appear to be influenced by beliefs. This suggests that small interventions can help reduce the differences in allocations of less-promotable tasks. In particular, beliefs may be perturbed by awareness of these differences so that documenting differences and pointing out the resulting inequities (and potential inefficiencies) may help alter the expectation that women more than men will volunteer. Further, in modifying the mechanisms by which we assign tasks we may also be able to reduce the effect of these differences. With homogenous skills it may be preferable to encourage turn-taking or to enforce random assignment, rather than to ask for volunteers. By understanding the forces that cause gender differences in task allocation it may be possible for managers to alter beliefs and the mechanisms used to assign low-promotability tasks, and this in turn may help improve the advancement of women.

References

- Ali, Muhammad, Carol T. Kulik, and Isabel Metz. 2011. "The Gender Diversity–Performance Relationship in Services and Manufacturing Organizations." *The International Journal of Human Resource Management* 22(07): 1464-1485.
- Altonji, Joseph G., and Rebecca M. Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics Vol 3, Part C* Edited by Orly C. Ashenfelter and David Card, 3143-3259, North Holland: Elsevier.
- Andreoni, James, and Lise Vesterlund. 2001. "Which Is The Fair Gender: Gender Differences In Altruism." *Quarterly Journal of Economics* 116(1): 293-312.
- Babcock, Linda, and Sara Laschever. 2003. *Women Don't Ask: Negotiation And The Gender Divide*. Princeton, NJ: Princeton University Press.
- Bear, Julia B., and Anita Williams Woolley. 2011. "The Role Of Gender In Team Collaboration And Performance." *Interdisciplinary Science Reviews* 36(2): 146-153
- Benschop, Yvonne, and Hans Doorewaard. 1998. "Covered By Equality: The Gender Subtext Of Organizations." *Organization Studies* 19(5): 787-805.
- Bergstrom, Ted, Rod Garratt, and Greg Leo. 2015. "Let Me, or Let George? Motives of Competing Altruists." <http://escholarship.org/uc/item/48m9547q>.
- Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics* 2(3): 228-255.
- Bertrand, Marianne, and Kevin F. Hallock. 2001. "The Gender Gap in Top Corporate Jobs." *Industrial and Labor Relations Review* 55(1): 3-21.
- Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *The American Economic Review* 91(2): 67-72.
- Black, Sandra E., and Philip E. Strahan. 2001. "The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry." *American Economic Review* 91(4): 814-831.
- Bliss, Christopher, and Barry Nalebuff. 1984. "Dragon-Slaying and Ballroom Dancing: The Private Supply of a Public Good." *Journal of Public Economics* 25(1): 1-12.
- Boschini, Anne, Astri Muren, and Mats Persson. 2012. "Constructing Gender Differences in the Economics Lab." *Journal of Economic Behavior & Organization* 84(3): 741-752.
- Braiker, Harriet B. 2002. *The Disease to Please: Curing the People-Pleasing Syndrome*. New York: McGraw-Hill.

- Brier, Glenn W. 1950. "Verification Of Forecasts Expressed In Terms Of Probability." *Monthly Weather Review* 78(1): 1-3.
- Chan, Curtis K., and Michel Anteby. 2016. "Task Segregation as a Mechanism for Within-job Inequality Women and Men of the Transportation Security Administration." *Administrative Science Quarterly*. 61(2):184-216.
- Costa-Gomes, Miguel A., and Georg Weizsäcker. 2008. "Stated beliefs and play in normal-form games." *The Review of Economic Studies* 75(3): 729-762.
- Croson, Rachel T.A. 2000. "Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play." *Journal of Economic Behavior & Organization* 41(3): 299-314.
- Croson, Rachel., and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47(2): 448-474.
- De Pater, Irene, Annelies E.M. Van Vianen, and Myriam N. Bechtoldt. 2010. "Gender Differences In Job Challenge: A Matter Of Task Allocation." *Gender, Work, and Organization* 17(4): 433-453.
- Diekmann, Andreas. 1985. "Volunteer's Dilemma." *Journal of Conflict Resolution* 29(4): 605-610.
- . 1993. "Cooperation in an Asymmetric Volunteer's Dilemma Game Theory and Experimental Evidence." *International Journal of Game Theory* 22(1): 75-85.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9(3): 522-550.
- Duffy, Sean, and John Smith. 2014. "Cognitive Load in the Multi-Player Prisoner's Dilemma Game: Are There Brains in Games?" *Journal of Behavioral and Experimental Economics* 51: 47-56.
- Eagly, Alice H., Wendy Wood, and Lisa Fishbaugh. 1981. "Sex Differences in Conformity: Surveillance by the Group as a Determinant of Male Nonconformity." *Journal of Personality and Social Psychology* 40(2): 384-394.
- Eckel, Catherine C., and Philip J. Grossman. 1998. "Are Men Less Selfish Than Women: Evidence From Dictator Experiments." *The Economic Journal* 108(448): 726-735.
- . 2002. "Sex Differences and Statistical Stereotyping In Attitudes toward Financial Risk." *Evolution and human behavior* 23(4): 281-295.

- . 2008. “Differences in the Economic Decisions of Men and Women: Experimental Evidence.” In *Handbook of Experimental Economics Results Vol 1*, edited by Charles Plott and Vernon Smith, 509-519. New York: Elsevier.
- Franzen, Axel. 1995. “Group Size and One-Shot Collective Action.” *Rationality and Society* 7(2): 183-200.
- Fischbacher, Urs. 2007. “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments.” *Experimental Economics* 10(2): 171-178.
- Gneezy Uri, Muriel Niederle, and Aldo Rustichini. 2003. “Performance in Competitive Environments: Gender Differences.” *Quarterly Journal of Economics*. 118(3): 1049–74
- Goeree, Jacob K., Charles A. Holt, and Angela K. Moore. 2005. “An Experimental Examination of the Volunteer’s Dilemma.” http://www.people.virginia.edu/~cah2k/vg_paper.pdf.
- Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. “The Homecoming of American College Women: The Reversal of the College Gender Gap.” *Journal of Economic Perspectives* 20(4): 133-156.
- Goldin, Claudia, and Cecilia Rouse. 2000. “Orchestrating Impartiality: The Impact of 'Blind' Auditions On Female Musicians.” *American Economic Review* 90(4): 715–41.
- Harrison, Glenn W., Jimmy Martínez-Correa, and J. Todd Swarthout. 2013. “Inducing Risk Neutral Preferences with Binary Lotteries: A Reconsideration.” *Journal of Economic Behavior & Organization* 94: 145-159.
- . 2014. “Eliciting Subjective Probabilities with Binary Lotteries.” *Journal of Economic Behavior & Organization* 101: 128-140.
- Harrison, Glenn W., Jimmy Martínez-Correa, J. Todd Swarthout, and Eric R. Ulm. 2015. “Eliciting Subjective Probability Distributions with Binary Lotteries.” *Economics Letters* 127:68-71.
- Healy, A., and J. Pate. 2009. “Cost Asymmetry and Incomplete Information in a Volunteer’s Dilemma.” Unpublished.
- Heilman, Madeline E., and Julie J. Chen 2005. “Same Behavior, Difference Consequences: Reactions to Men’s And Women’s Altruistic Citizenship Behavior.” *Journal of Applied Psychology* 90(3): 431-441.
- Holm, Håkan J. 2000. “Gender-Based Focal Points.” *Games and Economic Behavior* 32(2): 292-314.
- Hossain, Tanjim, and Ryo Okui. 2013. “The Binarized Scoring Rule.” *The Review of Economic Studies* 80(3): 984-1001.

- John, Oliver P., and Sanjay Srivastava. 1995. "The Big Five Trait Taxonomy: History, Measure, and Theoretical Perspectives." In *Handbook of personality: Theory and Research 2nd edition*, edited by Lawrence A. Pervin and Oliver P. John, 102-138. New York: Guilford.
- Kline, Patrick, and Andres Santos. 2012. "A Score Based Approach To Wild Bootstrap Inference." *Journal of Econometric Methods* 1(1): 23-41.
- Leo, Greg. 2014. "Taking Turns."
<http://www.gregcleo.com/TakingTurnsMain.pdf?attredirects=0&d=1>
- Maynard Smith, John. 1982. *Evolution and the Theory of Games*. United Kingdom: Cambridge University Press.
- Maynard Smith, J., and G. R. Price. 1973. "The Logic of Animal Conflict." *Nature*, 246: 15-18.
- Misra, Joya, Jennifer Hicke Lundquist, and Abby Templer. 2012. "Gender, Work Time, And Care Responsibilities among Faculty." *Sociological Forum* 27(2): 300-323.
- Mitchell, Sara McLaughlin, and Vivki L. Hesli. 2013. "Women Don't Ask? Women Don't Say No? Bargaining and Service in the Political Science Profession." *PS: Political Science & Politics* 46(2): 355-369.
- Murnighan, J. Keith, Jae Wook Kim, and A. Richard Metzger. 1993. "The Volunteer Dilemma." *Administrative Science Quarterly* 38(4): 515-538.
- Niederle, Muriel. 2016. "Gender." In *The Handbook of Experimental Economics Vol. 2*, edited by John H. Kagel and Alvin E. Roth, 481-562. Princeton: Princeton University Press.
- Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics* 122(3): 1067-1101.
- Ohlott, Patricia J., Marian N. Ruderman, and Cynthia D. McCauley. 1994. "Gender Differences in Managers' Developmental Job Experiences." *Academy of Management Journal* 37(1): 46-67.
- Oprea, Ryan, Keith Henwood, and Daniel Friedman. 2011. "Separating the Hawks from the Doves: Evidence from Continuous Time Laboratory Games." *Journal of Economic Theory* 146(6): 2206-2225.
- Organ, Dennis W. 1988. *Organizational Citizenship Behavior: The Good Soldier Syndrome*. Lexington, MA: Lexington Books.

- Polachek, Solomon William. 1981. "Occupational Self-Selection: A Human Capital Approach to Gender Differences in Occupational Structure." *The Review of Economics and Statistics* 63(1): 60-69.
- Porter, Stephen R. 2007. "A Closer Look at Faculty Service: What Affects Participation on Committees?" *The Journal of Higher Education* 78(5): 523-541.
- Roth, Alvin E., and Michael W. Malouf. 1979. "Game-Theoretic Models and the Role of Information in Bargaining." *Psychological Review* 86(6): 574-594.
- Sandberg, Sheryl. 2013. *Lean In: Women, Work, and the Will to Lead*. New York: Alfred A. Knopf.
- Santee, Richard T., and Susan E. Jackson. 1982. "Identity Implications of Conformity: Sex Differences in Normative and Attributional Judgments." *Social Psychology Quarterly* 45(2): 121-125.
- Smith, Cedric A.B. 1961. "Consistency in Statistical Inference and Decision." *Journal of the Royal Statistical Society. Series B (Methodological)* 23(1): 1-37.
- Weber, Elke U., Ann-Renée Blais, and Nancy E. Betz. 2002. "A Domain-Specific Risk-Attitude Scale: Measuring Risk Perceptions and Risk Behaviors." *Journal of Behavioral Decision Making* 15(4): 263-290.
- Wilhelm, Mark Ottoni, and René Bekkers. 2010. "Helping Behavior, Dispositional Empathic Concern, and the Principle of Care." *Social Psychology Quarterly* 73(1): 11-32.
- Woolley, Anita and Malone, Thomas. 2011. "What Makes a Team Smarter? More Women" *Harvard Business Review* 89(6): 32-33.