Conveniently Dependent or Naively Overconfident? An Experimental Study on the Reaction to External Help.

Zhang, Yinjunjie and Xu, Zhicheng and Palma, Marco

Tax and Transfer Policy Institute, Crawford School of Public Policy, The Australian National University, Canberra, ACT 2601, Australia, School of Economics, Henan University, Kaifeng, Henan 475001, China, Department of Agricultural Economics,Texas AM University, 2124 TAMU, TX77845

2018

Conveniently Dependent or Naively Overconfident? An Experimental Study on the Reaction to External Help*

Yinjunjie Jacquelyn Zhang† Zhicheng Phil Xu‡ Marco A. Palma§

Abstract

The rapid development and diffusion of new technologies such as automation and artificial intelligence makes life more convenient. At the same time, people may develop overdependence on technology to simplify everyday tasks or to reduce the level of effort required to accomplish them. We conduct a two-phase real-effort laboratory experiment to assess how external assistance affects subsequent revealed preferences for the convenience of a lower level of effort *versus* monetary rewards requiring greater effort. The results suggest that men treated with external help in the first phase tend to choose more difficult options with potentially higher monetary rewards. In contrast, after being treated with external help, women exhibit a stronger propensity to utilize the convenience of an easier task and are less likely to choose a more difficult option that carries higher potential earnings.

JEL Codes: C91, D81, J16

Keywords: gender difference, reaction to help, real effort

*We received useful comments from Angela de Oliveira, Rong Rong, John Spraggon, Lingbo Huang, Jie Zheng, and participants in the 2017 Asia-Pacific ESA Conference, the 4th annual Texas Experimental Association Symposium (TExAS), 2016 Beijing Normal University Conference on Experimental Economics, and seminar participants at UMASS Amherst.

†Department of Agricultural Economics, Texas A&M University, 2124 TAMU, College Station, TX 77843-2124, USA. Email: zyjj1982@tamu.edu

‡School of Economics, Henan University, Kaifeng, Henan 475001, China. Email: zhicheng.xu@vip.henu.edu.cn

§Department of Agricultural Economics, Texas A&M University, 2124 TAMU, College Station, TX 77843-2124. Email: mapalma@tamu.edu

1 Introduction

Artificial intelligence and smartphones integrate multiple features to facilitate everyday life to the point where many people are becoming addicted to their use (van Deursen et al., 2015). Today, remembering a telephone number or using a map to navigate to our destination are skills that are becoming obsolete. Middle school students develop dependency for information technology and the Internet to do their homework (Lei and Zhao, 2008). External help is not limited to technology. Helicopter parents provide excessive help to their children, who might consequently develop a dependency on their parents for doing almost everything. Nevertheless, there is also evidence that external help may have positive outcomes. For example, contrary to popular belief, a meta analysis conducted by Ellington (2003) shows that the use of calculators improves mathematical operational and problem solving skills. This paper aims to understand how the increasing reliance on external help may impact society.

Although economists have been interested in studying human behavior related to “help”, the focus has been on people’s willingness to offer help (e.g., altruistic behavior). However, whether and how external assistance affects preferences and subsequent decision-making of the help recipients still remains an open question. The immediate benefits of receiving help are straightforward, but there may also be unintended consequences on subsequent behavior and performance. Motivated by the potential externalities of receiving assistance on the help recipients, we conduct a two-phase laboratory experiment to investigate how external assistance to a real-effort task in the first phase affects individual preferences for trading-off effort *versus* monetary rewards in a subsequent task. The potential effects of receiving help may impact future behavior and performance in two opposite ways. Individuals may use external help to boost their confidence and motivation to complete a task independently and even pursue more difficult tasks in the future. Meanwhile, it is also possible that the convenience from a lower level of effort —of receiving help— may erode human capital and crowd out intrinsic work ethic. Namely, people may develop a stronger dependency on the external help, reducing their willingness to learn new skills and take on more difficult challenges. Determining the outcomes of external help is important for evaluating the welfare effects of business management strategies and policy interventions designed to provide people with external assistance.

We pay particular attention to potential gender differences in reaction to external help. Gender composition is unbalanced in many fields, ranging from industry and politics to academia.[1] There is ample evidence in the economics literature of significant gender differences in risk attitudes and competition. Relevant to our study, a large body of literature originating in psychology, documents substantial gender differences in the consequences of receiving help (see Section 2 for a comprehensive discussion). If men and women also experience differential impacts from external assistance, we believe it is critical to understand whether these asymmetric effects increase or reduce the prevalent gender gap.

Our laboratory experiment consists of two stages. Participants were randomly assigned to the treatment or control group. Subjects in the control group performed a paid real-effort task without any assistance, while subjects in the treatment group performed the same task with external assistance, receiving hints for the right answers that simplified the task significantly. The second stage introduced a different real-effort task. In order to elicit the subjective relative evaluation of monetary rewards against the convenience of external assistance (i.e., less effort), before the second task began, subjects were allowed to choose a payment schedule and effort level through a multiple price list (MPL) (Holt and Laury, 2002; Andersen et al., 2006). The MPL offered subjects an array of ordered scenarios (in rows) that differed in potential earnings and the amount of external help. For each row, subjects had to choose between option A, with 16 questions (accordingly lower potential earnings) and option B, with 24 questions (higher potential earnings). External help was again provided as hints that simplified the task. The number of hints in option A decreased

[1]For example, women hold only 6.4% of Fortune 500 CEO roles (see `http://fortune.com/2017/06/07/fortune-women-ceos/`). Female students tend to sort themselves out of STEM fields (see `http://www.joannejacobs.com/2014/03/fearing-bs-women-reject-stem-majors/`)

for each row of the MPL (from 16 to 0), while option B had a fixed number of hints (8). Consequently, the row in which a subject switched from option A to B provides a measure of his/her preference for the trade-off between the convenience of using external help and the extra effort required to obtain higher earnings (see the Appendix for a more detailed illustration).

The experimental design tests whether—after being treated with external help in the first stage—subjects develop a behavioral dependency on its convenience or boost their confidence and motivation, leading them to perform the real-effort task with less external assistance in the second stage. The results show that, after being treated with help, men tend to overestimate their cognitive capability and underestimate the effort required to perform the real-effort tasks. Although there are no differences in performance by gender, men are overconfident and less likely to use external help. Women, on the other hand, exhibit a stronger propensity to utilize the convenience and choose a less challenging task in the second stage. We further explore the underlying mechanism of how cognitive bias affects individuals' behavior by looking at differences in the switching patterns of the treatment and control groups conditional on the performance level.

The rest of the paper proceeds as follows. Section 2 discusses how our study relates to previous literature. Section 3 introduces the experimental design. Section 4 reports the general results. Section 5 concludes.

2 Related Literature

Using external help as the treatment links our study to an abundant literature in psychology examining "reactions to help." A review of literature helps to understand the roots of our findings. Fisher et al. (1982) argue that the effects of help are mixed, inducing either self-threatening or self-supportive experiences for the help recipients. On one hand, receiving help may hurt self-esteem by implied inferiority, inadequacy, and dependency. On the other hand, help can also be perceived as positive and supportive, often resulting in material gains (Nadler and Jeffrey, 1986).

Reactions to help differ by gender. The "threat to self-esteem" model suggests that when men receive help from a person with similar experience, it lowers their self-confidence. However, help can also provide stronger self-confidence if the giver has more experience (Fisher and Nadler, 1974; Nadler et al., 1976, 1979). Receiving help does not seem to harm the self-esteem and performance of women (DePaulo et al., 1981; Daubman and Lehman, 1993). Women are more inclined to admit that they need assistance and appreciate the help, while men experience more self-doubt. In our experiment, we find that after being treated with help, men have a stronger propensity to demonstrate their confidence by choosing a more challenging option, while women tend to develop greater dependency on the convenience of lower effort.[2]

There is a small but growing literature in experimental economics that discusses gender differences in responding to external advice. For instance, Brandts et al. (2014) point out that external advice from an experienced person has different impacts on men and women's work efficiency and competition entry in a real-effort task. They mainly focus on the impacts of external advice on the decision to enter a tournament, while our experiment examines the extent to which external assistance can affect confidence and effort in a subsequent task. Heikensten and Isaksson (2018) examine how the gender of the advisors influences individuals' advice-seeking decision and whether this impact is heterogenous across genders. While they focus on the gender of influencers, our design concentrates on the influencees' willingness to receive subsequent help after a training session and the potential gender differences from them. Notably, a major difference between previous studies and ours is that the external help in our experiment is provided by a *computer* rather than another person. Hence, the

[2]It is noteworthy that in the above-mentioned research in psychology, the experimental design deliberately leads subjects to believe that their performance is a reflection of their intellectual abilities. In most cases researchers also lead participants to believe that they performed significantly worse than their peers. In order to avoid contamination from potential self-doubt and negative feelings, we did not provide subjects negative or positive feedback about their performance until the end of the experiment.

results of our experiment are more suitable for understanding the effects of non-human help.

Our experiment also mirrors a large literature on gender differences in risk preferences and competition (Eckel and Grossman, 2008; Croson and Gneezy, 2009; Niederle and Vesterlund, 2011; Reuben et al., 2015; Buser et al., 2017; van Veldhuizen, 2017). A notable finding in this literature is that men and women have remarkable differences in their propensity to engage in competitive behaviors. To be specific, women shy away from competition, while men are more competitive, even in tasks in which they are not more capable than women (Niederle and Vesterlund, 2007; Wieland and Sarin, 2012; Buser et al., 2014). In field experiments of intellectual (Gneezy et al., 2003) and physical competition (Gneezy and Rustichini, 2004), men show greater effort and better performance in a competitive environment, while women's performance remains unchanged regardless of the environment's competitive level.

More closely related to the findings in our study, previous research suggests that men seem to gain self-esteem by demonstrating that they are better than others (Schwalbe and Staples, 1991; Josephs et al., 1992; Crocker et al., 2003). In contrast, Günther et al. (2010) find that women avoid competing with men, even in areas where women wrongly believe they have lower performance. Niederle and Vesterlund (2007) show that about one third of the gender gap in tournament entry can be explained by gender differences in confidence. Therefore, our experiment is complimentary to studies on gender differences related to overconfidence and self-esteem.[3] Overconfidence may be useful to explain our experimental result showing that after receiving help, men have greater willingness to take the challenge of a more difficult task.

Our experiment differs from previous literature in that most past studies have compared gender differences in competition with other people. In contrast, our study focuses on gender differences in reaction to external help from technology, whereby our results provide useful

[3]Note that these stylized findings may not be entirely driven by innate gender-specific characteristics. Women's under-performance in competitive environments also depends on the task (Günther et al., 2010; Dreber et al., 2011; Shurchkov, 2012; Wieland and Sarin, 2012), the gender composition of the competing group (Gneezy et al., 2003; Gneezy and Rustichini, 2004; Gupta et al., 2013), stereotype and information conditions (Iriberri and Rey-Biel, 2016), and cultural and social norms (e.g., patriarchal society vs. matrilineal society) (Gneezy et al., 2009).

insights into external help change in labor markets driven by technological development.

3 Experimental Design

The experiment has two stages. The first stage is a real-effort task consisting of ten questions with five lines of text; each line contains a random combination of 50 letters. We used the 26 lowercase letters of the English alphabet to construct the question. Participants were asked to count the number of times a predetermined letter appeared in the text. In the treatment, participants were provided with external help in the form of hints that significantly simplified the task. The presence of hints made all irrelevant letters less salient—although they were still present—to simplify the counting task (see Figure 1).

Insert Figure 1 here.

Participants first viewed sample questions, with or without hints, and then began to work on the ten questions. A timer displayed in the right corner of the screen counted the time used for each question. The timer gave participants a sense of the level of effort required to complete the task with and without help. Each participant had an equal chance of being randomly assigned to the *treatment* group, where they would work with the hints, or to the *control* group, where they would work without hints. In order to complete the first-stage task, participants had to correctly answer all ten questions. Although the task is not difficult, it requires effort to complete it. Obviously, higher effort is required in the control condition (without hints) compared to the treatment condition (with hints). Over the course of implementing the first-stage task, subjects were not allowed to proceed to the next question until they provide the correct answer for each question. Participants were allowed multiple attempts to enter the right answer to each question. They can only proceed to the next question when the correct answer was filled in. As such, at the end of the first stage, all participants earned $10 for completing the task, in which we avoid the potential income effects.

The main purpose of the first stage was to randomly treat half of the subjects with external help (i.e., the hints for the right answers that significantly simplify the task). We hypothesize that this hint treatment would influence an individual's preference for trading off payment for receiving in the subsequent stage.

In the second stage, each subject was randomly assigned with equal probability to either another real-effort task or a Raven's test. Again, participants were provided with sample questions before performing the task. The real-effort questions in the second stage were very similar to the questions in the first stage, with the only difference that the second stage used numbers instead of letters. An example of a Raven's test question is shown in Figure 2. Analogous to the real-effort case, the hints suppressed some irrelevant answers helping individuals by reducing the answer pool.[4]

Insert Figure 2 here.

In order to elicit preferences for external help and monetary rewards, we presented participants with a multiple price list (MPL) before starting the second stage. As shown in Figure 3, participants were asked to make a choice between option A and B in each of 17 scenarios. Option A had 16 questions and option B had 24 questions; hence option B always has higher potential earnings. The number of hints in option A is descending in the list from 16 in the first row to 0 in the last row, while the number of hints in option B is fixed at 8. Note that the attractiveness of option A decreases by each decision row. This can be easily illustrated by comparing the first two rows in Figure 3. In the first row, subjects face a trade-off between option A, which would pay \$16 with a very high probability, since there are 16 questions and 16 hints, while option B has higher possible payoff because of more questions (24) but also requires greater effort because of fewer hints (8) under the same time limit. In the second row, option A becomes less attractive compared to the first row because

[4] We use two types of task in the second stage. The goal was to detect whether behavioral patterns induced in the first stage would be significantly adjusted due to the similarity of the task in the second stage. Mann-Whitney U tests of the key indicators show there are no significant differences between the two types of tasks.

for the same number of questions (16) there are fewer hints (15). Accordingly, the subjects' willingness to select option A diminishes as the row number increases. The row number in which a subject switches from option A to B provides a measure of individual preferences for monetary rewards over external help. We argue that preferences are influenced by the hint treatment introduced in the first stage, which significantly simplifies the task. The earlier a subject switches from option A to B, the more evident that the subject is willing to forgo external help and choose a more difficult task with higher potential earnings. (See the Appendix section for a more sophisticated analysis.)

Insert Figure 3 here.

Subjects were informed that a lottery for the second stage would randomly determine one of the choice pairs to be realized. They had 20 minutes to complete the task, and each correct answer was worth $1. In order to collect earnings, subjects can not make errors in more than 25% of the questions.

Otherwise, no payment would be delivered during this stage. The payment criterion was used to discourage subjects from always choosing option B based on strategic behavior. Since option B always has a higher potential payoff than option A, subjects would have a higher chance of earning more money by choosing option B if there is no restriction on the accuracy rate. In other words, to increase the salience of external assistance, we increased the difficulty of option B by enforcing this rule. This restriction is analogous to real labor markets, where worker's performance is evaluated not only on the quantity, but also on the quality of their work. At the end of the experiment subjects filled a questionnaire, with demographic questions, including gender, race, religion, and ideology as well as self-evaluations regarding their performance in the experiment.

The experiment was computerized using the software 'z-Tree' (Fischbacher, 2007) and conducted at Texas A&M University. We used a between-subject design and each subject participated in only one session. The duration of each session was approximately 60 minutes, including sign-up, consent, decision making, and payment. Before entering the laboratory, participants were informed that they would receive a show-up fee of $5 upon completing the tasks and would also have the opportunity to earn extra payoffs based on their decisions and performance. However, they were not provided with any details about the experiment.

4 Results

We begin our analysis with descriptive statistics of the experimental results. Then, we explore whether external assistance has a significant effect on the decision makers' revealed preferences for monetary rewards and external help. Further, we test whether men and women react differently to external assistance using a difference-in-differences (DD) framework.

We assume that rational agents have only one switching point from option A to option B during the MPL stage.[5] After excluding 17 subjects who made multiple switches, a total of 160 subjects remain in the sample.[6] A balance check of the sample is presented in Table 1. The t-tests report that there are no significant differences between the treatment and control groups over a set of demographic covariates.

Insert Table 1 here.

In Table 2, we compare the average time spent per question between the treatment and control groups during the first stage. Introducing hints substantially improved the performance of both male and female participants. On average, participants in the treatment group spent significantly less time per question than participants in the control group. The difference in the time spent per question for the treatment and control group provides an objective measurement of the convenience provided by the external help. The experience in the first stage gives participants a reference point to make their switching choices in the MPL stage.

[5] The uniqueness of the switching point is proven by the theoretical framework provided in Appendix A.

[6] Given the sample size for a 2×2 design, we are able to detect effect size of as small as 0.26 with 80% of power.

Insert Table 2 here.

By design, there are two types of tasks in the second stage: another real-effort task similar to the first stage and a Raven's test. The purpose of using two different tasks is to detect whether behavioral patterns induced in the first stage will be significantly adjusted based on the similarity of the tasks in the two stages. Mann-Whitney U tests (Table 3) of the key indicators—including switching patterns, self-reported beliefs on second-stage performance, accuracy rate per question, and time spent per question—in both stages show no significant differences between the tasks. This suggests that the induced treatment effect on switching patterns is not related to differences in the tasks.[7] Next, we pool the data of the two types of tasks in the analysis of the second-stage behavior.

Insert Table 3 here.

4.1 Heterogeneous treatment effects on revealed preference

We show the overall comparison of the switching point between the treated and control group in the first row of Table 4. Subjects who received external help in the first stage do not show significantly different switching patterns compared to the control group. On average, both groups switched from option A to B between the 8th and 9th decision row.

Further investigation of gender differences in switching choices shows that the first-stage treatment affected the switching patterns of men and women in opposite ways, offsetting the overall effect.

Insert Table 4 here.

As shown in Table 4, treated men switched earlier than men in the control group. Male subjects in the control group, on average, made the switching decision between the 9th

[7] In the estimation of treatment effects within the DD framework, shown in Section 4.1, we further control for session and task combination fixed effects to show that our treatment effect results are not contaminated by differences in the tasks in the second stage.

and 10th question, while treated male subjects switched around the 6th and 7th question. This difference is significant at the 5% level. In contrast, women tend to be reluctant to switch too early if they received hints in the first stage, although the difference is subject to large variation. Since the switching points have multiple peaks and skewed distributions, we check the robustness of our findings by building confidence intervals using the bootstrapping method (Figure 4).

Insert Figure 4 here.

In order to take a closer look at the behavioral patterns, we explore potential changes in the decision maker's revealed preference (see Appendix A). Figure 5 displays the cumulative percentage of switching points at each decision row in the control and treatment groups. In both graphs, gender differences are more pronounced at the beginning, but gradually disappear at the end of the MPL.

Insert Figure 5 here.

For the control group, up until the 12th decision row, women's cumulative percentage of switching points is always higher than men's (panel a of Figure 5). Half of females who did not receive hints in the first stage switched before the 8th decision row, while this ratio for males is less than 35%. Men close the gap with women by the 13th decision row, where nearly 80% of both genders have switched. For the rest of the decisions (14th–17th), the cumulative percentage of switching points for men is slightly higher than women. This suggest that women tend to place a higher value in the potential monetary payout, while men tend to avoid higher effort.

Interestingly, this pattern is reversed with the external help treatment (panel b in Figure 5). Compared to the control group (without hints), women's switching points were delayed, while men switched much earlier.[8] Over 35% of male participants receiving hints in the first

[8] In our theoretical model, this indicates that after being treated with hints, α decreases for women but it increases for men, see the Appendix.

stage chose to switch to option B at the first row of the MPL, compared to only 11% of females. More than 60% of male participants switched by the 9th decision row, compared to only 45% of females. Women closed the gap with men at around the 11th decision row. The distinctive change in the switching pattern by gender again indicates that men are more likely to switch later in the control, but they are more likely to switch earlier in the treatment. Meanwhile, a significant number of treated females converged to switching between the 8th and 12th decision row.

We further test this relationship in the following difference-in-differences framework:

$$\text{Switching point}_{ist} = \omega Female_i + \theta Hint_t + \boldsymbol{\delta}(Hint_t \cdot Female_i) + \gamma_s + X_{ist}\beta + \epsilon_{ist}, \qquad (1)$$

where γ_s is session fixed effects,[9] $Female_i$ is gender indicator for subject i, and X_{ist} captures individual characteristics (see Table 1). $Hint_t \cdot Female_i$ is the interaction of external help treatment and the gender indicator, which is equal to 1 for female participants assigned to the treatment group and 0 otherwise. The parameter $\boldsymbol{\delta}$ is our key difference-in-difference estimator.

The point estimates are reported in Table 5. We implement the estimations by gradually adding controls for fixed effects of session or task type, and individual characteristics such as background covariates, self-evaluations and second-stage performance. [10] In response to the treatment of external help, female participants on average switched 4–5 decision rows later than male counterparts. The point estimates are statistically significant at the 5% level in column (1) and at the 1% level when controlling for session fixed effects in columns (2) and (3), and task type fixed effects in column (4). Therefore, using two different types of tasks (Real-effort task and Raven's test) in the second stage does not affect the robustness of the estimates, which is in line with the previous analysis presented in Table 3.

[9] We also control for the task type fixed effects in some specification for robustness check.

[10] As per our design, we control for individual ability using the performance in the second stage. We argue that individual ability is unlikely to be affected from the first stage to the second stage of the experiment, particularly since both tasks are very similar.

Insert Table 5 here.

4.2 Exploring the Potential Causes for the Gender Gap in Reaction to External Assistance

Thus far, our results have shown that the first stage hint treatment drives men and women to differ substantially in their switching patterns in the subsequent MPL stage. While male participants appear to place a higher value in the monetary payout, female participants seem to value more the convenience of external help. Next, we attempt to explore the possible mechanisms through which the hint treatment causes these divergent effects.

4.2.1 Performance in the second stage

We compare the performance of female and male subjects in the second stage (Table 6). Panel A reports the proportion of correct answers in the second stage overall and by gender. Although male subjects are willing to take a more difficult option, their performance is no different than the performance of females. This indifference holds even when the sample is divided by treatment assignment or by the number of questions selected in the second stage. In panel B, we compare the time spent per question (in seconds). Again, there are no significant differences by gender.

Insert Table 6 here.

In Figure 6, we further present the cumulative distribution of the proportion of correct answers for the overall sample. At each performance level, it depicts the share of individuals who solve at most that proportion of correct answers in the second stage. The distributions in the treatment and control conditions are shown in the two panels of Figure 7, respectively. Across the three graphs, the distributions closely track each other. It is unlikely that gender differences in ability drive the heterogeneous treatment effects.

Insert Figures 6 and 7 here.

4.2.2 Cognitive bias

We then test whether differences in individuals' cognitive biases about their ability drives the results. At the end of the second stage and before being notified about their earnings, participants were asked to report their beliefs regarding their performance relative to others. Figure 8 shows the mean gender comparison of self-evaluated performance.[11] The two-sided t-test suggests that there are no significant gender differences in the control group ($p = 0.303$). However, there is a significant difference in the treatment group ($p = 0.001$). A difference-in-difference estimation showed similar results (Table 7). Clearly, the treatment significantly boosted the confidence and subjective beliefs of men, despite no significant differences in the actual performance between genders (Table 6).

Insert Figure 8 and Table 7 here.

The greater confidence shown by men provides suggestive evidence that cognitive bias could be driving the earlier switching patterns exhibited by men. In contrast, women's self-evaluation on performance did not significantly change by the treatment. Women's late switching is not driven by changes in their beliefs about their ability. While receiving the treatment induces men to become overconfident, regardless of their true ability, they significantly underestimate the required effort to complete the task. Other possible mechanisms behind the results are examined in Appendix B.

5 Conclusion and Discussion

Everyday, people appear to rely more on external help from new technologies. In this laboratory experiment, we focused on the effects of external help on the trade-off between higher potential monetary rewards requiring greater effort and the convenience of lower effort from external assistance.

[11]Subjects are asked to evaluate their performance relative to the rest of the participants in the same session. 10 = better than 100% of others, 0 = no better than any others.

In particular, we find that after receiving help men tend to overestimate their cognitive ability and underestimate the necessary effort to perform a real-effort task. Consequently, men are more likely to choose a more difficult task with higher potential earnings after being treated with external help, but they do not perform better than women.

Women, on the contrary, tend to adjust their beliefs and decisions based on external supporting information in the opposite way. We argue that although external help may induce weak-performing women to utilize the convenience of external assistance, the possibility of strong-performing women to also develop a dependency cannot be ruled out (see the Appendix B for an elaborate analysis).

To some extent, the observed treatment effect differences by gender may be useful to explain why women are more risk averse and avoid competition, while men actively engage in competitive behavior. According to our results, external assistance makes women more likely to depend on it. This, in turn, might drive women to behave more conservatively. The behavioral bias exhibited by men indicates a refusal to external help, which ultimately becomes financially costly by reducing their earnings (see the Appendix B and C).

References

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutstrm (2006). Elicitation using multiple price list formats. *Experimental Economics 9*(4), 383–405.

Brandts, J., V. Groenert, and C. Rott (2014). The impact of advice on women's and men's selection into competition. *Management Science 61*(5), 1018–1035.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *Quarterly Journal of Economics 129*(3), 1409–1447.

Buser, T., N. Peter, and S. C. Wolter (2017). Gender, competitiveness, and study choices in high school: Evidence from switzerland. *American economic review 107*(5), 125–30.

Crocker, J., R. K. Luhtanen, M. L. Cooper, and A. Bouvrette (2003). Contingencies of self-worth in college students: Theory and measurement. *Journal of Personality and Social Psychology 85*(5), 894–908.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–474.

Daubman, K. A. and T. C. Lehman (1993). The effects of receiving help: Gender differences in motivation and performance. *Sex Roles 28*(11-12), 693–707.

DePaulo, B. M., P. L. Brown, S. Ishii, and J. D. Fisher (1981). Help that works: The effects of aid on subsequent task performance. *Journal of Personality and Social Psychology 41*(3), 478.

Dreber, A., E. von Essen, and E. Ranehill (2011). Outrunning the gender gap—boys and girls compete equally. *Experimental Economics 14*(4), 567–582.

Eckel, C. C. and P. J. Grossman (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization 68*(1), 1–17.

Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education*, 433–463.

Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fisher, J. D. and A. Nadler (1974). The effect of similarity between donor and recipient on recipient's reactions to aid1. *Journal of Applied Social Psychology 4*(3), 230–243.

Fisher, J. D., A. Nadler, and S. Whitcher-Alagna (1982). Recipient reactions to aid. *Psychological Bulletin 91*(1), 27.

Gneezy, U., K. Leonard, and J. List (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica 77*(5), 1637–1664.

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

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

Günther, C., N. A. Ekinci, C. Schwieren, and M. Strobel (2010). Women can't jump?an experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization 75*(3), 395–401.

Gupta, N. D., A. Poulsen, and M. C. Villeval (2013). Gender Matching And Competitiveness: Experimental Evidence. *Economic Inquiry 51*(1), 816–835.

Heikensten, E. and S. Isaksson (2018). Simon Says: Examining gender differences in advice seeking and influence in the lab. Technical report.

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

Iriberri, N. and P. Rey-Biel (2016). Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. *Barcelona GSE Working Paper: 880*.

Josephs, R. A., H. R. Markus, and R. W. Tafarodi (1992). Gender and self-esteem. *Journal of Personality and Social Psychology 63*(3), 391–402.

Lei, J. and Y. Zhao (2008). One-to-one computing: What does it bring to schools? *Journal of Educational Computing Research 39*(2), 97–122.

Nadler, A., A. Altman, and J. D. Fisher (1979). Helping is not enough: Recipient's reactions to aid as a function of positive and negative information about the self. *Journal of Personality 47*(4), 615–628.

Nadler, A., J. D. Fisher, and S. Streufert (1976). When helping hurts: Effects of donor-recipient similarity and recipient self-esteem on reactions to aid1. *Journal of Personality 44*(3), 392–409.

Nadler, A. and D. Jeffrey (1986). The role of threat to self-esteem and perceived control in recipient reaction to help: Theory development and empirical validation. *Advances in experimental social psychology 19*, 81–122.

Niederle, M. and L. Vesterlund (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annual Review of Economics 3*(1), 601–630.

Reuben, E., P. Sapienza, and L. Zingales (2015). Taste for competition and the gender gap among young business professionals. Technical report, National Bureau of Economic Research.

Schwalbe, M. L. and C. L. Staples (1991). Gender differences in sources of self-esteem. *Social Psychology Quarterly 54*(2), 158–168.

Shurchkov, O. (2012). Under pressure: Gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association 10*(5), 1189–1213.

van Deursen, A. J., C. L. Bolle, S. M. Hegner, and P. A. Kommers (2015). Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. *Computers in Human Behavior 45*(Supplement C), 411 – 420.

van Veldhuizen, R. (2017). Gender differences in tournament choices: Risk preferences, overconfidence or competitiveness? Technical report, Discussion Paper.

Wieland, A. and R. Sarin (2012). Domain specificity of sex differences in competition. *Journal of Economic Behavior & Organization 83*(1), 151 – 157.

Figures and Tables

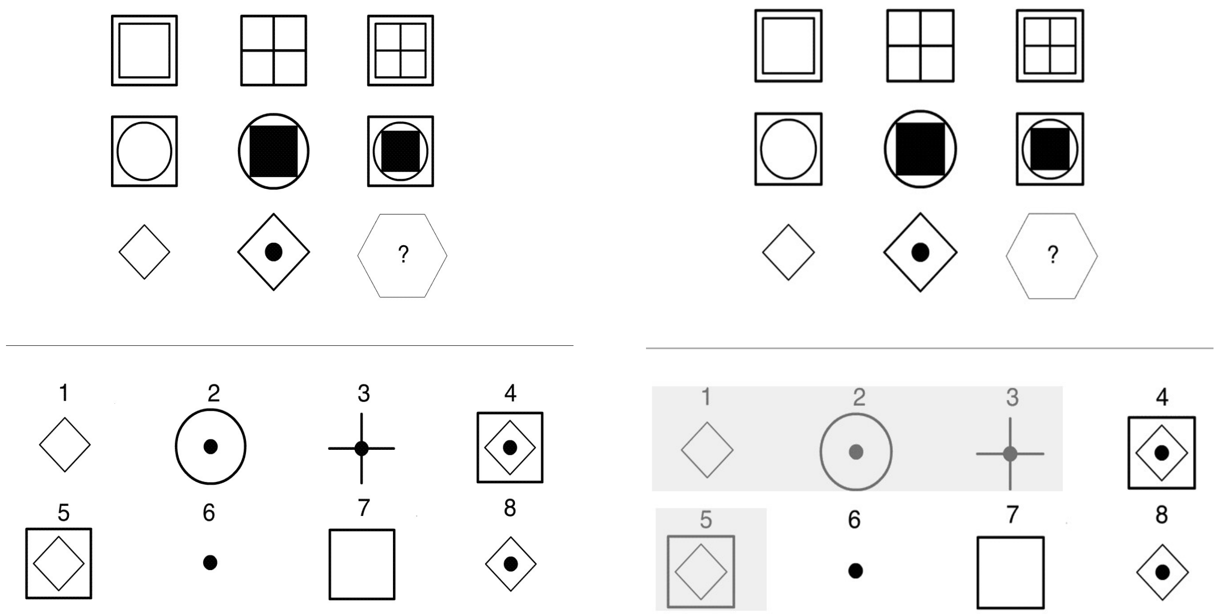
cfdkiogiicenmpyizchhomlhqcvfrqkwqwblokrtyhgiocaodp
poegvztuivirdbduzocreerhxyywwlqpxdmgrevqoaehhlbxgw
cmbuknagepuricjhevoszbamaauvvsjedspdnzvqrmjjuycrjn
tohvarjmfkvchzwgroezapuahnxdfjypxooplyzdfklhqpytuw
kntawkkcvnnsqgaiqxweihjothhhpzfvqvmfaxmzmwrcfrkctt

(a) Control Without Hints

cfdkiogiicenmpyizc**hh**oml**h**qcvfrqkwqwblokrty**h**giocaodp
poegvztuivirdbduzocreer**h**xyywwlqpxdmgrevqoae**hh**lbxgw
cmbuknagepuricj**h**evoszbamaauvvsjedspdnzvqrmjjuycrjn
to**h**varjmfkvc**h**zwgroezapua**h**nxdfjypxooplyzdfkl**h**qpytuw
kntawkkcvnnsqgaiqxwei**h**jot**hhh**pzfvqvmfaxmzmwrcfrkctt

(b) Treatment With Hints

Figure 1: An Example of the Real-Effort Task in the First Stage



(a) Control Without Hints

(b) Treatment With Hints

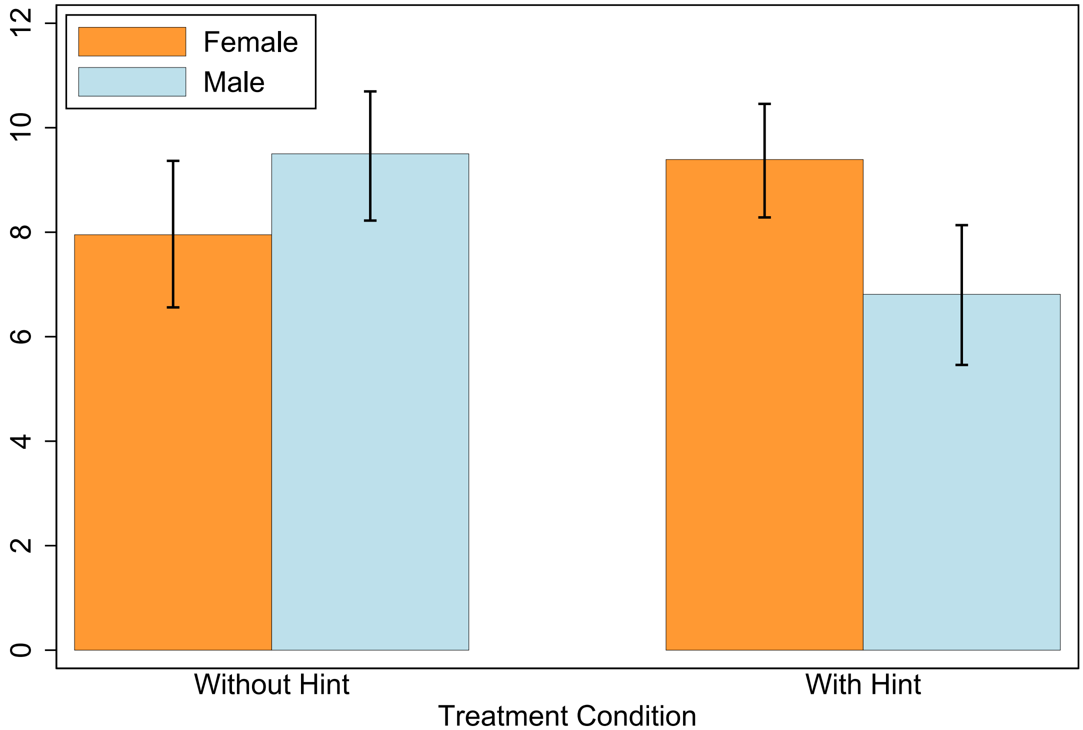
Figure 2: An Example Question from the Raven's Test Real-Effort Task

Reminder Clock [sec]: 594

Decision number	Option A	Option B	Choose A or B
1	16 Questions with 16 Hints	24 Questions with 8 Hints	A B
2	16 Questions with 15 Hints	24 Questions with 8 Hints	A B
3	16 Questions with 14 Hints	24 Questions with 8 Hints	A B
4	16 Questions with 13 Hints	24 Questions with 8 Hints	A B
5	16 Questions with 12 Hints	24 Questions with 8 Hints	A B
6	16 Questions with 11 Hints	24 Questions with 8 Hints	A B
7	16 Questions with 10 Hints	24 Questions with 8 Hints	A B
8	16 Questions with 9 Hints	24 Questions with 8 Hints	A B
9	16 Questions with 8 Hints	24 Questions with 8 Hints	A B
10	16 Questions with 7 Hints	24 Questions with 8 Hints	A B
11	16 Questions with 6 Hints	24 Questions with 8 Hints	A B
12	16 Questions with 5 Hints	24 Questions with 8 Hints	A B
13	16 Questions with 4 Hints	24 Questions with 8 Hints	A B
14	16 Questions with 3 Hints	24 Questions with 8 Hints	A B
15	16 Questions with 2 Hints	24 Questions with 8 Hints	A B
16	16 Questions with 1 Hints	24 Questions with 8 Hints	A B
17	16 Questions with 0 Hints	24 Questions with 8 Hints	A B

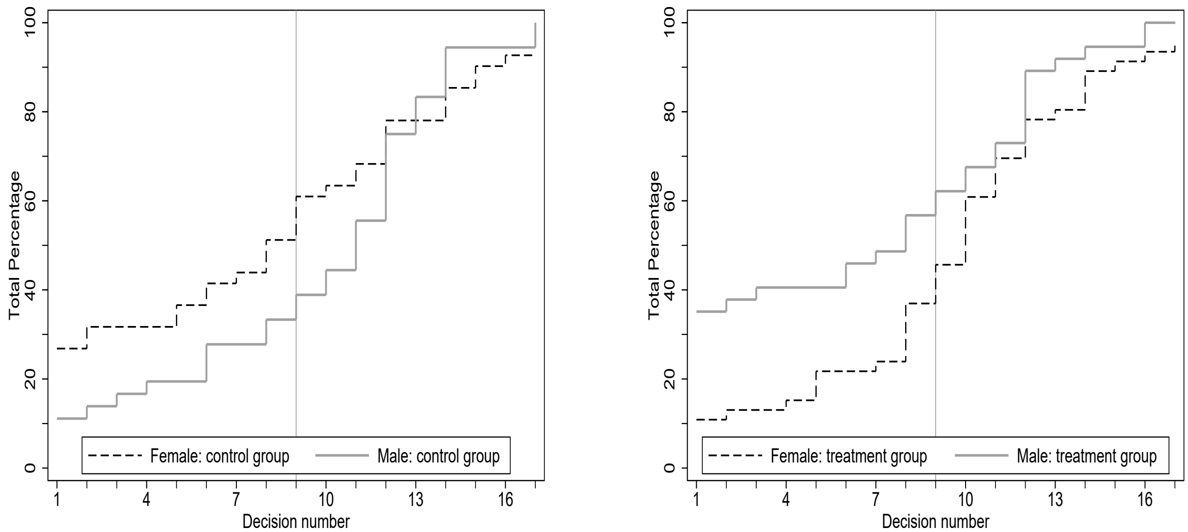
submit

Figure 3: Binary Choices in the Multiple Pricing List Offering



Note: 90% confidence intervals are based on bootstrapping with 1000 replications

Figure 4: Switching Point Gender Comparison by Bootstrap Method



(a) Control: Without Hints

(b) Treatment: With Hints

Figure 5: Cumulative Switching Point from Option A to B

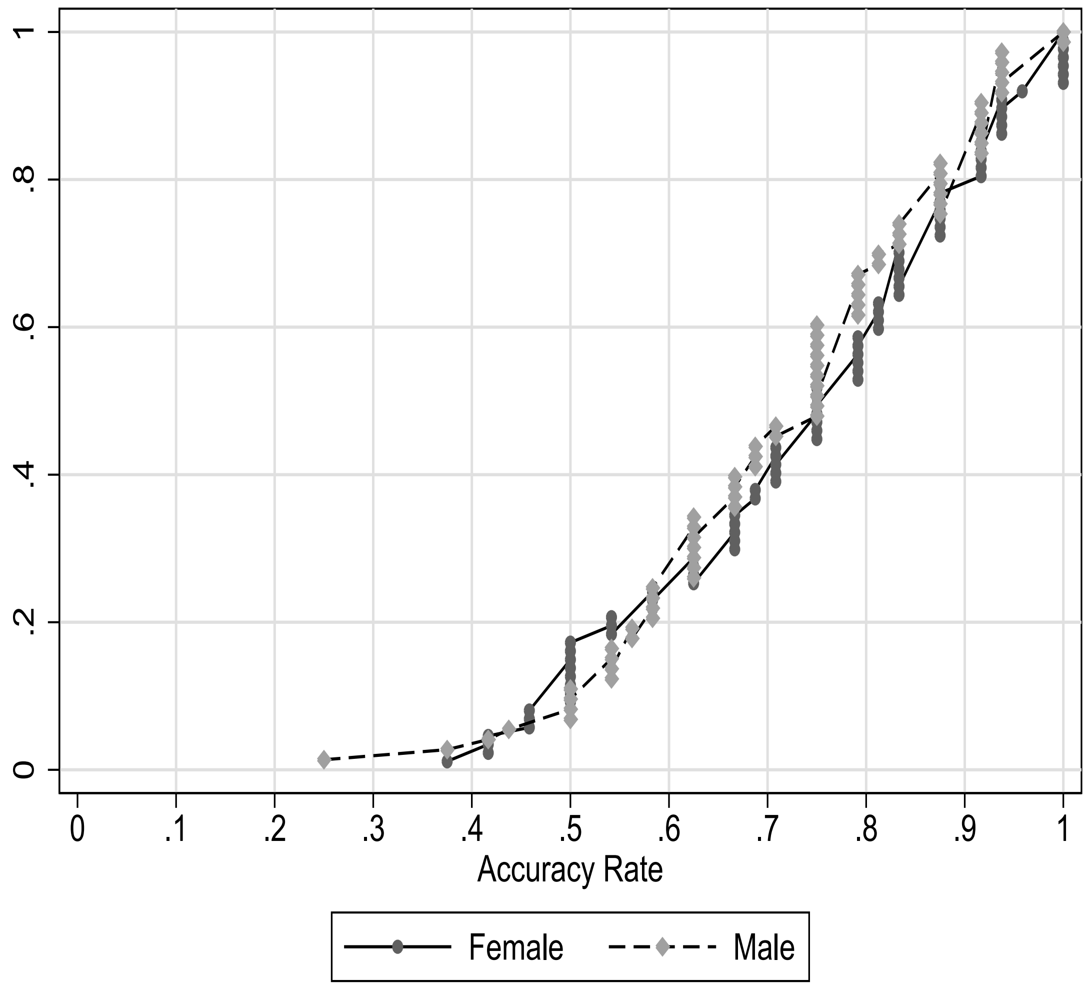
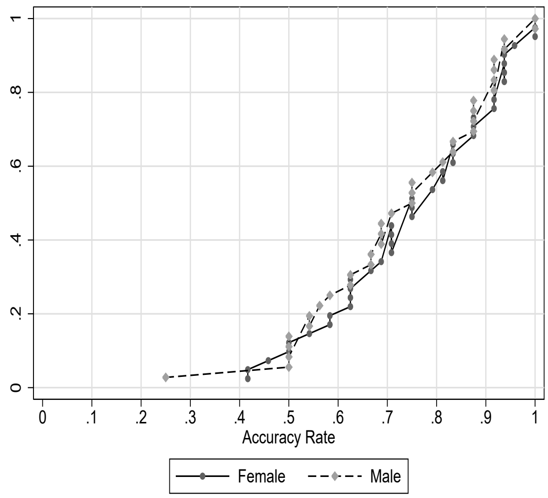
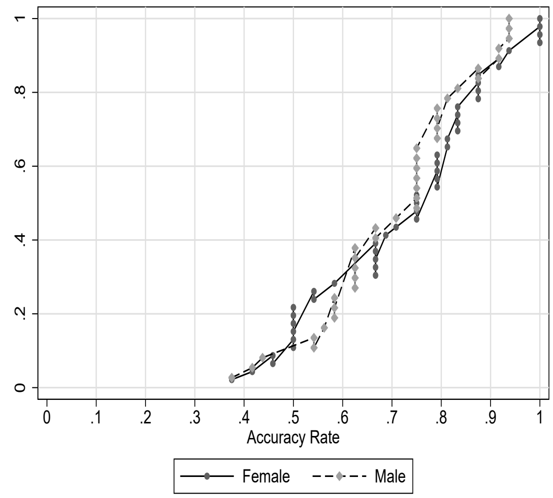


Figure 6: CDF of the Proportion of Correctly Solved Problems in the Second Stage: Pooled Sample

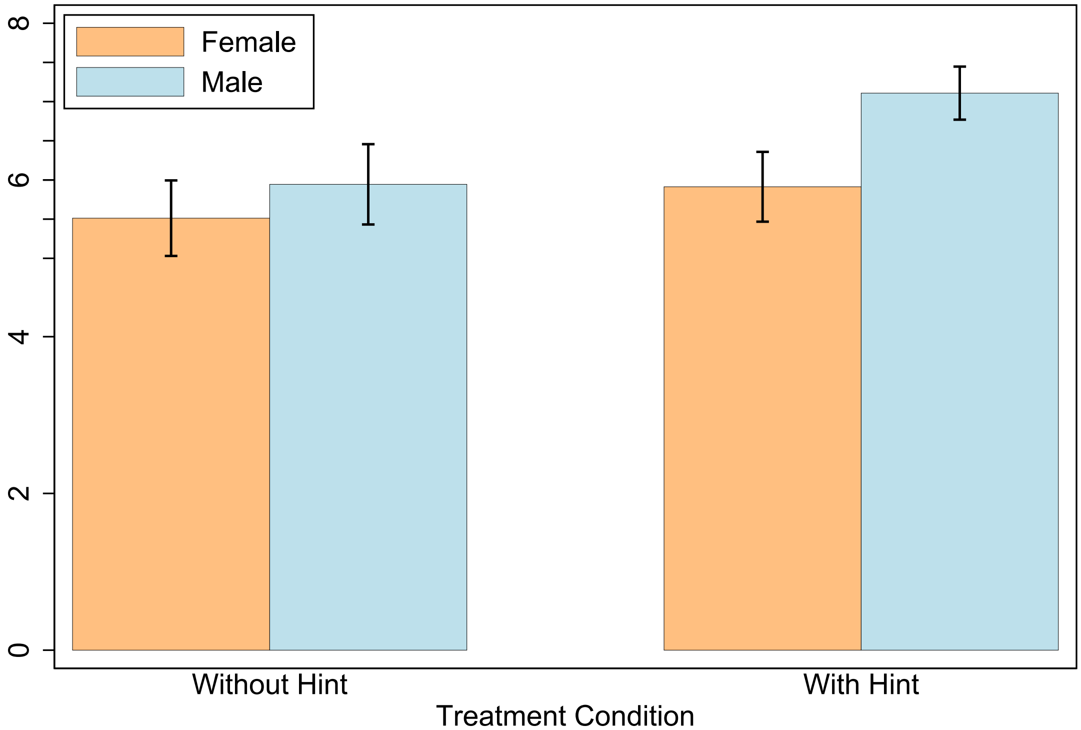


(a) Control Group



(b) Treatment Group

Figure 7: CDF of the Proportion of Correctly Solved Problems in the Second Stage: Treatment vs. Control Group



Note: Error bars represent the 90% confidence interval.

Figure 8: Self-evaluation of Second-Stage Performance by Gender

Table 1: Balance Check

Variables	Control	Treatment	p-value
Age	22.72(.46)	23.31(.41)	0.342
Male	0.47(.06)	0.45(.05)	0.784
Undergraduate	2.66(.05)	2.54(.06)	0.123
Ideology to right(1-5)	2.91(.14)	3.08(.13)	0.367
Belong to a religion org.	0.32(.05)	0.34(.05)	0.866
Family income			
<\$15,000	0.12(.04)	0.12(.04)	0.944
(\$15,000, \$35,000)	0.30(.05)	0.30(.05)	0.973
(\$35,000, \$60,000)	0.14(.04)	0.20(.04)	0.306
(\$60,000, \$100,000)	0.14(.04)	0.11(.03)	0.514
> \$100,000	0.30(.05)	0.27(.05)	0.639
Obs.	77	83	

Notes: Standard errors are reported in parentheses. p-values are reported for two-side t-tests. Mann-Whitney U tests report similar results.

Table 2: Time Spent per Question in the First Stage

	Control	Treatment	p-value
All subjects	63.85(2.16)	19.61(1.04)	0.000
	N=77	N=83	
Male	62.47(3.22)	19.95(1.73)	0.000
	N=36	N=37	
Female	65.07(2.94)	19.35(1.27)	0.000
	N=41	N=46	
p-value	0.552	0.775	

Notes: Time spent is measured in seconds. Standard errors are reported in parentheses. p-values are reported for two-side t-tests. Mann-Whitney U tests report similar results.

Table 3: Task Type Combination Comparison

	Effort + Effort	Effort + Raven	p-value
Switch point	8.78(4.95)	8.24(5.20)	0.733
Self-evaluation	6.06(1.94)	6.11(1.68)	0.945
Time spending per question in the 1^{st} stage	39.15(26.74)	42.05(26.67)	0.330
Time spending per question in the 2^{nd} stage	36.06(9.93)	36.45(11.10)	0.789
Percent of correct answers in the 2^{nd} stage	0.74(.18)	0.72(.15)	0.325
Obs.	63	97	

Notes: Standard deviations are reported in parentheses. p-values are reported for two-side Mann-Whitney U test. t-tests report similar results. Self-evaluation measures the self-reported evaluation of individual second stage performance compared to the rest of the participants in the session. 10 = better than 100% of others, 0 = no better than any others. Time is measured in seconds.

Table 4: Switching Point Comparison

	Control	Treatment	p-value
All subjects	8.68(.60) N=77	8.24(.55) N=83	0.592
Male	9.50(.77) N=36	6.81(.85) N=37	0.022
Female	7.95(.89) N=41	9.39(.67) N=46	0.194
p-value	0.198	0.018	

Notes: Standard errors are reported in parentheses. p-values are reported for two-side t-tests. Mann-Whitney U tests report similar results.

Table 5: Difference-in-Difference Estimates of Gender Gaps in Switching Points

	(1) Switching Point	(2) Switching Point	(3) Switching Point	(4) Switching Point
Hint Treatment	-2.689**	-2.735**	-3.952***	-4.036***
	(1.146)	(1.155)	(1.345)	(1.308)
Female	-1.549	-1.953*	-2.490*	-1.920
	(1.175)	(1.147)	(1.270)	(1.394)
Hint Treatment*Female	4.129**	4.272***	5.804***	5.357***
	(1.599)	(1.620)	(1.692)	(1.739)
Constant	9.500***	9.702***	6.110	7.442
	(0.767)	(0.791)	(7.167)	(7.301)
Session Fixed Effects	No	Yes	Yes	No
Task Fixed Effects	No	No	No	Yes
Demographic Variables	No	No	Yes	Yes
Observations	160	160	160	160

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ Robust standard errors are reported in the parentheses.

In the third column, we further control the self-reported evaluation and ability (proxied by the second stage performance) to exclude any confounders from individual confidence.

Tobit estimation (not reported here) censoring at the switching point between 0 and 17 yields similar point estimates across the three specifications.

Table 6: Gender Differences in the Second Stage Performance

	Female	Male	p-value
Panel A: percent of correct answers			
Pooled	0.74(.02)	0.72(.02)	0.521
	N=87	N=73	
Control	0.75(.03)	0.73(.03)	0.589
	N=41	N=36	
Treatment	0.72(.03)	0.71(.02)	0.693
	N=46	N=37	
16 Questions	0.81(.03)	0.74(.04)	0.140
	N=33	N=27	
24 Questions	0.70(.02)	0.71(.02)	0.604
	N=54	N=46	
Panel B: time spending per question			
Pooled	35.24(1.16)	37.57(1.21)	0.167
Treatment	34.15(1.66)	38.14(1.38)	0.077
Control	36.45(1.60)	36.99(2.02)	0.835
16 Questions	33.07(2.40)	38.14(2.45)	0.148
24 Questions	36.56(1.14)	37.24(1.29)	0.695

Notes: Standard errors are reported in parentheses. p-values are reported for two-side t-tests. Mann-Whitney U tests report similar results. Time is measured in seconds.

Table 7: Difference-in-Difference Estimates of Gender Gaps in Self-reported Performance

	(1) Self-evaluation	(2) Self-evaluation
Hint Treatment	0.214	0.267
	(0.372)	(0.381)
Male	0.415	0.509
	(0.413)	(0.462)
Hint Treatment*Male	1.144**	1.159**
	(0.522)	(0.538)
Constant	5.529***	4.210**
	(0.281)	(1.625)
Session Fixed Effects	Yes	Yes
Demographic Variables	No	Yes
Obs.	160	160

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Robust standard errors are reported in parentheses. Self-evaluation measures the self-reported evaluation of individual second stage performance compared to the rest of the participants in the session. 10 = better than 100% of others, 0 = no better than any others.

Appendices

Appendix A: Analytical Framework

To incorporate external assistance into a simple utility-maximizing framework that facilitates the estimation of the treatment effects of external assistance, we assume that a decision maker maximizes his/her utility by trading off between potential monetary rewards and the convenience from a lower level of effort. This trade-off is revealed by the choice of alternatives A or B through the 17 MPL choice pairs.

Theory

Consider a canonical Cobb-Douglas utility function in which each agent, i, trades-off between monetary payoff and "leisure":

$$U_i = M^{\alpha_i} L^{1-\alpha_i}, \tag{2}$$

where M refers to the expected monetary payoff; the budget constraint during the second stage of the experiment is represented by $W = L + E_{hint} + E_{nohint}$, where L indicates individual leisure.[12] A subject can allocate his/her total budget into three parts: the effort used to solve questions with hints (E_{hint}), the effort used to solve questions without hints (E_{nohint}), and the rest from which subjects can derive utility as "leisure." With $0 \leq \alpha \leq 1$, $1 - \alpha$ measures the agent's subjective valuation of the relative importance of "leisure."

The expected monetary payoff for a risk-neutral subject is the expected earnings from both types of questions: $M = H \cdot Pr_{hint} + (TQ - H) \cdot Pr_{nohint}$, where Pr_{hint} and Pr_{nohint} are the probabilities of correctly solving questions with and without hints, respectively, while H and TQ are the number of questions with hints and the total number of questions for each option. We assume that $Pr_{hint} > Pr_{nohint}$. In our sample, the actual probabilities can be obtained from the proportion of correctly solved questions in the second stage (i.e., $Pr_{hint} = 85\%$ and $Pr_{nohint} = 61\%$).[13]

Without any loss of generality, we define k as the required effort per question without hints. Time spent on each question can be viewed as a measurement for devoted effort. In the first stage, subjects take on average 20 seconds to solve each hint-facilitated question and 64 seconds for questions without hints. Then we assume the effort for each hint-facilitated question is $\gamma * k$, where $\gamma = 20/64 \approx 0.31$ (see Table 2).

[12] We interpret the total endowment W of each subject as a person's energy or total cognitive load, and we assume that this is likely to be similar over the participants in our sample.

[13] We use the exact probabilities to simplify the following derivation. The choice of these probabilities only affects the switching points and the magnitude of the treatment effects; it does not change the predictions and hypotheses associated with the theory.

With the log transformation of equation 2, we obtain the additive utility function for choosing option A and B as

$$V(A) = \alpha \cdot ln(0.85 \cdot H(A) + 0.61 \cdot (TQ(A) - H(A))) + (1-\alpha) \cdot ln(W - 0.31 \cdot k \cdot H(A) - k \cdot (TQ(A) - H(A))) \tag{3}$$

and

$$V(B) = \alpha \cdot ln(0.85 \cdot 8 + 0.61 \cdot (TQ(B) - 8)) + (1-\alpha) \cdot ln(W - 0.31 \cdot k \cdot 8 - k \cdot (TQ(B) - 8)) \tag{4}$$

For example, in the first row of the MPL, $H(A) = 16$, $TQ(A) = 16$, $H(B) = 8$, and $TQ(B) = 24$. Define $F \equiv V(A) - V(B)$. Then an agent chooses A if $F > 0$ and B otherwise. As a result, subjects would make the switch from option A to B when F changes sign from positive to negative or vice versa. Approximately, at the switching point we should have the condition $F \approx 0$. For a generic utility maximizer, this condition yields the following equality:

$$\begin{aligned} &\alpha \cdot ln(0.24 \cdot H(A) + 9.76) + (1-\alpha) \cdot ln(W - 0.31 \cdot k \cdot H(A) - k \cdot (16 - H(A))) \\ =&\alpha \cdot ln(16.56) + (1-\alpha) \cdot ln(W - 2.48 \cdot k - 16 \cdot k) \end{aligned} \tag{5}$$

By implicit differentiation of equation 5, we have $\partial H(A)/\partial\alpha = -F_\alpha/F_{H(A)}$, where

$$F_{H(A)} = \alpha \cdot \frac{0.24}{0.24 \cdot H(A) + 9.76} + (1-\alpha) \cdot \frac{0.69 \cdot k}{W - 0.31 \cdot k \cdot H(A) - k \cdot (16 - H(A))} \tag{6}$$

and

$$F_\alpha = ln(0.24 \cdot H(A) + 9.76) - ln(16.56) - ln(W + 0.69 \cdot k \cdot H(A) - 16 \cdot k) + ln(W - 18.48 \cdot k) \tag{7}$$

It is straightforward to show that $F_{H(A)} > 0$ and $F_\alpha < 0$, since $H(A) \in [0, 16]$, which further yield $\partial H(A)/\partial\alpha > 0$.

Property 1 *Participants have at most one switching point from option A to option B.*

By design, the number of hints in option A decreases for every subsequent decision. $F_{H(A)} > 0$ indicates that F is monotonically increasing in the number of hints ranging from 0 to 16, and the utility of option A is descending in the rows. This suggests that the switching direction between the two alternatives goes from option A to option B and the switching point is unique conditional on a given α. If a subject preferred option B over option A in the nth row, he/she would always choose option B after the nth row. Except for some subjects with extremely low α, who may never switch, theory predicts a unique switching point from option A to B.

Property 2 *A higher α results in earlier switching from option A to option B.*

This prediction is a result of the inequality condition $\partial H(A)/\partial \alpha > 0$ shown in section 5. Intuitively, it means that the position at which participants switch from option A to B is determined by the individual's evaluation of monetary rewards relative to the convenience of lower effort from the hint-facilitated questions. The lower a participant values the monetary payoff (smaller α), the more likely he/she would continue to enjoy the convenience of an easy task and switch at a later point (smaller $H(A)$).

Property 3 *Participants who switch from option A to option B at the first decision have the highest magnitude of α.*

The participants who switch from option A to option B at the first decision will always choose option B with higher potential monetary rewards along the MPL, regardless of the difference in the number of hints between the two options. This is because this type of participant weighs the expected payoff significantly higher than the convenience of lower effort from external help (greater α).

Our hypotheses include two arguments below.

Hypothesis 1 *Being treated with external assistance in the first stage can influence an individual's trade-off between monetary rewards and dependence on external assistance by working on a task with less effort.*

Receiving the external assistance in the first stage will most likely affect how individuals evaluate monetary rewards and the disutility from effort, which can be potentially mitigated by the hints. On the one hand, the external assistance in the first stage may result in stronger confidence and more emphasis on the potential monetary reward (greater α). As such, participants treated with hints in the first stage would be more likely to choose the more challenging option (i.e., option B with more questions but not necessarily more hints) in the second stage. Equivalently, the decision makers in the treatment switch from option A to B earlier than the control group. On the other hand, the external help in the first stage might also trigger an individual's dependency on the convenience generated by the lower amount of required effort (i.e., higher amount of external help) and choose option A, which is easier to complete.

Hypothesis 2 *Receiving external assistance has different treatment effects for female and male participants.*

This hypothesis is supported by the discussion in Sections 1 and 2, suggesting that the external assistance treatment in the first stage will drive women to be more likely to have a smaller α and switch to option B later than men.

Appendix B: Other mechanisms

Did ability affect the behavior of men and women differently?

In this section we further test whether the heterogeneous treatment effects by gender are related to different ways in which natural abilities determine men and women's switching choices during the MPL. Following the reasoning in Niederle and Vesterlund (2007), we analyze the switching decisions conditional on performance in the second stage.[14]

The four panels in Figure C1 present the average switching points of men and women in the treatment and control groups conditional on their performance quartile for the questions answered in the second stage. Recall that option B always has higher potential payoff and greater (or equal) number of difficult questions (non-hint-facilitated questions) than option A; before the 9th decision, option A always has more questions with hints than option B. Conditional on the same level of revealed preference, individuals with better (worse) performance level in the difficult questions should switch earlier (later), which would result in a downward-sloping pattern of the average switching point against the performance level in panels a and c.

In the case of easy questions, it is not obvious how an individual behaves according to his/her ability, since individuals with better performance in the difficult questions are also likely to be good at the easy questions, but not necessarily the other way around. If an individual has relatively better performance in both types of questions, it is still better for him/her to switch early, but if he/she can only perform well in the simple questions or in none of the question types, he/she would still be better off by switching later.

Across all the four graphs, the switching patterns do not correspond to the individual ability proxied by the performance quartile. This result suggests that cognitive bias might be critical to explaining the observed gender differences in the treatment effects. Overall, receiving the treatment in the first stage shifted the average switching point by each performance level to a later point for women but to an earlier point for men.

We first analyze the case of questions without hints in panels a and c in Figure C1. In the control condition, women with the lowest ability in the difficult questions on average switched at the 7th decision row, while treated women delayed switching to the 11th decision. For women with better performance in questions without hints, the treatment drove them to make a later switch by a smaller margin than those with weak performance. Women in the 2nd and 3rd quartile of performance level switched on average at around the 8th

[14] Note that the revealed preference of switching pattern is observed before the implementation of the second-stage game. Along the 17 decisions in the MPL, the decision selected for implementation is randomly determined. Second-stage performance is therefore exogenous to the individuals' switching decision in the MPL.

decision in both treatment and control conditions. However, treated women with the highest performance in the difficult questions delayed switching by about two more units than the control cohort, which indicates that strong-performing women are still switching later.

The switching point delay pattern among weak-performing women might suggest two potential channels for the treatment effects. First, it is possible that the treatment helps weak-performing women to reasonably update their beliefs about their own ability and adjust the switching decision to match their true ability. Second, weak-performing women in the treatment group might derive stronger dependency on the external assistance than women in the control condition. However, it is obvious that strong-performing women's switch point is not rational, the induced dependency on the convenience of external assistance from treatment could be the reason driving them to make later switching decision.

Interestingly, the treatment seems to work very differently for men. As seen in panel a of Figure C1, the treatment led men in the lowest-performing quartile to switch even earlier than women in the highest-performing quartile, providing strong evidence that men are over-confident about their true abilities.

In the case of hint-facilitated questions (panels b and d in Figure C1), treated women with relatively low performance level delayed their switching point by 2 to 3 units compared to women in the control group. It is likely that individuals with low ability in the hint-facilitated questions also perform worse than others in the questions without hints. We test this correlation using the proportion of correctly solved questions for weak-performing women.[15] The reported statistics of a Kolmogorov-Smirnov test and a two-sided t-test show that the difference is not statistically significant (p = *0.197* and p = *0.311*). Again, treated females with weak performance (1st and 2nd quartile) in the easy questions made switching decisions more consistent with their true ability but also developed stronger dependency on the convenience of the simplified task. Overall, treated women with the higher performance (3rd and 4th quartile) in the easy questions also delayed their average switching points, with women in performance level 3rd switching late by 2 units.

We then compare differences in switching points between weak- and strong-performing women in both types of questions during the second stage.[16] This difference is not significant (two-sided t-test, p = *0.191*) for strong-performing women but it is significant at the 10% level (two-sided t-test, p = *0.077*) for weak-performing women. The corresponding test statistics for males are different. The difference is significant for both strong-performing men (two-sided t-test, p =*0.039*) and weak-performing men (two-sided t-test, p = *0.009*)

[15] Weak performance refers to a performance worse than the 3rd quartile.

[16] For questions with/without hints, in the following tests weak-performance refers to performance level of 1 or 2 and strong-performance refers to performance levels of 3 or 4.

at the 5% level.

Were the heterogenous switching patterns chosen out of profit maximization?

To further exclude the mechanism that profit maximization is confounded with dependency, we analyze the share of participants earning positive profits in the second stage conditional on performance level. Recall that participants were required to answer at least 75% of the tasks correct in order to be paid in the second round. If women with the lowest ability strategically adjust their switching point in order to maximize their profit, we would observe a significant increase in the probability of obtaining positive earnings. As shown in Appendix Figure C2, this is not the case. The reported statistics of a two-sided t-test indicates the difference is not statistically significant either by performance level or in the pooled data. For the lowest performing men and women, adjustment in their switching pattern after receiving help, actually decreases the probability of receiving positive earnings, although the statistical test is not significant.

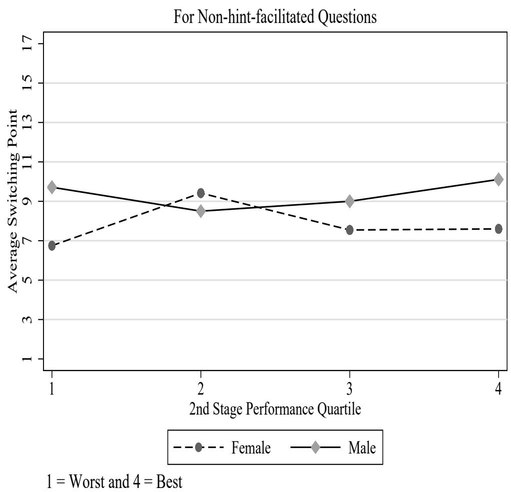
Figure C3 presents the aggregate share of men and women who earned positive payment in the second stage. Both genders have no significant differences between the treatment and control, which further exclude profit maximization as the underlying mechanism for the observed heterogeneous treatment effects.

Taken together, the treatment of external drive women overall to switch late. Although this could lead women with low ability to make their switching decisions more related to their own ability, it might imply that treatment drives the high-performing women to develop stronger dependency on an easier task.[17] The behavioral bias from the treatment drives men—especially weak-performing men—to switch too early. Given the findings of the discrepancy between choices and abilities, we know that besides the monetary payoff, each individual attaches a different degree of importance to the convenience of an easier task while making switching decisions. This subjective valuation might be affected by the treatment intervention, so even men and women with the same ability respond to the treatment in very different ways due to the induced cognitive bias.[18]

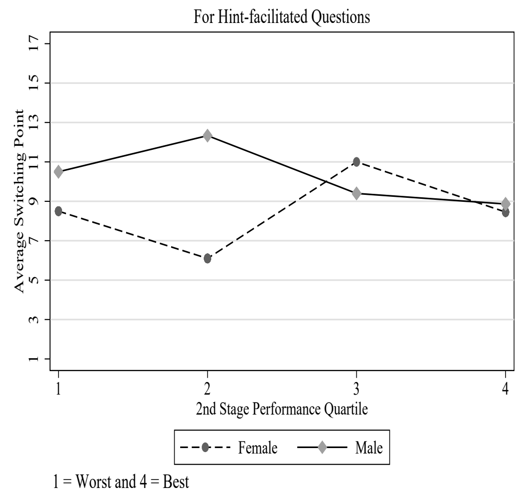
[17] Though we observe treated women with the highest performance level also conservatively switch later, the difference is not statistically significant due large variation. It is possible that strong-performing women also develop a tendency to depend on the convenience of an easy task, but with a smaller magnitude compared to weak-performing women.

[18] See the Appendix A for the theoretical part, this subjective valuation is measured in α.

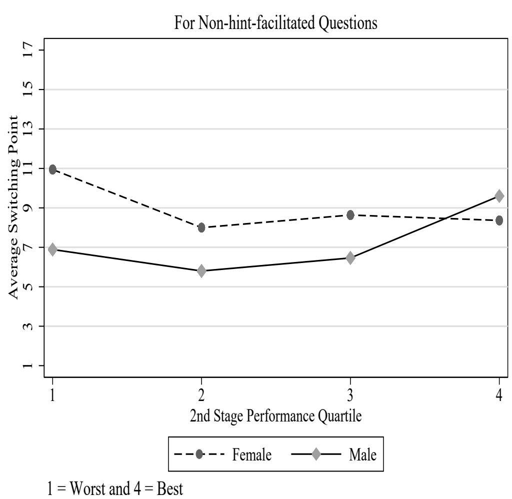
Appendix C: More Figures and Tables



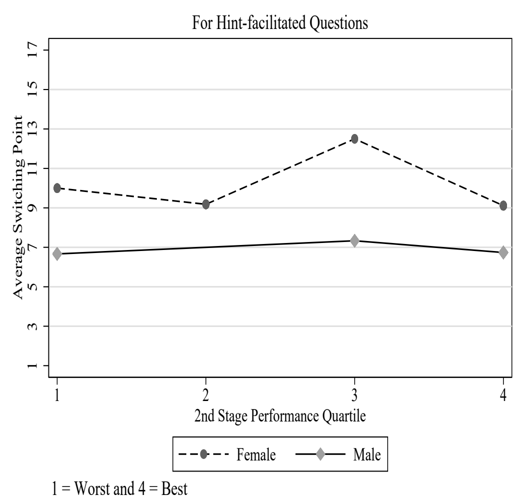
(a) Control Group



(b) Control Group

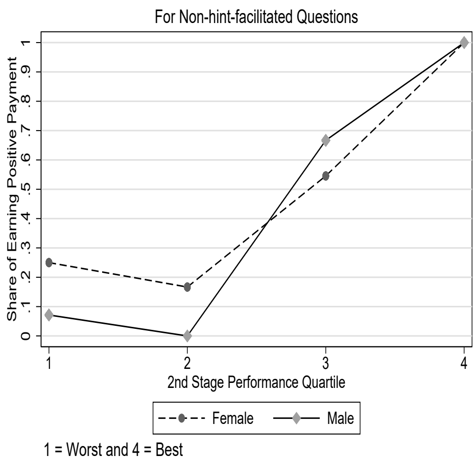


(c) Treatment Group

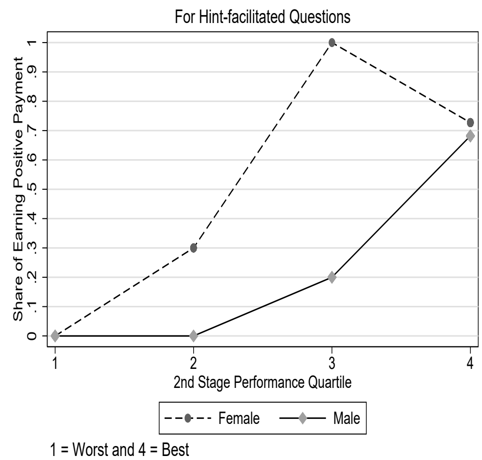


(d) Treatment Group

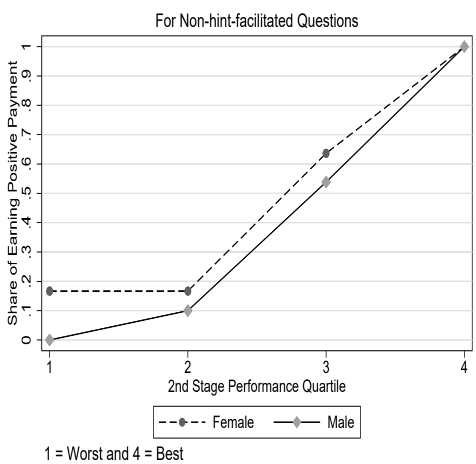
Figure C1: Switching Point Conditional on the Performance in the Second Stage: Treatment vs. Control Group



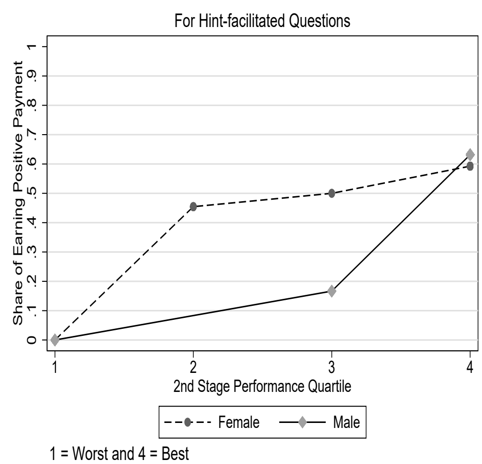
(a) Control Group



(b) Control Group

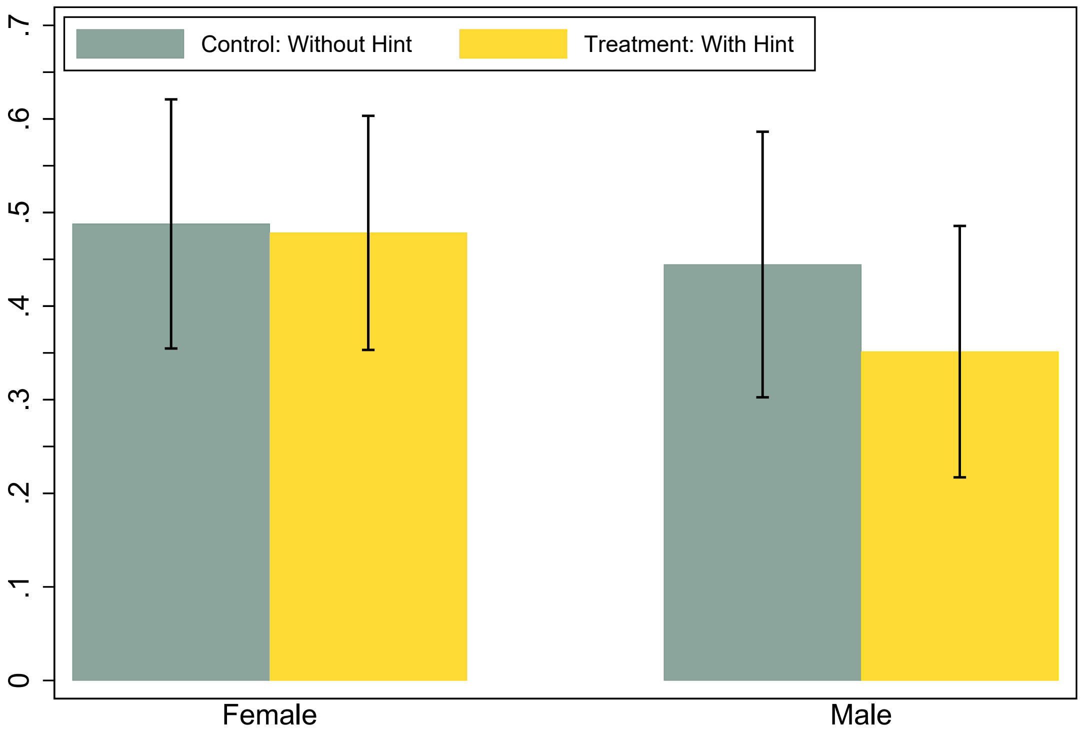


(c) Treatment Group



(d) Treatment Group

Figure C2: Share of Positive Profits in the Second Stage Conditional on Performance: Treatment vs. Control Group.



Note: Error bars represent the 90% confidence interval.

Figure C3: Share of Positive Profits in the Second Stage by Treatment.

Appendix D: Experimental Instructions

General Instruction

Before the session begins, you will carefully read the basic instruction in 15 minutes. Please feel free to ask questions if you are confused. During the experiment you are not allowed to communicate with other participants. If you have a question, please raise your hand. We will come to answer your questions.

Sometimes you may have to wait a short while before the experiment continues. Thanks for your patience and cooperation.

Upon the completion of the experiment, you will receive a participation fee of $5. You will also receive some extra payment based on your responses to the questions. At the end of the experiment your total earnings will be paid out to you in cash.

Before the end of the session, we will ask you some general questions about yourself. Your responses are helpful in that they can be used to explain some of the decisions you make in the experimental exercises. Please note that your responses will not be linked to your name, nor made available to anyone outside the research team. Your ID number is used to match your responses so that they are not confused with anyone else's, and will be used to determine your earnings from the experiment.

We ask you not to talk with anyone else today except for the designated researchers conducting this experiment.

We expect that the entire session will take about one hour. Your participation is completely voluntary. You may ask questions at any time during the experiment.

First Stage Instruction

In this stage, you are going to work on one of the two types of tasks. One type of task will list questions with hints which will assist you to finish them, while the other type of task will list questions without hints. A sample for both type of tasks will be on the next screen for your reference.

After you review the sample question, the computer will randomly, with probability one half, assign you to work on either one type of the tasks.

This experiment is completely anonymous: neither the other participants, nor the organizer will be able to know what your decision was.

Questions in both of the tasks require you to count the number of occurrences of a prespecified letter appearing in a table of several lines' random combination of 26 letters. There are 50 letters in each line, you will need to go through all these lines to correctly count the frequency of this exact letter.

In the type of task without hints, you are asked to go through every letter across these lines counting for the number of the requested letter. In the type of task with hints, all irrelevant letters will be suppressed to narrow down the counting area with the purpose of assisting you to easily complete the task.

There are 10 questions in total and you have to correctly answer all the questions within 15 minutes to earn $10.

Participants see the sample question on the screen (Figure 1). In this example, they are asked to count the number of letter 'h'.

After they finish the first-stage task of ten questions, they enter the second stage

The Second Stage Instructions (Real Effort Sessions)

PLEASE CAREFULLY READ BELOW INSTRUCTION AND EXAMPLE IN THE NEXT PAGE SINCE IT MATTERS FOR YOUR FINAL PAYOUT.

Now you are in the second stage. In this stage, you are going to work on another batch of task. Each question requires you to find a pre-specified number from a table of several lines' random number combination. There are 50 numbers in each line, you will need to go through all these lines to correctly count the number of times the pre-specified digit appears. Please click Proceed to see the sample question.

Then participants see an example similar with the first stage with replacement letters by numbers. The participants in the raven test sessions see the following instructions instead.

The Second Stage Instructions (Raven Test Sessions)

PLEASE CAREFULLY READ BELOW INSTRUCTION AND EXAMPLE IN THE NEXT PAGE SINCE IT MATTERS FOR YOUR FINAL PAYOUT.

Now you are in the second stage. In this stage, you are going to work on another task of test questions. Each question requires you to select the one choice in the answer panel that best fits in the blank position of the question. Please click Proceed to see the sample question.

In the below tables, you will need to select the best fit answer. The left panel is the example of question without hint. A full set of choices will be presented to you. The right panel is the example of question with hint. Some irrelevant choices will be suppressed with the purpose of assisting you to narrow down choice pool.

Multiple Pricing List Instruction

There is a list of option pairs, “A” and “B”. Option “A” has 16 questions and option “B” has 24 questions.

Both option “A” and “B” have some questions with hint and others without hint. The number of hints for option “A” varies across the list. The number of hints for option “B” is 8 and keeps constant through the list.

You are required to check the one option that you prefer to implement for EACH option pair in the list. Only ONE pair in the list will be randomly chosen for execution and your choice of option in that pair will be presented to you for implementation later.

You first have 10 minutes to make selections across the list, then 20 minutes to finish the task.

You will see the details in the next screen.

Each correct answer will be worth $1.

To earn payout, the number of mistakes that you are allowed to make is at most one fourth of the total number of questions. Specifically, in Option A you are ONLY allowed to give at most 4 wrong answers and in Option B you are ONLY allowed to give at most 6 wrong answers. You will receive the payment corresponding to the number of your correct answers, only if your number of correct answers exceeds 12 (including 12) in Option A and 18 (including 18) in Option B. Otherwise, you will NOT receive payment for this stage.

You will be ONLY aware of how many correct answers in the end of this stage. No notification will show at each question.