28/09/2022

Gender and the demand and supply of advice
Aman-Rana, Chaudry, Delfino

Pre-Analysis Plan (PAP)

DATA COLLECTION

TIMELINE

The beginning of data was delayed because of the need to do further piloting in the chosen university and because of logistical issues with internet devices and experimental software. The actual start date of the intervention is the 28th of September.

RECRUITMENT

We initially visited the university and met the Head of department to give them an explanation regarding the project and then requested them to allow us to conduct study. We then shared a flyer to be put up for students across the department. Lists of students willing to participate from each course were shared with the research team by the university administration, including possible dates for conducting the study.

For logistical and internet reasons, we had to limit the number of participants per session to 16 people, which we will equally split between men and women if possible. If less than 16 people show up, we keep the full group. In case that more than 16 students of each gender show up, here is how the sampling works every day: students who agreed to participate come to the location communicated to them (auditorium or classroom). There are two bags with numbers from 1 to the total number of female and male students, respectively. An enumerator asks each person (privately and one by one) to draw a number. If the number is lower than 8, the person stays. If higher, the person leaves and will be called back for a different session later on (if she still wants to participate). A small fee is provided also to those who are asked to leave. We pay extra attention not to make gender salient, thus gender is never mentioned or shown on the bags.

SAMPLE SIZE

We expect to be able to reach a sample of 200 people in the partner university (University of Punjab) and to get at least other 150 people from a different one.

POWER ANALYSIS

We perform power analysis with repeated measures at the individual level, two subgroups (men/women) and three conditions: match type (preferred/random), match gender (man/woman) and uncertainty of hint supply (yes/no). We will need a sample size of 322 people to detect an average difference in hint-taking behaviour across conditions of 0.33 standard deviations and a power of 0.9, assuming a correlation in measures between levels of each condition of 0.8 (0.5 for the gender of the match condition) . We would need 452 subjects for differences scaled by a half and the same power.

EXPERIMENTAL DESIGN

The experiment consists of a within-subject design, with three conditions. One condition is whether a person is matched with a male or female helper. Another condition is whether the helper is random or

among the preferred by the test-takers (see below). The third condition is whether there is uncertainty on the availability of the hints. Furthermore, the helpers also play in two conditions: whether they get a fixed payment for helping or whether their payment depends on the test-taker's performance.

The following paragraphs describe the design in more details.

Participants in the study will be allocated to two different roles: test-takers and helpers. Everyone plays in both roles. Test-takers have to reply to multiple-choice knowledge questions and helpers have to provide hints. The knowledge questions are divided in three knowledge areas: specific subject studied by the participants, Cooking and Sports. The former knowledge area can be economics, IR, political science or sociology depending on the major the participants are enrolled in in the university. Helpers can decide whether to give a hint or not, but have no control on what the hint says: all the hints are taken from a pool prepared by the researchers, in order to guarantee the same average quality of advice across questions.

The experiment is implemented through tablets without any physical face-to-face interaction between participants.

Individuals who participate in this study will go through five parts, after a first round of practice. Each parts of the experiment is described as follows:

1. In part 1, participants are test-takers and solve 12 multiple-choice questions (4 in each knowledge area). In this part, participants are helped by the computer. Three out of four questions in each knowledge area have a hint available, one will not. For all the questions with hints available, participants are sure to get a hint if they ask for it. The order of the questions and knowledge area will be block-randomized.

2. In part 2, participants are test-takers and solve 12 multiple-choice questions (4 in each knowledge area). In this part, participants are helped by the computer. Three out of four questions in each knowledge area have a hint available, one will not. For all the questions with hints available, participants are NOT sure to get a hint if they ask for it, and they know there is a 66% chance that the computer will provide a hint within each knowledge area. The order of the questions and knowledge area will be block-randomized. *NOTE: For logistical reasons related to the duration of the experiment, this part will be implemented only in some sessions of the experiment.*

3. In part 3, participants will be helpers and will be randomly matched with 4 different test-takers. Using a strategy-method, for each match we ask the helper to choose how many hints to provide in each knowledge area to each test-taker (between 0 and 3). They have a maximum of 10 hints to allocate across the four test-takers in a given knowledge area, but they can choose to allocate less than that amount. We tell the helper that their choice will be implemented should the match realize in one of the following rounds of the game, and that they will be able to see the number of hints asked for by the test-takers. If the match is realized, the helper will be paid a flat rate of 450 PKR. We additionally ask the helper to predict i) the performance of the test taker in each knowledge area without hints available, ii) the number of hints the test-taker will ask for, iii) the performance of the test taker in each knowledge area when hints are available (in aggregate, considering both questions with and without hints). The order of the questions and knowledge areas will be block-randomized.
NOTE: depending on the group size and gender ratio, the number of matches may be lower than 4 for a given helper. If the helper is matched with 3 test-takers, the hints available are 7. If the helper is matched with 2 test-takers, the hints available are 5. If the helper is matched with 1 test-taker, the hints available are 2.

3. Part 4 is exactly like part 3, but with one main difference. If a given match is realized, the helper will be paid 75 PKR for each correct answer given by the test-taker.

The order of parts 3 and 4 will be randomized.

4. In part 5, participants will be test-takers and will be matched with 4 different helpers whose choices have been elicited in rounds 3 or 4 (we will pick round 3 or 4 randomly). While matched with a given helper, test-takers will again solve 12 multiple-choice questions (4 in each knowledge area). Thus in total they will solve 48 questions in this part. Three questions in each knowledge area will have a hint available, one will not. If the test-taker wants to ask for a hint, he/she has to press a button. The test-taker knows that the helper will know the number of hints s/he asked for. In this part, the hint is always available if the person asks for it, so there is no uncertainty on the supply of the hint. The order of the questions and knowledge areas will be block-randomized.

5. Part 6 of the game is exactly like part 5, but with one main difference. In this part, the hint may not be released even when the test-taker asks for it. This depends on the choice made by the helper in parts 3 or 4. For instance, suppose that A is a test-taker and B is the helper. B declared in part 3 that he/she wants to give 0 hints in Cooking to A. This means that when A presses the button to ask for a hint in cooking, the system will tell him/her that the helper has not released the hint for him/her. Thus there is uncertainty on the supply of hints coming from the helper's choices. The order of the questions and knowledge areas will be block-randomized. In this part, we show the test-taker the choice of hint allocation that the helper had to make and ask to predict the supply of hints for each knowledge area. This question is monetarily incentivized.

The order of parts 5 and 6 will be randomized.

In parts 1, 2, 5 and 6 we ask the test-taker to predict i) his/her performance in each knowledge area without hints available, ii) the number of hints s/he will ask for, iii) the performance in each knowledge area when hints are available (in aggregate, considering both questions with and without hints).

In parts 5 and 6 the matches will be such that each 'test taker' will be paired with four types of helpers: a random woman, a random man, a preferred female helper and a preferred male helper. All the helpers will be selected from the same class where the experiment happens and from amongst the set of 16 participants present in an experimental session. "Preferred female helpers" and "preferred male helpers" are determined in the survey conducted prior to the experiment, where we ask participants to rank 10 classmates they would like to have as helpers.

In parts 3 and 4, the matches between helpers and test-takers are as follows. Each helper is shown up to a maximum of 4 test-takers. Two of them have selected the helper as "preferred" in the initial survey ranking, and two of them have not selected the helper as their preferred. For each of these two categories, one test-taker is a man and one is a woman.

The matches are implemented as follows: Each helper will have 8 slots available for test-takers. Out of these 8, 4 slots (2 men and 2 female test-takers) are for those test-takers for whom that helper was a preferred one (2 men and 2 female test-takers). The rest are allocated to randomly chosen test-takers. The matches are implemented using a random serial dictatorship. The first randomly selected test-taker was matched with the most preferred helper on their list, the second one was matched with their most preferred helper

unless that helper's slots are full. If the slots are full then that test-taker is allocated their second most preferred helper. This process then continued until all test-takers who were due to be allocated a helper were paired with one.

A single part and pairing is randomly drawn for the final payment.
At the end of the experiment and before seeing the payment, every person is shown a summary table with the hint-taking behavior of the test-takers they have helped in each subject (three columns: one on the hints supplied, one on the hints taken by the test-taker and their difference). The same table is also sent by email within 24 hours to everyone.

OUTCOMES

Primary Outcomes

Our first **overall outcome variable** of interest is the performance by participants in knowledge questions: whether a person provided the right answer to a question (as well as the number and % of correct answers to knowledge questions), across and within knowledge areas.

For the **demand for advice**, our main outcome variable is hint-taking: whether a person requested a hint on a question (as well as the total number and % of hints asked by the test-taker to reply to knowledge questions), across and within knowledge areas.

For the **supply of advice**, our main outcome variable is the total number of hints and the % number of hints that a helper decided to provide to a test-taker, across and within knowledge areas.

Construction of primary outcomes

For the variables on performance and demand of advice, we will proceed as follows. In each round, a test-taker is matched with one partner (the helper) and has to reply to 12 knowledge questions (4 in each subject: cooking, sports and subject of study).

First, we will assign an indicator variable equal to one to each question answered correctly (for performance) and a second indicator for whether the person asked for a hint (if the hint was available, for demand for advice). Moreover, we will also look separately at performance in questions that had hints available (9 out of 12 in each round) or that had no hints available. This latter variable allows us to study the direct impact of a treatment on performance when help is not available.

Second, to construct the % of correct answers in a knowledge area, we will divide the number of correct answers in that area over 4. Similarly, to construct the number or % of hints asked by the test-taker overall, we will count the number of hints the test-taker asked for across all the subjects (divided by the number of hints available).

To elicit the supply of hints, we use a strategy method: we show to each helper a set of 4 test-takers they could be matched with during the game. For each knowledge area (e.g., cooking), we ask them to allocate a maximum of 10 hints across the test-takers. In some cases if they are matched with less than 4 test-takers we ask them to allocate hints as follows: a max of 7 hints if they are matched with 3 test-takers, a max of 5 hints if they are matched with 2 test-takers and a max of 2 hints if they are matched with just 1. Helpers can allocate from a minimum of 0 to a maximum of 3 hints per person in that particular knowledge areas, and can also decide to allocate less than 10 hints in total. We will count the percentage of hints given to a certain test-taker over all the hints available to distribute (overall and within a certain knowledge area). We will also count the number of hints as well as the share of hints given to a certain test-taker over the maximum that could be given to him/her (overall and within a certain knowledge area). We will further

explore whether helpers are willing to distribute less than the total for the sake of distributing an equal number of hints to everyone (or other distribution criteria).

Secondary Outcomes

We have two families of secondary outcomes:

- Expectations of supply: we will compare the expected supply of hints with actual supply, to understand whether there are misperceptions related to the helpers' behaviour.
- Willingness to pay for avoiding being seen asking for advice. We have two questions on this. One question uses a BDM mechanism and another is a multiple price list to elicit the willingness to pay for avoiding being seen asking for advice, in order to quantify frictions in knowledge flows within teams. *Our preferred measure is the question using the incentivized BDM mechanism, while the MPL is to be considered exploratory (some pilot data indicate noise in these data).*

Construction of secondary outcomes

Expectations of supply

We tell test-takers that the helper had to allocate 10 hints across the given 4 test-takers (or fewer hints across fewer test-takers, depending on the number of realized matches as explained above). We ask them to predict how the helper distributed the hints, within each knowledge area. We incentivize a correct answer with a monetary bonus of 50 Rs.

We will construct two measures of misperceptions:

1) OwnHints: for a given test-taker, we will compute the difference between actual hints that the helper sent to him/her and the expected number of hints, overall and within each subject.

A positive or negative difference indicates the presence of misperceptions in the supply of hints to theirselves. We will look separately as misperceptions depending on the helper's gender.

2) OverallHints: for a given test-taker, we will compute the average difference between actual hints that the helper sent to each test-taker and the expected number of hints (across all test-takers), overall and within each subject. For instance, if the helper is asked to help 4 test-takers, this variable will be the difference (actual – expected) supply for every test-taker, averaged across all four test-takers. A positive or negative difference indicates the presence of misperceptions in the distribution of hints across different test-takers. We will further compare the difference between actual and expected supply separately for female and male test-takers, to check whether misperceptions are especially related to one's gender. Moreover, we will look separately as misperceptions depending on the helper's gender.

Willingness to pay

The BMD question is computed as follows:

After finishing part 5, each participant is shown again the list of helpers s/he was matched with during the experiment and asked whether s/he is willing to pay some money to stop us from telling them the number of hints they took. They can select from 0 to a maximum of 4 helpers out of the list. For each selected helper, we then ask how much money they are willing to pay, between 1 and 75. A random draw is implemented: should the number be higher, their hint-taking behavior will be still shown to the helpers and they won't pay anything. Should the number be lower, they will pay the stated amount and we will not communicate the hints they took to the given helper in the final table.

Our outcome variables here are twofold:

- Whether the person is willing to pay to hide hint-taking from any helper (0/1 variable)
- For how many helpers the person is willing to pay to hide hint-taking (from 0 to 4)
- The share of male helpers among those that the person is willing to pay to hide hint-taking from (from 0 to 100%)

- The share of "preferred" helpers among those that the person is willing to pay to hide hint-taking from (from 0 to 100%)
- The average amount of money that the person is willing to pay for male/female helpers (ratio)

The Multiple Price list is as follows:
We ask people to choose repeatedly between two options, A and B. They are told to think back to their experience as a test-taker: they have to answer one question and have the option to get a hint at a cost. In option A, they take the hint privately. This means that nobody will see that they took a hint. In option B, they take the hint in front of a helper. This means that the helper will know whether they took the hint or not. In either option, the hint is exactly the same and they decided to take the hint (there is no option of NOT taking the hint). We propose the choice between A and B for 7 times. The cost of option A is fixed at 5 and the cost of option B is going from 1 to 7. We repeat the question twice, once with the person matched with a female helper and one with the person matched with a male helper (randomly drawn from the list of all participants in the session). We also randomize across subjects whether the anchored option A is "nobody will see the hint" or whether it is "the helper will see the hint". We are going to measure the switching point from B to A for every person when matched with a man or a woman.

Additional measures
For exploratory heterogeneity and to look into mechanisms, we also measure the following dimensions:

- Expectations of own and others' performance: we ask people how they think they will perform as well as how they think their test-takers will perform. We ask expected performance with two questions: i) for the case with no hints available, ii) in aggregate, considering questions with and without hints available. We will look into whether performance expectations explain the supply of advice as well as the expected supply of advice from others, by gender and subject.

- Expectations of hint taking: we will ask test-takers how many hints they expect to take, and helpers how many hints they expect the test-takers – the ones they are matched with - to ask for in each subject (out of the hints available). We will look at the extent to which these expectations, by the test-taker and the helpers, align with each other and whether this can explain supply. This has to be considered as an additional measure of potential misperceptions in the need for help, which can contribute to explain supply.

- Confidence in own performance: for each single multiple-choice question, we ask the test-taker to answer the following: "I think my answer as a ___ % of being right. A randomly drawn computer should answer for me if its accuracy is greater than that." This gives us an indication of the confidence in own performance. We will explore whether being matched with different helpers affect confidence, conditional and unconditional on performance.

- Confidence in expectations of supply: this will be measured by asking directly the confidence level in the expected supply question (only for the test-taker's own hints). We will look at this by gender of the test-taker and gender of the helper, by subject.

- People's preferences for helpers to be matched with during the game. At the beginning of the experiment, we ask people to provide a ranking of their top 10 classmates that they would like to have as helpers during

the activity (specifying the reason). This is the ranking which is used for the actual matches that happen during the experiment with "preferred" helpers.

- Usefulness of hints: after asking for a hint (and receiving it), we ask the test-taker if they thought the hint was useful (yes/no). We will check the balance in the perceived usefulness of hints across parts of the experiment, by subject and gender, to make sure that differences in performance or demand for advice are not explained by this.

- Whether people would like to change their supply decisions after knowing the hint-taking behavior of their helpers. This measure is asked at the very end of the experiment, after people are made aware of the hint-taking behavior of their helpers. Given the length of the experiment, we cannot exclude that answers will be noisy because of fatigue. Thus this measure has to be considered exploratory.

We implement a final survey after the end of part 6 which asks the following additional families of variables:

- Stereotypical associations between a knowledge area (the three used in the experiment) and gender.
- Willingness to hide hint-taking behavior from different matches, by subject
- Ranking of the top 5 people that a person would have liked to help during the game*. NOTE: this question is non-mandatory, thus we will consider whether to use it or not depending on response rates.*
- Demographics and social relationships with peers
- GPA
- Gender attitudes (WVS questions)
- Trust in men and women
- Personality (big 5) and preferences (from Global Preference Survey, Falk et al., 2018)
- An incentivized measure of risk aversion using the Gneezy-Potters-Charness investment task (Charness & Gneezy, 2010; Gneezy & Potters, 1997)
- Ranking of helpers that they worked with on the ease with which they could ask for help
- Reasons for giving hints and asking for hints (or not asking)

EMPIRICAL STRATEGY

Main empirical specification

Our objective is to measure how being matched with a person of the same or different gender affects: i) advice seeking, ii) advice provision and iii) performance, and whether this differs for men and women. So all the main specifications have to be intended separately for men and women.

Our first specification is:

$$y_{ipm} = \alpha + \sum_{j=1}^{2} \beta_j \, Match_{ipm}^{j} + \phi_p + \delta_i + \epsilon_{ipm}$$

where y_{ipm} is the outcome of interest for individual *i,* in part *p* and match type *m*. The outcome can be the total number of hints asked, total number of correct answers and total number of hints provided across knowledge areas in a given part of the experiment and match *m*.

We will also enrich our specification by considering differences across subjects, using the following:

$$y_{ispm} = \alpha + \sum_{j=1}^{2} \beta_j \, Match_{ispm}^{j} + \lambda_s + \phi_p + \delta_i + \epsilon_{ispm}$$

where y_{ispm} is the outcome of interest for individual *i,* in subject *s,* in part *p* and match type *m*. The outcome can be the total number of hints asked, total number of correct answers and total number of hints provided in knowledge area *s* in a given part of the experiment and match *m*.

The variable $Match_{ispm}^{j}$ is defined as 0 for the baseline round (replying to questions alone), 1 for a match with a woman and 2 for a match with a man. $\lambda_s, \phi_p, \delta_i$ are knowledge area, part and individual fixed effects. We will do robustness with a specification at the question level. We will test for each $\beta_j = 0$ and for $\beta_1 = \beta_2$.

In a second step, we will modify our specification to split the sample by parts 5 and 6 (and removing the part fixed effects) to study the effect of uncertain supply. Besides the tests above, we will test the equality of the effect of a match with a man or woman on outcome y_{ispm} in parts 5 and 6 (without or with uncertainty of supply). A fully interacted model will be used for robustness.

In a third step, we will modify our specification to be:

$$y_{ispm} = \alpha + \sum_{j=1}^{4} \beta_j \, Match_{ispm}^{j} + \lambda_s + \phi_p + \delta_i + \epsilon_{ispm}$$

Where now the variable $Match_{ispm}^{j}$ is defined to be 0 in the baseline condition (replying to questions alone), 1 for being matched with a "preferred" male partner, 2 for a "random" male partner, 3 for a "preferred" female partner and 4 for a "random" female partner. We will test for each $\beta_j = 0$, for $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$, and for $\beta_1 = \beta_3$ and $\beta_2 = \beta_4$.

In a fourth step, we will repeat the split by parts 5 and 6 using this four-level definition of the *Match* regressor.

For the supply of advice, our main specification is:

$$y_{ispm} = \alpha + \beta \, Match_{ispm}^{Male} + \lambda_s + \phi_p + \delta_i + \epsilon_{ispm}$$

Where y_{ism} is the number of hints (or share of hints available) that a person *i* gave to different test-takers in knowledge area *s*. The variable $Match_{ispm}^{Male}$ is defined as 1 when the helper is giving hints to a man and 0 when to a woman. λ_s, δ_i are knowledge area and individual fixed effects and ϕ_p is an indicator for being in part 3 or 4.

We will also look into the interaction of helping a man depending on whether the man chose the helper as "preferred" or not ($Match_{ispm}^{Pref}$):

$$y_{ispm} = \alpha + \beta_1 \, Match_{ispm}^{Male} + \, \beta_2 \, Match_{ispm}^{Pref} + + \beta_3 \, Match_{ispm}^{Male} * \, Match_{ispm}^{Pref} + \, \lambda_s + \phi_p + \, \delta_i + \epsilon_{ispm}$$

To understand whether incentives to supply advice matter, we will split the sample by part 3 and 4 (fixed-payment or performance-related payment) and test the difference of the beta coefficients between these parts.

Estimation of heterogeneous treatment effects

Our analysis of heterogeneous treatment effects will focus (1) gender and (2) knowledge area. Beyond that, we will explore the data and report other potential heterogeneities that may be relevant to understanding mechanisms of treatment or that may be relevant for policymaker in this context (such as the gender stereotype of a knowledge area). To discipline this exploratory analysis we will consider using the generalized random forest algorithm by Chernozhukov, Demirer, Duflo, and Fernandez-Val (2018).

We expect gender and the knowledge area to impact performance as well as demand for and supply of advice. Heterogeneous treatment effects will be estimated by splitting the sample by gender and, separately, by gender and knowledge area.

Multiple hypothesis testing correction

Following Benjamini, Krieger, and Yekutieli (2006), we will use false discovery rate corrections to account for multiple hypothesis testing across our primary outcome variables. We will do FDR corrections for primary outcomes, reflecting our belief that results in each domain are of separate interest. Since our mechanism outcomes are included to help us identify channels through which the treatment affected our primary outcomes, we will not include the secondary outcomes in the multiple hypothesis testing adjustment for the primary outcomes.

Because of the presence of multiple treatments, we will also follow List, Shaikh and Xu (2019) to correct for FWER across primary outcomes.

MAIN HYPOTHESES

We have the following main hypotheses, drawing on the working paper of Aman Rana et al. (2022).

WOMEN

1) Women matched with a male helper will have lower performance than women matched with a female helper.
2) We hypothesize that lower performance may be (at least in part) explained by
    a. lower willingness to ask for help, especially in stereotypical female-subjects.
    b. lower supply of hints by men, especially in stereotypically female-subjects.
3) We will test whether lower willingness to ask for help is due to correct expectations that men won't provide help, or whether is due to misperceptions of help provision.

4) We will try to understand mechanisms by distinguishing whether women do not ask for help due to image concerns or misperceptions of supply (and/or fear of rejection) by exploring differences in experimental parts with certain or uncertain supply.
5) We expect women to have a higher willingness to pay to hide their hint-taking behavior from men.
6) We will investigate whether the willingness to supply hints is based on any helper's (mis)perceptions of ability or their (mis)perceptions about how many hints would be needed by the female test-taker.
7) We will test two different policies that may help alleviate the gap in performance and demand/supply for advice in mixed-gender pairs: choosing a "preferred" helper and having the helper be compensated based on performance.

MEN

Hypotheses for men may be symmetric to women's hypotheses, or not. In any case, we expect men's behavior to also be sensitive to the stereotypical nature of the knowledge area as well as the gender of the match.