Pre-analysis Plan

Does Better Information Reduce Gender Discrimination in the Technology Industry?[*]

Ashley Craig and Clémentine Van Effenterre

Thursday 16th February, 2023

Background

We aim to study patterns of behavior of people participating in coding interviews, and their impact on gender disparities in labor market outcomes. We have partnered with an online platform that provides job applicants in the technology industry with an opportunity to practice their interview and coding skills. Our project asks the following **research question**: Does the identity of the evaluator, and the information available to them, affect the evaluation of the code quality written by the applicant?

Our experiment will use a large set of de-identified code blocks written by a set of men and women on the platform. This would span coders of different skill and problems of different levels of difficulty. For each code block, we will have access to the platform's objective measures of performance including sub-test results (e.g., whether it runs, whether it produces correct answers to unit tests etc.).

Using these data, we will ask evaluators to judge the quality of the code using the same Likert scales as on the platform. The evaluators will be Bachelor or Masters level computer science students who have familiarity in the relevant programming languages. If we cannot source enough of the students, we may expand our pool to include other skilled evaluators (e.g., hired via Qualtrics).

Depending on the treatment condition, the evaluator will be aware of the gender or other basic information about the programmer who wrote the code (but will never be given identifying information). Using these evaluations, we will ask: (i) whether there are perceived differences in the quality of the code written by men and women; and (ii) how those perceived differences change when the evaluator is aware of the gender of the coder. This will let us test whether there are any unobservable dimensions of performance that are correlated with gender and driving the residual gender gaps that we see in our data. To examine particular

[*] Craig, University of Michigan, ashcraig@umich.edu, Van Effenterre: University of Toronto, c.vaneffenterre@utoronto.ca. The investigators have no conflicts of interest to declare.

dimensions of performance, we will also pre-register different dimensions of the code (length, time of program execution, number of comments, maintainability of code).

The goal of this experiment is to assess whether unobserved differences in the code written by men and women drive the gender gaps that we see, because male and female software developers code differently; or whether gender bias plays an important role. It will also shed light on whether blind recruitment is enough to improve diversity in hiring. In an additional incentivized version of the experiment, we will ask evaluators to predict the subjective ratings of candidates, and we will condition their payoff on how close their prediction is from the score provided by the actual evaluator. This version will allow to isolate an individual's preference from his/her beliefs about what evaluators on the platform value.

Experimental Design

Recruitment Our subject population is comprised of recent graduates or students currently enrolled in computer science programs. We will recruit evaluators through universities' graduate (and potentially undergraduate) programs. Most personal correspondence will occur via email. Our recruitment email discloses that we are studying the role of coding skills on future labor market outcomes, but does not explicitly mention gender.

We will pay participants a fixed completion fee of $10, at a piece rate of $10 script they evaluate, plus bonus payments of $2 for each accurate predictions they make for the objective code quality measure and hireability measure per code block. The ten best evaluators will earn a cash prize of $500. Additionally, we provide a symbolic but potentially powerful incentive selecting a set of 3 evaluators to the Creative Destruction Lab 2023 Super Session which brings together world-class entrepreneurs, investors and scientists with high-potential startup founders. CDL Super Session days will provide real networking opportunities and exposure to key players in the industry. We present the full questionnaire in Appendix.

Sample To construct the sample of code blocks, we leverage a dataset obtained from our partnered platform spanning observations from January 2018 to May 2022 (see Table 1 Panel B). Like our previous dataset, this dataset contains the subjective ratings and objective measure of coding quality. From this sample, we use first names to identify gender using predictions from genderize.io. This leaves us with 352,317 session-participant pairs, and 62,875 unique participants. Of these, 21.20 percent are probabilistically identified as female, and 78.80 percent as male. We restrict to three main languages: C++, Java and Python. A novel feature of our dataset is that we can link this information to the code blocks written by each participant in each session. An example of such a code block is shown in Figure 1 Panel B, together with the question (Panel A) and the unit tests (Panel C). Our final sample will be stratified by gender and coding performance.

Stratification To build our final sample of codes, we stratify both on gender (male/female) and on performance (100 percent of the unit tests are correct, versus less than 100 percent). We also limit ourselves to distinctively white names. We finally keep a random sample of codes.

Dimensions of the code quality We pre-register several observable dimensions: the length of the codes (number of lines), the duration to run, the number of comments. We will aim to explore whether these dimensions will explain any gender gaps we see even when gender is unobserved by the evaluator.

Randomization Let N be the number of evaluators and P the number of problems by evaluator. As mentioned before, our sample of code blocks is stratified by gender and performance, such that $\frac{P}{2}$ code blocks are written by women, among which $\frac{P}{4}$ are high-score codes according to the platform objective device. Each evaluator i is assigned a set of P problems in a random order. We use a within-subject design. We define $R_j = 0$ for a blind problem j (if the gender of the coder is not revealed), $R_j = 1$ for a non-blind problem j (if the gender of the coder is revealed). For each evaluator i, the gender of the coder will be revealed for half of the problems. To account for potential priming effect, we plan to randomize whether the gender of the coder is revealed in the first or in the second half of the study:

1. For half of evaluators, problems will be blind, then non-blind.

$$\forall i = 1, ..., \frac{N}{2} \begin{cases} \text{for } j = 1, ..., \frac{P}{2} & , R_{ij} = 0 \\ \text{for } j = \frac{P}{2}, ..., P & , R_{ij} = 1 \end{cases}$$

2. For the other half, problems will be non-blind, then blind.

$$\forall i = \frac{N}{2}, ..., N \begin{cases} \text{for } j = 1, ..., \frac{P}{2} & , R_{ij} = 1, \\ \text{for } j = \frac{P}{2}, ..., P & , R_{ij} = 0 \end{cases}$$

Testing the salience of the main treatment In the piloting phase of the experiment, we will ask participants to predict the gender of the coder after completing their evaluation. In theory, they should be 100% accurate in the non-blind treatment and at least 50% accurate in the blind treatment. If we find less than 100% accuracy in the non-blind condition, this would give us a sense of the share of inattentive evaluators, in which case we will re-evaluate the salience of gender in our treatment condition.

Measure of Priors To measure participants, we exposed them to three different vignettes before the perform their evaluation tasks. We ask them to predict the potential performance of three different hypothetical coders. We cross-randomize the first name (alternating gender) and the skill level for each vignette. The vignette are constructed as follows:

52% of the codes you will potentially see resulted in a perfect score and passed all the unit tests. We ask your opinion about the potential performance of different hypothetical coders. If your guess is within 5% of the truth, we will send you an additional reward!

"[First Name] holds [Skills]. According to you, what is the percent chance that [First Name]'s code passed all the unit tests?"

Skills	First names
a M.Sc in computer science and has 2 years of work experience	Katie/Tom
a Ph.D. in mathematics and has no industry experience	Alexa/Mickael
a B.Sc. degree in computer science	Corinne/Matt

Hypotheses Tested

Primary

- H1: Code blocks are evaluated differently if the gender of the coder is known.
- H2: Code blocks written by women are evaluated differently when we reveal the gender of the coder, with the gender gap increasing.
- H3: Individual gender bias varies significantly across evaluators.

Secondary

- H4: The gender identity of the evaluator affects their bias.

- H5: The difficulty of a given coding problem affects the evaluator′s bias.
- H6: The level of the coder′s performance affects the degree of bias.
- H7: Prior bias as assessed by the vignettes correlates with the evaluator′s bias in ratings.
- H8: The characteristics of a given coding problem affects the evaluator′s bias.
- H9: The race of the coder affects the degree of gender bias.

Because we are testing multiple hypotheses, we will use techniques that limit the false discovery rate such as correcting the p-values following the standard approach (e.g., Benjamini et al. 2006; Anderson 2008).

Outcomes

Our primary outcome is evaluators' subjective ratings of the code quality. For each block of code, respondents will be asked to rate problems on a scale from 1 to 4. For all primary hypotheses, we will use these responses as our main dependent variable.

We also have a secondary outcome: evaluators' prediction of the candidate's score from the automated evaluation tool. This is a continuous variable from [0,1]. A third outcome variable is evaluators' prediction of the candidate hireability score. This is measured on a Likert-scale from 1 to 4, and allows us to draw a more direct link between our findings and hiring outcomes (in addition to using the Revelio data).

Additionally, we will measure how much time respondents spend on each question to measure fatigue and inattention, and how this varies over time. For example, if discrimination is significantly larger in the latter half of the code blocks, that would provide suggestive evidence that implicit bias plays a role in our findings.

Econometric Specifications

To test H1, we will use the following specification:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 \times T_{ij} + \beta_2 \times T_{ij} \times NBB_i + \beta_3 \times NBB_i \\ &+ \sum_{j=1}^{P} \gamma_j \left(order \times \mathbb{1}_j \right) + \pi_{p(j)} + \delta_i + \epsilon_{ij} \end{aligned} \tag{1}$$

where Y_{ij} is a discrete variable from 1 to 4 which captures the ratings of evaluator i for code block j; T_{ij} is an indicator for whether gender is revealed to the evaluator; NBB_i is an indicator for the randomly assigned "non-blind-blind" condition of the code that the evaluator sees; $\pi_{p(j)}$

are problem fixed-effects; and δ_i are evaluator fixed effects. In some specifications, we include controls. Since code blocks characteristics are randomly drawn, including these variables in the analysis should not affect our estimates but could increase precision. Standard errors will be clustered at the evaluator level.

In equation 1, the coefficient of interest in β_1, which measures the average differences in subjective ratings for code blocks where the gender of the coder is revealed or not, controlling for the order NBB_i. The coefficients γ_j capture the effect of the order in which the code was evaluated, to account for learning and fatigue.

To test H2, we will use the following specification, which is very similar to 1 but interacts the key variables with gender indicators:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 \times T_{ij} \times female\_coder_j + \beta_2 \times T_{ij} \times NBB_i + \beta_3 \times female\_coder_j \\ &+ \beta_4 \times NBB_i + \sum_{j=1}^{P} \gamma_j \left(order \times \mathbb{1}_j\right) + \pi_{p(j)} + \delta_i + \epsilon_{ij} \end{aligned} \tag{2}$$

The coefficient of interest is β_2, which measures the differential effect of revealing the gender of the coder on subjective ratings, depending on what that gender is.

To test H3, our strategy entails first estimating the difference between male and female average gaps between a non-blind and a blind evaluation, following the literature on teacher grading bias (Terrier, 2020; Lavy and Sand, 2018):

$$E[Y_i|NB] - E[Y_i|B] = \beta_0 + \beta_1 \times female\_coder_i + \beta_2 \times NBB_i + \pi_{p(j)} + \epsilon_i \tag{3}$$

Here, β_1 is the coefficient of interest. Gender bias is defined as the average gap between non-blind and blind ratings for female-written codes, minus this same gap for male-written code. Because a gender bias is estimated for each evaluator, the number of observations for each estimation is limited. To address the sampling error, we plan to adapt the framework by Kline and Walters (2021). The authors use empirical Bayes (EB) analysis and data from correspondence studies to identify higher moments of the distribution of job-level callback rates as a function of the number of resumes sent to each job, and propose shape-constrained estimators of these moments. The key behavioral restrictions of their framework can be applied to our Likert-scale measure, so we will use their approach to investigate heterogeneity in gender bias.

To test H4 to H9, we will use variant of Model (2) where treatment effect on gender bias is interacted with, respectively, the gender of the evaluator, the difficulty and characteristics of the code, the coder's performance, the evaluator's bias measured through their priors, and the race of the coder.

Robustness and Additional Data Quality Checks

We plan to use an ordered probit specification. This specification is more flexible than OLS, allowing the discrete steps between Likert-scale points to vary in size. The coefficients reflect the effect of each characteristic on a latent variable over the Likert-scale space, and cutpoints are estimated to determine the distance between categories. We will also test the robustness of the results by re-weighting these regressions using observable characteristics to match our initial sample of workers from the platform we partnered with. Finally, we will perform a series of standard data quality checks to ensure that the randomization was successful and that the experimental design was not compromised. For example, we will perform a balance test by regressing respondent characteristics on indicators for the main randomized treatments. We will also use double/debiased machine learning methods to increase precision.

Registration Timeline

IRB clearance was obtained from the University of Toronto (RIS Human Protocol Number 41662) and the University of Michigan (Human Protocol Number HUM00204184). On December 9, 2022, the investigators uploaded the first copy of the pre-analysis plan (PAP) that outlines the main hypotheses.

Tasks	Start Date	Duration
Piloting phase	November 15th, 2022	2 weeks
Staggered delivery of survey experiment, week-by-week	February 20th, 2023	2 months
Analysis of survey responses and write-up.	April 20th, 2023	6 months

On February 14, 2023, approximately two months before the experiment began, we made the following changes to the PAP:

- Updated the timeline to reflect minor delays in the roll-out
- Included H9 to test for heterogenous effects by the race of the coder;
- Added details about robustness checks.

References

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484).

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli, "Adaptive Linear Step-up Procedures that Control the False Discovery Rate," *Biometrika*, 2006, *93* (3), 491–507.

Kline, Patrick and Christopher Walters, "Reasonable Doubt: Experimental Detection of Job-Level Employment Discrimination," *Econometrica*, 2021, *89* (2), 765–792.

Lavy, Victor and Edith Sand, "On the Origins of Gender Gaps in Human Capital: Short- and Long-Term Consequences of Teachers' Biases," *Journal of Public Economics*, 2018, *167(C)*, 263–269.

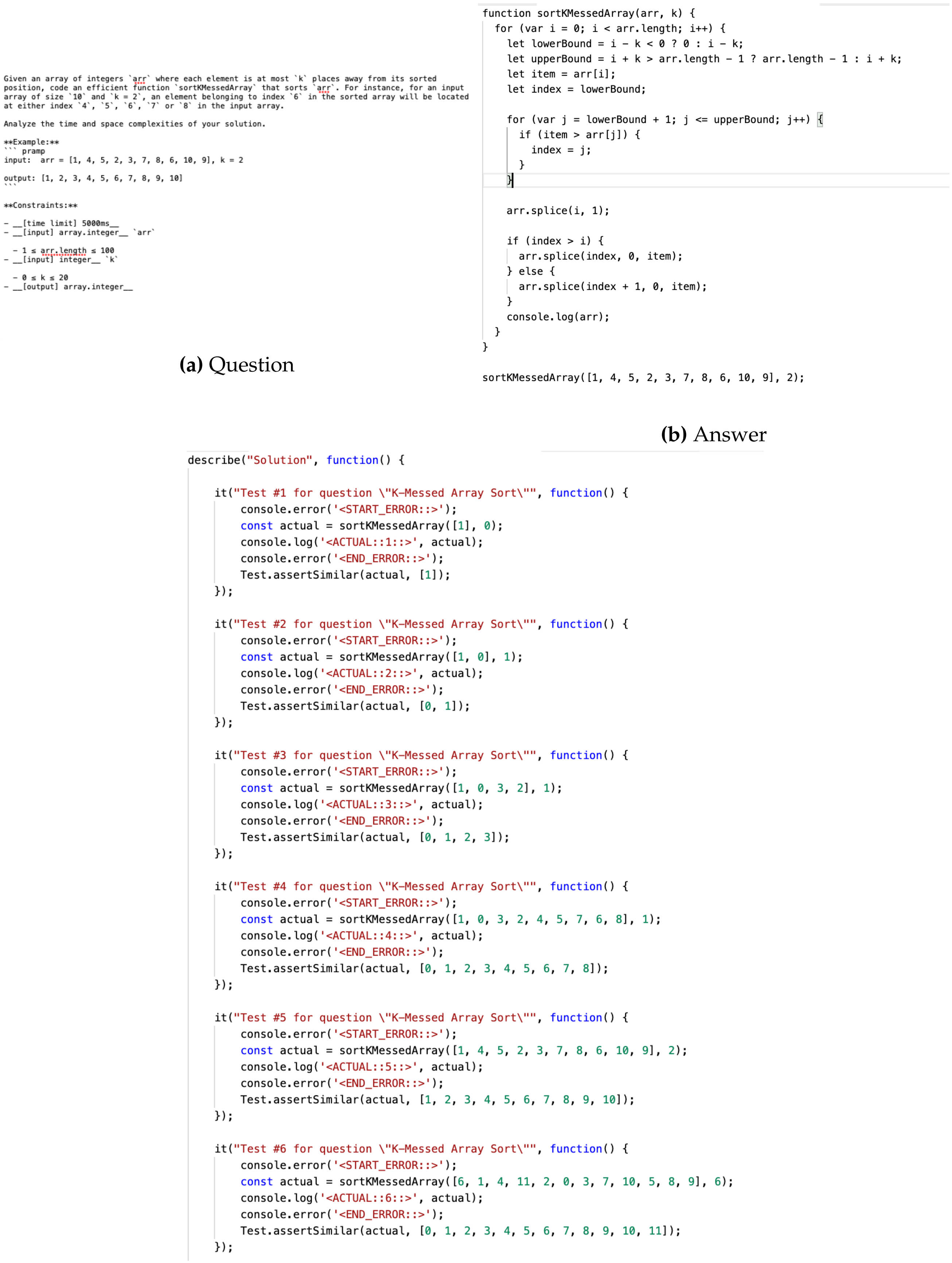
Terrier, Camille, "Boys Lag Behind: How Teachers' Gender Biases Affect Student Achievement," *Economics of Education Review*, 2020, *77.*

Appendix A: Tables and Figures

Table 1: Descriptive Statistics

January 2018-May 2022 — Follow-up Experiment	
Number of session-participant	352,317
Number of unique participants	62,875
Number of problems	39
Share of female participants	21.20%

Figure 1: Example of Code — K-Messed Array Sort



~~~
Given an array of integers `arr` where each element is at most `k` places away from its sorted
position, code an efficient function `sortKMessedArray` that sorts `arr`. For instance, for an input
array of size `10` and `k = 2`, an element belonging to index `6` in the sorted array will be located
at either index `4`, `5`, `6`, `7` or `8` in the input array.

Analyze the time and space complexities of your solution.

**Example:**
``` pramp
input:  arr = [1, 4, 5, 2, 3, 7, 8, 6, 10, 9], k = 2

output: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

**Constraints:**

- __[time limit] 5000ms__
- __[input] array.integer__ `arr`

  - 1 ≤ arr.length ≤ 100
- __[input] integer__ `k`

  - 0 ≤ k ≤ 20
- __[output] array.integer__
~~~

(a) Question

```
function sortKMessedArray(arr, k) {
  for (var i = 0; i < arr.length; i++) {
    let lowerBound = i - k < 0 ? 0 : i - k;
    let upperBound = i + k > arr.length - 1 ? arr.length - 1 : i + k;
    let item = arr[i];
    let index = lowerBound;

    for (var j = lowerBound + 1; j <= upperBound; j++) {
      if (item > arr[j]) {
        index = j;
      }
    }

    arr.splice(i, 1);

    if (index > i) {
      arr.splice(index, 0, item);
    } else {
      arr.splice(index + 1, 0, item);
    }
    console.log(arr);
  }
}

sortKMessedArray([1, 4, 5, 2, 3, 7, 8, 6, 10, 9], 2);
```

(b) Answer

```
describe("Solution", function() {

    it("Test #1 for question \"K-Messed Array Sort\"", function() {
        console.error('<START_ERROR::>');
        const actual = sortKMessedArray([1], 0);
        console.log('<ACTUAL::1::>', actual);
        console.error('<END_ERROR::>');
        Test.assertSimilar(actual, [1]);
    });

    it("Test #2 for question \"K-Messed Array Sort\"", function() {
        console.error('<START_ERROR::>');
        const actual = sortKMessedArray([1, 0], 1);
        console.log('<ACTUAL::2::>', actual);
        console.error('<END_ERROR::>');
        Test.assertSimilar(actual, [0, 1]);
    });

    it("Test #3 for question \"K-Messed Array Sort\"", function() {
        console.error('<START_ERROR::>');
        const actual = sortKMessedArray([1, 0, 3, 2], 1);
        console.log('<ACTUAL::3::>', actual);
        console.error('<END_ERROR::>');
        Test.assertSimilar(actual, [0, 1, 2, 3]);
    });

    it("Test #4 for question \"K-Messed Array Sort\"", function() {
        console.error('<START_ERROR::>');
        const actual = sortKMessedArray([1, 0, 3, 2, 4, 5, 7, 6, 8], 1);
        console.log('<ACTUAL::4::>', actual);
        console.error('<END_ERROR::>');
        Test.assertSimilar(actual, [0, 1, 2, 3, 4, 5, 6, 7, 8]);
    });

    it("Test #5 for question \"K-Messed Array Sort\"", function() {
        console.error('<START_ERROR::>');
        const actual = sortKMessedArray([1, 4, 5, 2, 3, 7, 8, 6, 10, 9], 2);
        console.log('<ACTUAL::5::>', actual);
        console.error('<END_ERROR::>');
        Test.assertSimilar(actual, [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]);
    });

    it("Test #6 for question \"K-Messed Array Sort\"", function() {
        console.error('<START_ERROR::>');
        const actual = sortKMessedArray([6, 1, 4, 11, 2, 0, 3, 7, 10, 5, 8, 9], 6);
        console.log('<ACTUAL::6::>', actual);
        console.error('<END_ERROR::>');
        Test.assertSimilar(actual, [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]);
    });
```

(c) Tests

Notes: This figure presents an example of code excerpt that will be used in the experiment. Panel A displays the question, Panel B the written code block, and Panel C the series of unit tests that generate the objective measure of performance.

Appendix B: Questionnaire

Informed Consent

Overview

You are being asked to take part in a research study being done by a group of researchers from the University of Michigan and the University of Toronto. This is a survey for academic research in social sciences. Your participation is invaluable for our research. If you choose to participate and to complete the survey, you will be financially compensated with a minimum of $50. As a participant, you will be asked to evaluate pieces of code written by others, and answer a short follow-up questionnaire. We expect that participation will take around 60 minutes. In each part, you will receive clear instructions and will be told how your decisions in that part will influence your earnings in the study. You will also have the opportunity to learn about your performance as evaluator.

Non-Deception Statement

This study does not deceive you by providing misleading or incorrect information. All our communications are truthful, but we may not always reveal all information. Specifically, there are different versions of this study. While you will be fully informed about the version of this study that you have been randomly assigned to, you will not be informed about different versions of this study that other participants are in.

Voluntary Participation, Privacy, and Point of Contact

Your participation is completely voluntary. You can agree to take part and later change your mind. Your decision will not be held against you. Note that the data you provide in this study will be anonymized prior to analysis. Your information will be kept entirely confidential and accessed only by the research team, and only as necessary to conduct the research. In the future, this non-identifiable data may be shared with other researchers or published. All information identifying you as a study participant will be destroyed upon the conclusion of the study. However, the anonymized information you provide may be maintained indefinitely.

The principal investigator of this study is Ashley C. Craig from University of Michigan. If you have any questions, concerns, or complaints, or think this research hurt you, talk to the research team at `ash@ashleycraig.com`. If you have questions about your rights as participants, you can contact the Research Oversight and Compliance Office — Human Research Ethics Program at ethics.review@utoronto.ca or 416-946-3273. You can also contact the University of Michigan IRB (Health Sciences and Behavioral Sciences) at 734-936-0933 or `irbhsbs@umich.edu`, quoting eResearch #HUM00204184.

The research study you are participating in may be reviewed for quality assurance to make sure that the required laws and guidelines are followed. If chosen, (a) representative(s) of the Human Research Ethics Program (HREP) may access study-related data and/or consent materials as part of the review. All information accessed by the HREP will be upheld to the same level of confidentiality that has been stated by the research team. If you would like a summary of the results of this research (once the study has been completed), please email `ash@ashleycraig.com`.

Compensation

You will receive $10 if you complete the survey and an additional $10 for each code segment you evaluate. Additionally, we will ask you to make a series of predictions. You will have the opportunity to gain $2 for each accurate prediction. Your total earnings will be distributed within one week after the completion of the survey. If you are interested, you can receive individualized feedback about the quality of your performance as an evaluator.

Based on their performance, the best ten evaluators win a $500 prize. The three best evaluators will also be invited to the Creative Destruction Lab 2023 Super Session in Toronto, which brings together world-class entrepreneurs, investors and scientists with high-potential startup founders. Organized in June 2023, the CDL Super Session days will give you with meaningful networking opportunities and exposure to key players in the industry. If there are ties in evaluation performance, the recipients of the prize and these invitations will be chosen randomly from among the set of evaluators with equal best accuracy scores. You may print a copy of this information for your records.

Yes, I would like to voluntarily participate in this experiment.

I am interested in receiving individualized feedback on my performance as an evaluator.

- Yes
- No

For the purposes of payment and the $500 cash prize, and to be considered for an invitation to the Creative Destruction Lab, please type your email below. We will not use your email for any purposes other than the provision of these rewards.

[Type here]

Please make sure you are willing and ready to sit through this study uninterruptedly and undistractedly before starting it. We ask you to please focus on the tasks of this study and thank you for your cooperation.

General Roadmap

This study consists of 4 evaluation tasks, followed by a few questions. The evaluation parts will ask you to give a score from 1 to 4 for scripts, both of which are solutions to a given coding question. The coding question will be outlined before the script.

Attention Checks

Note that this experiment contains attention checks. These questions are there to ensure you are paying attention as you take this survey. The answers to those attention check questions will not be ambiguous, will not be a trick question, and will not be timed. If you answer an attention check incorrectly or not within the provided time, you may be dismissed without pay.

Here is your first attention check. In the space below, please spell the word "human" backwards. Please use all lowercase letters and insert no space between the letters.

[Type here]

1. What best describes your present situation regarding your education?
   - I am currently a student
   - I have completed at least one degree
   - I was previously enrolled in a degree program but did not complete it
2. What is your highest level of education (including enrolled)?
   - High School diploma or GED
   - Some college, but no degree
   - Associates or technical degree
   - Bachelor's degree
   - MA, MSc or MEng
   - PhD
   - Prefer not to say
3. What is or are the area(s) of your highest degree? (multiple answers are allowed)
   - Computer Science
   - Computer Engineering
   - Mathematics
   - Information Systems / M.I.S.
   - Statistics
   - Other Exact Sciences Degree (e.g. physics, chemistry, astronomy)
   - Other Technology Related Degree
   - None
   - Other
4. What is the institution where you received or will receive your highest degree?

   [Drop down menu]

5. How would you describe your knowledge of these programming languages? Basic-Intermediate-Advanced
   - Python
   - Java
   - C++
6. During this study, you will be asked to evaluate a series of human written code blocks. Please select the coding language you are most proficient in.
   - Python

- C++
- Java

Before you start, we want to ask you a series of quick questions. The code excerpts were automatically subjected to a series of unit tests. These determined whether the code ran, and produced correct answers in pre-defined test cases.

Overall, 52% of the code blocks you will potentially see resulted in a perfect score and passed all the unit tests. We ask your opinion about the potential performance of different hypothetical coders. If your guess is within 5% of the truth for coders like those described, you will receive an additional reward!

- Katie/Tom holds a M.Sc in computer science and has 2 years of work experience. According to you, what is the percent chance that Katie's code passed all the unit tests?
- Alexa/Michael holds a Ph.D. in mathematics and has no industry experience. According to you, what is the percent chance that Alexa's code passed all the unit tests?
- Corinne/Matt holds a B.Sc. degree in computer science. According to you, what is the percent chance that Matt's code passed all the unit tests?

BEGINNING OF TASK

We are now going to ask you to evaluate a series of codes. These codes were written by actual software developers. We will provide you with the initial question and their written answers.

For each piece of code, we ask you to give your personal opinion about the quality of code, by providing a rating between 1 (lowest) and 4 (highest). At the end of all code evaluation, we will ask you to explain how you decided on your rating. You will gain a $10 additional bonus for each code you evaluate.

Additionally, we will ask you to make a series of predictions. You will have the opportunity to gain $2 for each accurate prediction.

Code Block 1

1. How would you rate the quality of the code (1 lowest, 4 highest)?
   - 1
   - 2
   - 3
   - 4
2. Can you let us know why you gave this score to the code ?

Text Box

3. A series of unit tests were used to evaluate this code. How many out of 10 unit tests do you think were passed? If your guess is within 5 percentage points of the truth, you will gain $2 and will increase your chances of participating to the Creative Destruction Lab Meeting and winning the cash prize.

- Drop Down menu

4. How confident are you about this prediction?
    - Not confident at all
    - Not confident
    - Somewhat confident
    - Confident
    - Very confident
5. This coder received a hireability score from another human evaluator based on this coding performance. We ask you to guess this hireability score, on a scale from 1 to 4. If your guess the correct score, you will gain $2 and will increase your chances of participating to the Creative Destruction Lab Meeting and winning the cash prize. How would you predict the hireability score of the code (1 lowest, 4 highest)?
    - 1
    - 2
    - 3
    - 4
6. How confident are you about this prediction?
    - Not confident at all
    - Not confident
    - Somewhat confident
    - Confident
    - Very confident

    According to you, what is the percent chance that the candidate was later invited for an interview for a role involving coding?
    - Cursor between 0 and 100

People often consult internet sites to learn about employment opportunities in tech. We want to know which sites you use. We also want to know if you are paying attention, so please select Glassdoor and Crunchbase regardless of which sites you use. When looking for employment opportunities, which is the one website you would visit first? (Please only choose one).

- LinkedIn
- Hired
- Glassdoor
- Crunchbase
- ZipRecruiter
- TripleByte
- Underdog
- Angel

Code 2 to 4 — Repeat

FOR PILOT ONLY What is your prediction of the percent chance that the last candidate was a woman?

- Cursor between 0 and 100

Follow-up questions

1. In which country do you currently reside?
    - Canada
    - USA
    - Other (choose)
2. How do you describe yourself?
    - Male
    - Female
    - Non-Binary / third gender
    - Prefer to self-describe: (type)
    - Prefer not to say
3. What is your year of birth?
    - Drop down menu
4. What best describes your employment status of the last three months?
    - Working full-time
    - Working part-time
    - Unemployed and looking for work
    - A homemaker or stay-at-home parent
    - Student
    - Retired
    - Other
5. How many year of working experience do you have?
    - Drop down menu
6. On a scale of 1-4 how prepared do you believe you are able to evaluate others' code?
    - 1
    - 2
    - 3
    - 4

1. In the box below, explain how you made your decisions today. Please answer in one or more full sentences.
    - Text Box

2. If you had to guess, what do you think was this study about? Please answer in one or more full sentences.
   - Text Box
3. Do you have any comments or feedback related to this study? (optional)
   - Text Box
4. Was there anything confusing about this study? (optional)
   - Text Box

Congratulations, you completed the main portion of the experiment! Once you have completed the questionnaire, you will reach the end of the experiment and learn about your total payment.

END of Questionnaire