# Pre-Analysis Plan:
# Racial Discrimination in Seeking Advice

**Vojtěch Bartoš, Ulrich Glogowsky, Johannes Rincke**

May 11, 2020

## Contents

# 1 Introduction

## 1.1 Abstract

We design an online experiment to study racial discrimination in seeking advice. In round 1 of the experimental design, subjects face a real effort task that is difficult to solve without prior advice from an expert. We offer subjects the option to watch a tutorial before working on the task. The main treatment variation is the race of the advisor (black vs. white), signalled by the skin color of a hand appearing at the beginning of the tutorial. We vary the skin color of a given hand model using video post-production techniques. This allows us to keep all features of the hand other than skin color constant between treatment arms. In round 1, we analyze how subjects' willingness to pay for advice depends on the race of the advisor, and how the race of the advisor affects advice utilization. In round 2, subjects watch another tutorial containing advice about a different strategy to solve the real effort task. To elicit preferences, we let subjects choose between two different advisors. Using an information treatment stating that the content of both tutorials is identical, we identify the extent of taste-based discrimination in seeking advice.

## 1.2 Motivation

Seeking the advice of others is a fundamental ingredient to solving complex tasks and decision problems. People seek advice from colleagues, coaches, and consultants on how to respond to professional challenges, from brokers, real estate agents, and physicians on how to manage problems relating to their wealth and health, and from friends, neighbors, and family on how to deal with a countless number of daily life issues. Advice seeking is also crucial in education, where students learn through the advice of teachers.

In many contexts, advice has been found to improve decisions [Schotter, 2003; Celen *et al.*, 2010; Brandts *et al.*, 2015] and to enhance performance when working on complex tasks [Blau *et al.*, 2010; Bettinger and Baker, 2014]. However, recent research has identified important social barriers that may often inhibit the beneficial effects of advice. Specifically, it has been shown that the willingness to seek and the receptiveness to advice depends on the perceived social distance between advice provider and advice seeker.[1] A special focus has been on racial congruence. For instance, black patients randomly assigned to black medical doctors are more likely to take up preventive services compared to when they are matched with a white doctor [Alsan *et al.*, 2019]. Similarly, several studies have

[1] It has been hypothesized that by seeking advice, one signals (to other people or the self) incompetence and dependence on others [Lee, 1997]. The social cost of seeking advice may increase with the social distance to the advice provider. The role of individual characteristics for seeking advice have also been studied. See, for instance, Heikensten and Isaksson [2018] on gender and Kramer [2016] on self-confidence.

documented that the racial congruence of students and teachers improves student learning [Dee, 2004; Egalite *et al.*, 2015; Lusher *et al.*, 2016]. The importance of congruence between advice provider and advice seeker may also be part of the explanation why members of minorities rarely make it to advisory positions. In the U.S., for example, the share of African-Americans among financial advisors, medical doctors, and teachers is less than 4, 6 and 7 percent, respectively, and the share of African-Americans among news anchors is less than 6 percent. Hence, African-Americans are strongly under-represented in many advisory positions relative to their U.S. population share of around 14 percent.[2] However, it is not yet understood why homogenous advisor-advisee pairs lead to advice being utilized more effectively. Potential channels include within-group communication being more effective, advisors of the same race functioning as role models, or in-group favoritism.

Against this backdrop, this research project deals with racial discrimination in advice-seeking. We focus on three main questions. First, given the importance of advice for task performance and decision making, we study to what extent advice-seeking is inhibited by racial discrimination. For that purpose, we analyze how subjects' willingness to pay (WTP) for advice depends on the race of the advisor. Second, we study how racial discrimination affects advice utilization. We do this by investigating how performance in a task for which advice is given depends on the race of the advisor. Importantly, we exogenously vary the advisor's race to avoid self-selection of advisees into specific advisor-advisee pairs. Third, we study the channels through which the race of the advisor affects advice seeking by separating statistical from taste-based discrimination.

To answer the aforementioned questions, we implement an online experiment using subjects recruited via Amazon Mechanical Turk. The experiment has two rounds. In round 1, subjects first watch a short video trailer. The trailer provides basic information on the characteristics of a real effort task, a sliding tile puzzle. In the trailer, a hand of the advisor explaining the sliding tile puzzle shows at several instances. This feature allows us to randomly vary the skin color of the advisor between black and white. After watching the trailer, subjects state their WTP to watch the full tutorial on how to solve the puzzle. Our design makes sure that almost all subjects watch the full tutorial before working on the puzzle for a fixed time. This design feature enables us to avoid selection effects when studying how the race of the advisor affects subjects' behavior and performance in the real effort task.

Round 2 serves to identify the channels. Participants watch another trailer that is presented by a new advisor and presents a faster way to solve the sliding puzzle. As in round 1, we randomly vary the skin color of the new advisor. After watching the trailer, subjects state their preference over whether the advisor from the first-round or the new advisor would de-

[2]Figures on the share of African Americans and other minority groups in the respective occupations are taken from American Society of News Editors [2018], Center for Financial Planning [2018], Association of American Medical Colleges [2014], and the U.S. Census Bureau.

liver the full tutorial, together with the WTP to avoid the less preferred advisor. To identify the extent of taste-based discrimination, we introduce an information treatment. Before the elicitation of preferences over advisors, a randomly selected subset of subjects are informed that the content of the material presented by the two advisors is identical. Providing this information, any remaining differences in the stated preference over advisors of different race can be attributed to differences in subjects' tastes. Importantly, the experimental design offers a plausible deniability of participant's preferences for an advisor of certain skin color, as the choice between advisors is framed as the choice between the first-round and second-round advisors.

A novel feature of our experimental design lies in the fact that we use video post-production techniques to vary the skin color of the advisor. Specifically, we employ Hispanic models when shooting the videos with the hand sequences. Using post-production techniques, we then produce different videos by changing the skin color of a given model to either black or white. This ensures that all other features of the hands are constant between treatment groups and allows for unconfounded causal inference about the effect of skin color on subjects' behavior in the experiment.

Besides offering one of the first studies on discrimination in advice seeking, we also add to the literature by introducing an experimental design that extends and refines the so-called correspondence methodology to study discrimination. As discussed in the recent surveys by Bertrand and Duflo [2017] and Neumark [2018], a big advantage of the correspondence method lies in the fact that it allows for the study of real market interactions.[3] On the downside, the design of most correspondence studies following the example of Bertrand and Mullainathan [2004] suffers from the outcome variables typically being very coarse. In most field experiments, the main outcome variable being studied is the call-back rate. Other, and potentially more informative, measures of discrimination that would require a prolonged interaction between fictitious applicants and the subjects studied are rarely available. Also, correspondence studies rely on signalling the applicant's race, often by using black-sounding and white-sounding names on resumés. One typically cannot rule out the possibility that the subjects associate other characteristics with black-sounding names rather than race per se. Finally, even including all information typically provided on resumés does not guarantee that all productivity-related characteristics of applicants are held constant.

Our experimental design is a refinement of the correspondence methodology that improves the approach in all three dimensions discussed above. First, our experiment generates a very rich set of outcomes, including the WTP for advice, survey measures for the perceived quality and a ranking of advisors of different race, various measures for advice

[3]The correspondence methodology shares this advantage with the audit methodology as the second main experimental paradigm to study discrimination [Ayres and Siegelman, 1995; Neumark *et al.*, 1996].

utilization and performance on the task once advice has been received, and measures for subjects' attention while the advice is given. We also link our experimental data to the outcomes of an Implicit Association Test (IAT) subjects are invited to take after participating in the main experiment. Based on all these outcomes, our study provides a comprehensive analysis of how racial discrimination affects how people seek and make use of advice. Second, showing a hand of the advisor in a video sequence is a very direct signal of the advisor's race. Moreover, the technique we used to produce the videos ensures that all other features of the hand are held constant, avoiding possibly confounding differential perceptions that cannot be excluded in many other settings. Third, the videos showing either a black or a white hand have exactly the same content. This means that, using the info treatment that informs subjects about the content-wise equivalence of the videos in round 2, we can identify the extent of taste-based discrimination.

Several aspects of our work, including the focus on the acquisition and utilization of new information, link our paper to the work of Bartos *et al.* [2016] on attention discrimination. In different field experiments, Bartos *et al.* [2016] use online resumés and personal websites of fictitious applicants to track the information acquisition behavior of employers and landlords, respectively. Besides providing evidence of discrimination against minority applicants in both labor and housing markets, they demonstrate that the allocation of attention to minority applicants depends on the specific market environment. Another aspect relating our work to the recent literature is that we pin down the cost of not using optimally the information contained in advice. In that respect, the previous contribution closest to ours is Hedegaard and Tyran [2018], who show that ethnic discrimination in the workplace is highly responsive to the opportunity cost of choosing a less productive co-worker. One advantage of our experimental design over Hedegaard and Tyran [2018] is that we can rule out complementarities in production between the advisor and the worker. In terms of how the race signal is transmitted and the quality of the outcomes studied, our work links to Doleac and Stein [2013], who use pictures showing a hand of a fictitious seller to distinguish between black and white sellers and track the sales of iPods through local online markets all the way to completion. While Doleac and Stein [2013] use pictures of different hands that might signal characteristics of the seller other than race, we manipulate the skin color of the hands using video post-production techniques and thereby make sure that all other features of the hands are held constant across treatments.

Effective policies against racial imbalances require a precise understanding of the underlying channels. So far, few papers in the literature on race and advice seeking have tried to identify the channels through which the racial congruence (or difference) in advisor-advisee pairs affects behaviors and outcomes. For instance, Alsan *et al.* [2019] cannot precisely pin down why black men are more likely to take up preventive treatment if interacting with a

doctor of the same race. The results, however, point to the driver of this racial differential being better patient-doctor communication during the encounter rather than discrimination. While (part of) the effect might be driven by differences in the behavior of doctors, our design shuts down all possible supply-side effects. In the context of education, the study by Baker *et al.* [2018] stands out for its clear identification of discrimination as the relevant channel. The authors exogenously assign race and gender (signalled through typical names) to fictitious participants in online courses and demonstrate that course instructors are substantially more likely to respond to requests from white males. Focusing on racial imbalance in giving advice provides a perspective that is complementary to our analysis of discrimination in seeking advice.

### 1.3 Research Questions

The overarching research question is to what extent advice-seeking is inhibited by racial discrimination. Specifically, we investigate the following primary research questions:

1. How does the advisor's race impact individuals' advice-seeking and advice-utilization behavior?

2. How does the advisor's race affect individuals' performance, conditional on having received advice?

We also consider the following secondary research question:

3. Through which channels does the advisor's race impact advice seeking, i.e., what part of the differential in seeking advice from advisors of different race can be attributed to statistical discrimination, and what part to taste-based discrimination?

## 2 Experimental Design

We plan to recruit participants on Amazon Mechanical Turk, or MTurk, to study (a) how the advisor's race affects participant's advice seeking behavior, (b) how the advisor's race affects advice utilization, and (c) through which channels the race of the advisor affects advice seeking. In the following, we briefly sketch the most important aspects of our design before we lay out the details in more depth.

**Task** We design an easy to understand, yet comprehensive real effort task, a sliding tile puzzle. The task is designed such that participants benefit from getting advice. Furthermore, it allows us to track participant performance and to measure whether advice received was

used. Also, note that the incentive structure in the experiment is set such that participants are rewarded for every task solved within a limited time period. We also manipulate the size of incentives by manipulating the amount of money participants receive for each correctly solved task. Section 2.1 describes the task in detail.

**Manipulation of Advisor's Race** We manipulate the advisor's race by presenting the advice in an online video with an advisor explaining a strategy how to solve the task before the participants are asked to perform. In the video, a hand of the advisor appears at several instances. The advisor's skin color is either white or black. We use video post-production techniques that allow us to manipulate the skin color of a Hispanic model. This ensures that the skin color is truly the only difference between advisors. Put differently, all other dimensions remain constant. This allows us to make unconfounded causal inference about the effect of skin color on individual behavior. Section 2.2 describes treatment manipulation and details of the randomization procedure in more detail.

**Measurement of Advice Seeking Behavior** To measure the extent to which the advisor's race affects participant's advice seeking, we measure participant's willingness to pay for advice. Initially, participants only see a part of the video, a trailer. The trailer does not reveal any information beyond stating the objective of the task. It also refers to the remaining video that offers a strategy to solve the task. The participant then has a chance to pay for watching the remaining part of the video using a Becker-DeGroot-Marchak (BDM) willingness to pay elicitation method [Becker *et al.*, 1964; Becker and DeGroot, 1974]. Section 2.3 describes this mechanism in detail.

**Identification of Channels** Our design allows us to study the channels through which potential race-specific behavior occurs. We focus on two leading theories of discrimination: taste-based discrimination [Becker *et al.*, 1957] and statistical discrimination [Phelps, 1972; Arrow, 1973]. To separate both forms of discrimination, we introduce a round 2 and an information treatment. We then study if individuals choose white over black advisors (when they have a choice).

More specifically, round 2 is structures as follows: Participants see another trailer that presents a faster strategy to solve the puzzle. We also inform participants that the strategy is faster.[4] This trailer is presented by a new advisor and it is of the same video and voice quality as the trailer in round 1. As in round 1, the advisor's hand shown is either black or white. Our design allows for all combinations of black and white advisors across the two rounds.

[4]Appendix D explains how we ensure that the second strategy is, indeed, faster.

Individuals can then choose one out of two advisors. This allows us to identify whether individuals prefer black or white advisors. Specifically, we ask participants which advisor they prefer: the first-round advisor or the second-round advisor (i.e., the advisor who presented the second-round trailer). To elicit this choice in an incentive-compatible way, we use a three-stage procedure. First, we ask participants to rank the two advisors, allowing for indifference. Second, we introduce the following lottery. With probability 95 percent the new advisor from the second-round trailer will deliver the remaining video, and with probability 5 percent participants enter an additional lottery. The remaining video will be delivered by the preferred advisor with a probability of 70 percent, and otherwise by the non-preferred one.[5] Third, we elicit participants' willingness to pay for the preferred advisor in the case that the non-preferred advisor is drawn. For elicitation, we use the BDM mechanism as in round 1. This procedure allows us to study counterfactual behavior of how participants behave when being exposed to either their preferred or their non-preferred advisor. In the real effort task, we measure the subjects' performance and whether they follow a strategy described in either of the two rounds.

To identify the extent of taste-based discrimination, we also introduce an information treatment. Before individuals rank the advisors in the second round, a randomly selected half of participants are informed that the two advisors use *exactly the same script* when recording the video (emphasis added here, not in the instructions). We also inform them that the contents of the two tutorials are identical, including the layout of the puzzle, the steps taken to solve it, and the wording used to explain the strategy. As a result, participants' beliefs about the quality of the second-round video should be independent of whether it is presented by the first- and second-round advisor. Importantly, beliefs should be also independent of the advisors' race. We then say that there is taste-based discrimination if participants who, for example, can choose between a black first-round and a white second-round advisor more likely prefer the second-round advisor compared to participants who can choose between two white advisors. Put differently, we exploit the fact that the participants may be randomly exposed to a black and or a white advisor either in round 1 or round 2. Also, note that the experimental design offers a plausible deniability of participant's preferences for an advisor of certain skin color. On an individual level, we can never say whether the participant prefers an advisor because of the advisor's skin color or because of the (non-)familiarity with the advisor.

Section 2.4 describes the procedures and additional measures we collect, and Section 2.5 discusses the sampling procedure.

[5] If participants are indifferent between the advisors, we classify one of the advisors as a preferred one with equal chance. The procedure then follows as if a preferred advisor was initially selected. Instructions clearly mention this.

**Measurement of Beliefs** As previously explained, beliefs are important to discriminate between statistical and taste-based discrimination. In round 1, we elicit how many puzzles participants expect to solve in 5 minutes after having watched the full tutorial. Equivalently, in round 2, we elicit how many puzzles participants expect to solve in 5 minutes after having watched a tutorial presented by the first-period advisor or the second-period advisor. Section 4.3 describes the details.

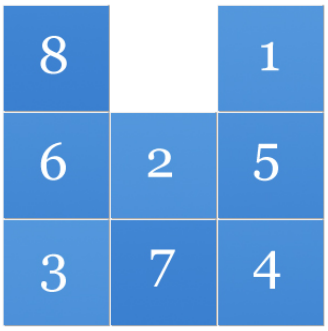
## 2.1 Task

The real effort task we use is a sliding tile puzzle (see Figure 1 for an example). The task consists of a three-by-three grid, with eight numbered tiles and one space left empty. The space allows the tiles to be moved around. The puzzle appears with the tiles placed in an unordered way. The goal of the game is to rearrange the tiles into numerical order by sliding them successively into the empty space. The puzzle is solved once the correct order has been achieved. After that, the tiles are reshuffled again. The participant earns a piece rate for each puzzle solved. In each round, participants have five minutes to solve as many puzzles as possible.

The game has several important features. First, it is simple in appearance, and the goal is easy to understand. Second, by setting equal starting positions of tiles in the puzzle across all participants, we have perfect control over the difficulty of the task. Third, there are various easy-to-learn strategies that differ in complexity. In the tutorials, the advisors present such strategies. Participants benefit from following the strategies presented, as for most individuals a sliding tile puzzle is quite difficult to solve without any guidance.[6] Hence, the choice not to watch a video is costly. The costliness further varies with the piece rate that we also manipulate in our design. Fourth, the puzzle has a unique solution and hence a simple count of puzzles solved within a given time frame can be used as a measure of overall performance. Lastly, we can measure whether individuals follow a strategy that was presented by an advisor. From unique consecutive patterns in the data, we can infer if individuals used either of the presented strategies. In summary, we measure as outcomes each participant's performance, and whether the participant used one of the two strategies presented by the advisors (see Section 4.2).

[6] We conducted a pre test in which participants solved puzzles (a) after having watched the full tutorial and (b) without having watched the full tutorial. In ten minutes, participants who did watch the full tutorial were, on average, able to solve 7.4 (9.5) puzzles in the first (second) round. Without the full tutorial, they solved 3.4 (4.6) puzzles less in the first (second) round.

**Figure 1:** Sliding tile puzzle




## 2.2 Treatments

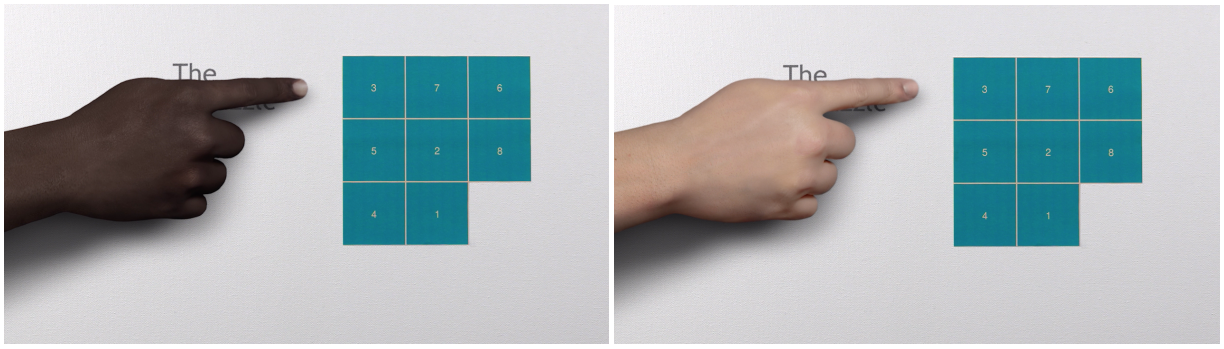
**Skin color** We create two types of videos, a *trailer* and a *main video*. The main video explains one of the two strategies that can be used to solve the slider tile puzzle intuitively and efficiently. The second-round main video explains a faster way of solving the task relative to the first-round main video.

We reveal the skin color of the advisor in the trailer videos. These videos are 30 second long clips that start with a close-up of a screen where the slider tile puzzle is presented. Several seconds after the start of the video, a hand of the advisor enters the screen and points at the slider, explaining that this is the task that the participant will be solving and also the puzzle's goal. The hand then also hovers over the puzzle when explaining how the task works (i.e., what moves are possible). At the end of the trailer the advisor explains that he would present a strategy how to solve the puzzle in the main video.

To produce the videos, we recruited two male actors of Hispanic origin, around 30 years old. Both actors followed strict instructions that prescribed their hand movement and we recorded a version of a trailer video for both rounds with both actors. A post-production company then manipulated the skin color using special video techniques to produce equivalent white and black hands. Importantly, all videos used in the experiment show hands with manipulated skin color (i.e., the hands' original skin color was made lighter for the videos featuring the 'white' hand and made darker for the videos featuring the 'black' hand. To remind participants of the treatment, the hands also appear shortly at the beginning of the main videos. See Appendix C for a detailed summary of how the videos were produced. Figure 2 presents a screenshot of the videos for one of the actors. Sub-Figure 2a shows a black skin color transformation, while Sub-Figure 2b shows a white skin color transformation.

As mentioned previously, out experiment consists of two rounds. In each round, the participants are exposed to a different hand. We randomly assign the two actors across rounds. When producing the videos, we recorded two different voices, one with a white US

**Figure 2:** Skin color manipulation using post-production video techniques



**(a)** Black hand **(b)** White hand

native, and one with a black US native.[7] The voices are randomly assigned to the videos across rounds. We allow for all skin-color and voice combinations across the two rounds. Table 1 summarizes the resulting 16 combinations. Importantly, in round 1, the combination of actor, skin color and voice is the same between trailer video and full tutorial. In round 2, however, subjects make choices that affect which of two advisors they have seen (first-round advisor and advisor from the second-round trailer video) presents the full tutorial (for details, see section 2.3). Therefore, Table 1 refers to the random assignment of actor, skin color and voice in the trailer videos.

Also, note several further features of our design. First, the allocation of hand types and voices is orthogonal to the skin-color treatment, the main variation of interest. Second, the participant neither faces the same actor with a different hand color nor two different actors with the same voice across both rounds. Third, we designed the study such that we can observe round 2 participants who shared a common round 1 history (say, an actor 2 with voice 1 and a black skin tone; highlighted in Table 1) but face a different skin color in round 2 (an actor 1 with voice 2 and with either black or white skin tone). This feature allows us to make causal claims about round 2 behavior.

**Piece rate** Orthogonally to changing advisor's skin color, we also randomly manipulate the piece rate for each completed puzzle. In the the low piece rate treatment, participants earn \$0.5 for each completed task. In the high piece rate treatment, the piece rate is \$1. This

[7] In a separate survey with 100 mTurkers, we testwhether the participants perceive the hands as "naturally" looking. Specifically, we present still frames to subjects and ask subjects 1) to describe the hands, 2) to guess the race of the respective person, and 3) to state whether or not they believe the skin color of the hands presented was manipulated ex post. To each subject, we present two different randomly selected hands. In contrast to the experiment, we draw from a set of hands not only containing the four hands used in the experiment, but also equivalent still frames of the two hands with original skin color. This allows us to also test whether subjects are more likely to perceive as manipulated the skin color of the hands used in the experiment relative to hands without any manipulation.

**Table 1:** Treatment allocation

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Round 1** | | | | | | | | | | | | | | | | |
| Black / White | B | | | | | | | | W | | | | | | | |
| Actor | 1 | | | | 2 | | | | 1 | | | | 2 | | | |
| Voice | 1 | | 2 | | 1 | | 2 | | 1 | | 2 | | 1 | | 2 | |
| **Round 2** | | | | | | | | | | | | | | | | |
| Black / White | B | W | B | W | B | W | B | W | B | W | B | W | B | W | B | W |
| Actor | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| Voice | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |

*Notes:* This table describes the full set of 16 possible combinations of skin colors, actors, and voices used across rounds 1 and 2. B stands for black, W stands for white. The numbers represent different actors and voices. The highlighted light grey cells show an example representing treatment combinations in which participants are exposed to an advisor of different skin color in round 2, while sharing the common round 1 treatment history (actor 2 with black skin color and voice 1).

allows us to manipulate the costs of discrimination and estimate how participants respond to it.

**Information** Orthogonally to the skin color and piece rate treatments, we randomly manipulate the information we provide to individuals in round 2, before they are asked to rank the advisors. In the information treatment—on top of instructions how to rank the advisors—we tell participants that "when recording the tutorials, both instructors followed the same script. Therefore, the contents of the two tutorials are identical, including the layout of the puzzle, the steps taken to solve it, and the wording used to explain the strategy." We implement this treatment to shut down any remaining differences in beliefs about the quality of the advice. This allows us to separate taste-based discrimination from statistical discrimination.

## 2.3 Willingness to Pay Mechanism

**Round 1** In round 1, we ask participants to state their willingness to pay for the video. The elicitation mechanism we use motivates participants to report their true willingness to pay, as misreporting leads to a utility loss. The mechanism works as follows. On top of the show up fee, we add an extra $1 to participants' endowments. The participants can use any amount of this additional money to pay for the full video ($x$). They can do so in increments of $0.01. To set the exact maximal price they are willing to pay, they use a slider bar on their screens.

The computer randomly draws a price $p$, ranging from $0.00 to $1.00. If $x \geq p$, the participant pays $p$ and watches the full video. If $x < p$, the participant pays nothing but does

not have access to the full video. Instead, the participant would watch an uninformative video showing fish swimming in the sea. The probability of drawing a price of 0 is set at 95 percent, and each other price in increments of one cent up to $1.00 has a remaining positive probability following a uniform distribution. The instructions say that each price is drawn with a positive probability. We do not inform participants about the true underlying probability distribution.

By comparing the willingness to pay between individuals in the black and white treatments, we obtain a measure for individuals' willingness to discriminate.

**Round 2** In round 2, after watching the trailer, participants are promised to see the full video for sure. This time, they are asked to choose between two advisors: the first-round advisor, referred to as the "*first instructor*", or the advisor they have just seen in the round-two trailer, referred to as the "*second instructor.*" The selection works as follows:[8] Participants are first asked to rank the two advisors or to indicate indifference. Participants then enter a lottery that determines which advisor is selected. With probability 95 percent—unknown to the participants who only know that the probability is positive—the second advisor will present the full tutorial (case 1). With the remaining probability of 5 percent, they enter an additional lottery (case 2). Here, the main video will be delivered by the preferred advisor with a probability of 70 percent—known to participants—and by the non-preferred advisor otherwise. Hence, the participants' ranking of the advisors matters in expectation.[9] Finally, participants are asked how much they would be willing to pay to get the preferred advisor in case the non-preferred advisor would be selected by the lottery in case 2. We elicit the willingness to pay using the same Becker-DeGroot-Marschak mechanism as in round 1.[10] In this round, however, we draw the price $p$ from a uniform distribution.

## 2.4 Experimental Procedures

A summary of the timeline is presented in Appendix A. The sequence of events in round 1 is as follows: After a participant enters our website, general instructions appear. Participants login using their MTurk worker ID and give informed consent. Next, participants fill out a short demographic survey used for stratification purposes. Conditional on their responses, participants are assigned to skin color, piece rate, and information treatments.

After completing the survey, participants are redirected to the next page with general instructions, and a timer is switched on. New pages load automatically when the time

[8] Participants are first informed about how the full elicitation mechanism works, and then make their choices.

[9] If participants are indifferent between both advisors, the advisors are assigned with equal probabilities.

[10] Toussaert [2018] uses a similar method to elicit willingness to pay for a commitment device.

for a previous page runs out. Therefore, all participants proceed at exactly the same pace regardless of their choices, keeping the opportunity cost of time fixed. Furthermore, as an attention check, we also measure whether the participant has the tab with the experiment open in his or her browser, and if not, when and for how long this is the case.

As part of the instructions, participants are informed about the general procedures, and the sequence of the experiment (see Appendix E for instructions). They learn that they will 1) see a trailer, 2) be asked for their willingness to pay for the full video that they then 3) either watch or not, and 4) solve the task. They also learn that a similar sequence of the four steps will be repeated in a second round. The instructions also explain how the payoff is calculated. Participants learn that only one randomly selected round will be payoff relevant.

The individual payoff is calculated as the $4 show-up fee (called reward), plus the bonus payment earned for solving the tasks at a given piece-rate ($0.5 or $1), plus the additional endowment of $1 for the willingness to pay procedure, minus the price drawn by the willingness to pay elicitation mechanism (if the stated willingness to pay is higher than the price drawn).[11, 12] Of course, only values for the payoff relevant round are considered.

After having read the instructions, participants see the first-round trailer. Then, we elicit participant's willingness to pay for the full tutorial. Conditional on the price drawn and the stated willingness to pay, participants are either redirected to the main video (vast majority) or to the uninformative video. After participants have stated their willingness to pay, a short survey elicits beliefs about the expected number of tasks solved when watching the full trailer or the entertainment video. Subsequently, participants have the possibility to evaluate the advisor (see Subsection 4.4.1).[13] In the next step, participants are redirected to the actual sliding tile puzzle task and work on it for five minutes. We record all the moves the participant makes. This allows us to classify the strategies used by the participant.

The entire sequence is repeated in round 2, with one main difference: Participants now choose between the two potential advisors using the willingness to pay method described in Section 2.3. Furthermore, the belief question after stating the willingness to pay now elicits beliefs about the expected number of tasks solved conditional on watching either of the two advisors.

At the very end, we administer another short survey on basic demographics, political party affiliation, general experience with the sliding tile puzzle, characteristics of both advisors, and own assessment of strategy use in the slider tile puzzle task across rounds (see

[11] An average HIT on MTurk requiring a minute of the user's time pays 5-10 cents, which corresponds to an hourly wage of $3-6. Because there seems to be a positive correlation between payment and data quality, we decided for paying above the average wage.

[12] Participants also learn that the advisor is not paid the money they give up using the willingness to pay mechanism. This shuts down a confounding channel of social preferences, such as altruism towards the advisor. This is relevant if participants form beliefs about the advisor's income or wealth.

[13] Participants who do not watch the full trailer have an option to reveal that they did not watch the trailer.

Subsection 4.5.1). In the last step, the payoff-relevant round is selected, participants are informed about the total amount earned, and the experiment concludes after participants copy a unique completion code that appears at the last page of the experimental website back to MTurk.

Several weeks after completing the HIT, participants are invited to take part in another HIT. In this HIT, participants complete a version of a race implicit association test [Greenwald *et al.*, 2003], a method widely used in social psychology. It assumes that the strength of individuals' associations between pairs of concepts correlates with the speed with which they can classify the concepts in a rapid categorization exercise. In our case, participants classify images of black and white faces with positive and negative words. As a measure of implicit bias, we use the D-score calculated as in Lane *et al.* [2007] (see Subsection 4.5.2).

## 2.5 Sampling

**Subject Pool:** The participants are recruited via MTurk. MTurk is an online web-based platform for recruiting and paying subjects to perform tasks that require human effort. As highlighted by Bohannon [2011], more and more social scientists exploit this online environment for conducting surveys or experiments. And according to an article in *The Economist*, MTurk is so popular amongst psychologists that "[it] is transforming the science of psychology."[14] Recently, the Pew Research Center also documented the increasing importance of MTurk as data source [Pew Research Center, 2016]. The center conducted a study highlighting that in one representative week in 2015, 36% of the unique requesters were either graduate students, professors, or other academic groups. That was somewhat more than the 31% for businesses. Thus, by now, many workers are familiar with participating in various academic studies.

**Sample Characteristics:** According to Stewart *et al.* [2015], the participant population is about 7,300 individuals. The following table shows the characteristics of the Kuziemko *et al.* [2015] sample. Because they also focus on the US and because we plan to use a similar sampling strategy, we expect to get a similar sample. The table also compares the summary statistics to a nationally representative sample of US adults contacted by a Columbia Broadcasting Company (CBS) poll in 2011 and the American Life Panel (ALP).

[14] See "The Roar of the Crowd" (The Economist, May 26, 2012). Also economists started to recruit subjects through MTurk. See Horton [2011], Horton *et al.* [2011], Kuziemko *et al.* [2015], and Oster *et al.* [2016] for prominent examples.

TABLE 1—SUMMARY STATISTICS AND COMPARISON TO OTHER POLLING AND ONLINE DATA

| | mTurk sample (1) | CBS election poll (2) | American Life Panel (3) |
|---|---|---|---|
| Male | 0.428 | 0.476 | 0.417 |
| Age | 35.41 | 48.99 | 48.94 |
| White (non-Hispanic) | 0.778 | 0.739 | 0.676 |
| Black | 0.0756 | 0.116 | 0.109 |
| Hispanic | 0.0444 | 0.0983 | 0.180 |
| Other racial/ethnic group | 0.0759 | 0.0209 | 0.0410 |
| Employed (full or part) | 0.465 | 0.587 | 0.557 |
| Unemployed | 0.123 | 0.104 | 0.103 |
| Married | 0.397 | 0.594 | 0.608 |
| Has college degree | 0.433 | 0.318 | 0.309 |
| Voted for Obama | 0.675 | 0.555 | 0.559 |
| Political views, conservatives (1) to liberals (3) | 2.176 | 1.586 | |
| Observations | 3,741 | 808 | 1,002 |

*Notes:* This table displays summary statistics from our mTurk omnibus surveys in column 1 along with (weighted) averages based on a 2011 CBS news survey in column 2 and RAND's online American Life Panel (ALP) in column 3. We are grateful to Ray Fisman for providing us with summary statistics from the ALP.

**Recruitment:** We recruit participants, called workers on MTurk, for a scientific study on e-learning. Figure B.1 in the Appendix shows the description on MTurk. While participants know that we conduct a study, they are not aware of discrimination in seeking advice being the true purpose of the study. This gives us a natural setting to study the effect of race.[15]

**Sampling Restrictions:** Our sample restrictions are as in Kuziemko *et al.* [2015]: First, our study is only accessible to workers who are US residents.[16] Second, to exclude robots, only workers with a completion rate of at least 90 percent were allowed to take the survey. Third, to improve the quality of our data, we focus on individuals who completed at least 100 HITs.[17,18] Fourth, we tell participants that payment is contingent on completing the study and providing a private key visible only at completion. In case we run out of participants, we will relax these sample restrictions.

[15]This distinguishes our work from standard correspondence experiments in the line of Bertrand and Mullainathan [2004]. Other studies that also exploit natural settings are Hoff and Pandey [2006], Mobius and Rosenblat [2006], and Rao [2019].

[16]Workers whose IP addresses are not consistent with our country location settings are prevented from participation.

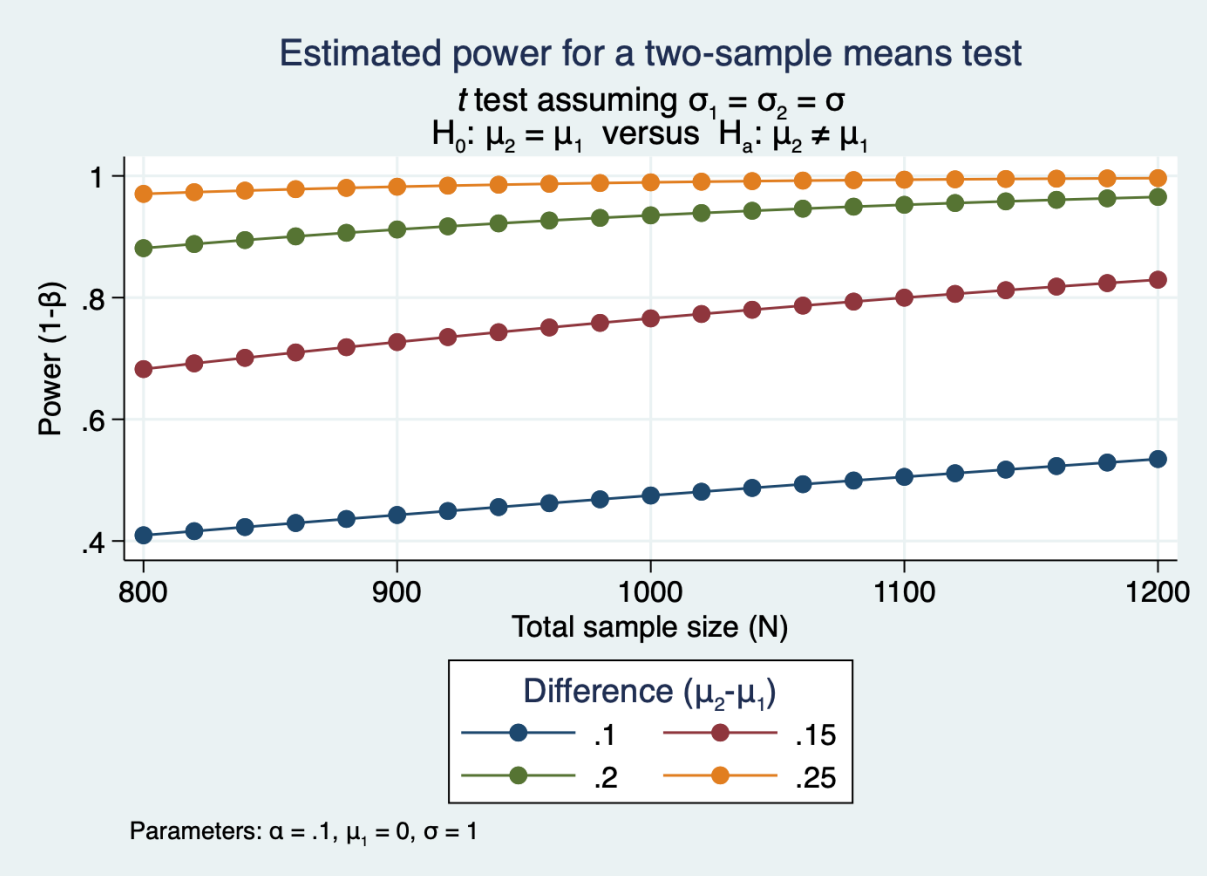
[17]To prevent workers to participate multiple times, we block each participant's IP addresses after participation. Further, we use the "hyperbatch" option of the company "CloudResearch" that allows us to launch our study in batches of 9 in parallel, as opposed to one after the other. The hyperbatch option automatically ensures that (a) individuals can participate only in one of the offered HITs and (b), at the same time, enables us to allow many participants to take your study at once.

[18]We exclude all participants who do not pass a simple comprehension check at the page presenting instructions for the experiment. This feature is aimed at excluding autonomous programs, or bots, that could contaminate our data.

**Stratification:** We stratify the treatment allocation by participants' race (black, white, other), education (no college degree, some college degree or higher), and state (South, other) following the US Census classification).

## 2.6 Power calculations

To detect an economically meaningful effect of 0.15 of a standard deviation at a 10 percent significance level with power of 80 percent, we would require an overall sample of 1100 individuals. Using the data from the pilot study with 32 individuals, this would correspond to detecting an effect of \$0.06 for the willingness to pay in round 1 ($WTP1$; mean=0.48, SD=0.37), an effect of \$0.04 for the willingness to pay in round 2 ($WTP2$; mean=0.20, SD=0.30), and an effect of 0.52 for the number of puzzles solved in round 1 ($NP$; mean=4.28, SD=3.49). The figure below presents power calculations for two-sample means comparison t-tests for a range of parameters for a variable following a standard normal distribution.




Since the other primary outcome we consider is an ordered categorical variable, we cannot use a simple means comparison. Instead, we follow Whitehead [1993] and show that given a sample of 1100 individuals, a 10 percent significance level, and power of 80 percent,

we would be able to detect a "reference improvement" of approximately 0.05. To illustrate what this means, consider our data from the pilot. The share of individuals choosing option 1 ("I prefer the first instructor to the second instructor") was 40.6 percent, the share of individuals choosing option 2 ("I prefer the second instructor to the first instructor") was 28.1 percent, while the remainder of the sample was indifferent (option 3). The proposed sample would allow us to detect a distribution of 47.5 percent choosing option 1, 22.8 percent choosing option 2, and the remainder of the sample being indifferent, and similar. I.e. an effect of a shift of approximately 6 percentage points in either direction.

# 3 Empirical Analysis

In this section, we specify how we will analyze our data once they are available. All the results will be reported in the paper or the Appendix. To further explore the relationships that emerge from our registered regressions, we anticipate running additional specifications. In these instances, we will follow Casey *et al.* [2012] and label the associated results as non-registered in the paper.

## 3.1 Randomization Checks and Other Design Checks

### 3.1.1 Test for Balance

We will test for treatment balance along all the available strata variables (see Subsection 4.1) and individual characteristics elicited in the final survey (see Subsection 4.5.1). We proceed in several steps:

- Considering each strata variable and each individual characteristic as a separate outcome variable, we regress this outcome on a full set of treatment dummies. Each treatment has the attributes: advisor round 1 [black hand, white hand], advisor round 2 [black hand, white hand], piece rate [low, high], information [no, yes]. We, hence, consider $2^4 = 16$ treatments. We include 15 dummies in our regressions. A F-test for joint significance will then presented for each regression. We expect that all coefficients are jointly statistical insignificant.[19]
- We will also follow the suggestions of Rosenbaum and Rubin [1985] and Imbens and Wooldridge [2009] and use a test of standardized differences to analyze covariate (in)balance. Rosenbaum and Rubin [1985] highlight that a standardized difference greater than 20 is “large” and points to imbalances. Given the high number of potential comparisons, we focus on our main comparisons: black vs white hand round 1; black vs white hand round 2; low versus high piece rate; no information vs information.

[19] By insignificance we understand p-values exceeding 0.1 throughout this document.

### 3.1.2 Attrition Checks

We also test for systematic attrition in several steps.

- We will check whether strata variables of individuals who do not complete both rounds are comparable to those of individuals who complete the study. For that purpose, we will consider standardized differences.
- We regress an indicator $A_i$ for individuals $i$ who have completed both rounds on the vector of strata variables. This will inform us about whether subjects from specific strata are more or less likely to attrit.
- We use probit and OLS models and regress $A_i$ on the full set of 15 treatment dummies. We then use F-tests to test if all coefficients are jointly insignificant.

### 3.1.3 Manipulation Checks

We test to what extent the participants' beliefs about the advisors' race is manipulated by the skin-color treatment. As part of the final survey, we ask if participants remember (a) the first advisor's race (Caucasian: yes, no, don't know; African American: yes, no, don't know) and (b) the second advisor's race (Caucasian: yes, no, don't know; African American: yes, no, don't know). From this information, we create two dummy variables. The first indicates that a participant correctly remembered the race of the first advisor (*FAC*). The second indicates that a participant correctly remembered the race of the second advisor (*SAC*). We also construct a dummy variable for subjects who remember the first advisor's skin color as black (*FAB*), and a similar dummy for the second advisor (*SAB*). Using these indicators, we perform the following manipulation checks:

- We expect *FAC* and *SAC* to indicate imperfect manipulation and use binomial tests to test the hypothesis that at least 95 percent of all individuals were correctly manipulated. In case we reject the null hypothesis, we will run the following regressions: First, we will regress the indicator for subjects who state that they remember the first advisor as being black, *FAB*, on an indicator for treatments featuring a black advisor, *B*.[20] Similarly, we will regress the indicator for subjects who state that they remember the second advisor as being black, *SAB*, on an indicator for treatments featuring a black advisor in the second round (*WW* and *WB*).
- Considering both variables indicating correct beliefs about the advisor's race (*FAC* and *SAC*) separately, we will also regress the respective indicator on a dummy *B* indicating that the advisor was indeed black (OLS and probit models). From this analysis, we will learn if reporting mistakes depend on an advisor's race. We also regress *FAC* and

[20] For subjects stating that they do not remember the advisor's race, the indicator will be coded as missing. These individuals are excluded from the main analysis.

*SAC* on the 15 previously mentioned treatment dummies, and use a F-test for joint significance.

### 3.1.4 Check if Information Treatment Equalizes Beliefs

We test whether our information treatment is able to equalize participants' second-round beliefs about their performance under the first-period and the second-period advisor. See Subsection 4.3 for variable definitions.

- We first consider the dummy variable $DB2$, which takes a value of one if the believed performance under the first-period advisor equals the one under the new advisor. Focusing on participants in the information treatment only, we also use binomial tests to test the hypothesis that 90 percent of all individuals were correctly manipulated.
- Considering observations in the information and no information treatments, we also regress $DB2$ on (a) a dummy $B$ indicating that the advisor was black, (b) a dummy $I$ indicating whether an individual received the information treatment, and (c) an interaction term between $B$ and $I$ (OLS and probit models). From this analysis, we learn whether our information treatment tends to equalize participants' beliefs and also how this effect depends on an the advisor's race. We also regress $DB2$ on the 15 previously mentioned treatment dummies and use a F-test for joint significance.
- To further study the structure of how our information treatment impacts beliefs, we consider an additional outcome $BD2$. This variable is defined as the difference between the expected performance under the second-round advisor and the new advisor. We then regress $DB2$ on the 15 previously mentioned treatment dummies and use a F-test for joint significance.

### 3.1.5 Check for Voice and Actor Effects

The advisor's voice and aspects of the advisor's hand other than skin color could impact individuals' choices and behaviors. To test for this possibility, we will estimate versions of our main specifications (outlined in Subsection 3.2) that control for voice and hand dummies.

### 3.1.6 Instructional Manipulation Checks

Our design also implements a standard instructional manipulation checks [Oppenheimer *et al.*, 2009], measuring if participants pay attention to instructions. More precisely, when presenting our experimental instructions, we ask participants to answer a single survey question.[21] However, in the instructions explaining the basics of our study, we request participants to ignore this question. Instead, regardless of what the true answer is, we ask them

[21] The question is: "How many MTurk HITs have you ever participated in?"

to fill in a specific number. Participants who do not fill in this number will be excluded from the rest of the study; i.e., we will not collect further data for these individuals.

## 3.2 Treatment Effects

**Research Questions:** Using simple treatment comparisons, our design allows us to examine the following topics that correspond to our research questions defined in Subsection 1.3:

Primary research questions:

1. How does the advisor's race impact individuals' advice-seeking and advice-utilization behavior?
2. How does the advisor's race affect individuals' performance?

Secondary research question:

3. Through which channels does the advisor's race impact advice seeking?

**Estimation Strategy:** In this paper, we apply a regression-based estimation approach of treatment effects. We choose the estimator according to the type of outcome.

- If the outcome is continuous, we use OLS.
- In the case of binary outcomes, we estimate linear probability and probit models.
- In the case of count data (e.g., number of puzzles solved with advisor's strategy), we use OLS and Poisson models.
- If the outcome is fractional (e.g., share of puzzles solved with the strategy proposed by the advisor), we use OLS and also follow Papke and Wooldridge [1996].
- If the outcome is ordinal, we estimated ordered logit models.
- In the case of categorically distributed dependent variables, we estimate multinomial logistic or conditional logit regressions (whatever is appropriate).
- When analyzing the individuals' preference for the second-round advisor (preference for first-period advisor, second-period advisor, or indifference), we will use a rank-ordered logit model allowing for indifference between the two options [Allison and Christakis., 1994].
- If we consider multiple outcomes that measure a similar construct, we follow Kling *et al.* [2004] and calculate average (standardized) effect sizes across multiple outcomes. We also report OLS results for each equation.

Some further estimation details are as follows:

- If we find that many individuals are inattentive to the advisor's race (i.e., they falsely report the advisors' race), we also estimate LATEs using $2SLS$ models. Particularly, in this case, we construct an indicator taking a value of one if an individual reported

that the advisor is black. We then instrument this variable with our treatment dummy $B$ to obtain the LATE.

- We will also perform mediator analyses in our paper. We follow the methodology developed in Imai *et al.* [2008], Imai *et al.* [2011], and Imai *et al.* [2013]. In particular, if we can precisely control mediators, we apply the estimator developed for the parallel design [Imai *et al.*, 2013]. Otherwise, we use a $2SLS$ approach and also estimate bounds using the non-parametric approach for binary mediators developed in Imai *et al.* [2013].

**Inference:** We use Huber-White standard errors. Whenever appropriate, we will cluster standard errors at the individual level. In addition, we will also use randomization inference to test the exact null of no treatment effect [Fisher, 1937; Athey and Imbens, 2017]. The tests will use 1000 random draws. In some specifications, we will estimate effects of more than one treatment on our outcomes or we examine multiple outcomes. In these instances, we correct for multiple hypothesis testing along the lines of the method proposed by List *et al.* [2016].

### 3.2.1 Topic 1: Advisor's Race and Behavior to Seek and Utilize Costly Advice

Our first goal is to analyze how an advisor's race impacts an individuals advice-seeking and advice-utilization behavior (Research question 1, see subsection 1.3). We proceed as follows:

**Outcomes:** We consider four types of outcomes, two primary and two secondary:

Primary outcomes:

- To measure costly advice seeking, we consider an individual's willingness to pay for being advised by the first advisor ($WTP1$). This outcome is collected in round 1 only. See Subsection 4.2.1 for further details.
- Our design also includes variables to measure from *whom* individuals tend to seek advice, which are measured in round 2 only. See Subsection 4.2.2 for a description of how the variables are constructed.
  - $IR$ indicates whether individuals (a) prefer the first-period advisor, (b) the second-period advisor, or (c) are indifferent.
  - $DI2$ takes a value of one if individuals indicate a preference for the second-period advisor (and zero otherwise).

Secondary outcomes:

- We use three variables to measure whether and to what degree individuals utilize advice. See Subsection 4.2 for further details.
  - To study the extensive margin of whether an individual utilizes advice, we consider dummy variables which take a value of one if an individual solved at least one puzzle with the strategy proposed by the advisor. We construct two dummies: $US1$ and $US2$. $US1$ ($US2$) indicates that a participant used the strategy explained by the first (second) advisor. We will consider both dummies when analyzing data from each of the two rounds.
  - To study the overall effect on advice utilization, we study the number of puzzles solved with the strategy proposed by the first advisor ($NS1$) and second advisor ($NS2$). These measures potentially reflect a higher performance due to advice utilization. We will consider both variables when analyzing data from each of the two rounds.
  - To study performance-adjusted advice utilization, we consider the share of all completed puzzles solved with the strategy proposed by the first advisor ($FS1$) and the second advisor ($FS2$). We will consider both variables when analyzing data from each of the two rounds.

- We also include a further secondary outcome measuring from *whom* individuals tend to seek advice in round 2. See Subsection 4.2.2 for a description of how the variable is constructed.
  - $WTP2$ measures the second-round willingness to pay for being advised by the preferred advisor given that the less preferred advisor has been selected by the mechanism. As subjects state their willingness to pay for either the first-round or the second-round advisor (conditional on their previous ranking of the two), we make a linearity assumption and define $WTP2$ as being equal to an individual's stated willingness to pay for being advised by the second advisor if the second advisor is preferred ($WTP2$), and equal to minus the individual's stated willingness to pay for being advised by the first advisor if the first advisor is preferred. $WTP2$ thus measures the willingness to pay for being advised by the second-round advisor.

**Sample Round 1:** We restrict the sample to individuals who completed both rounds of the experiment. We also exclude individuals who failed to answer our attention check correctly. Furthermore, we only use observations for whom a zero price was drawn by the willingness to pay mechanism in the first round, and for whom the stated preference over advisors does not play a role in the second-round mechanism (95% probability to get the new advisor, ir-

respective of own preference). Lastly, we exclude individuals who responded to final survey questions on remembering advisor's race with the option "I did not see the video or I don't remember" (for both $FS9$ and $FS12$ defined in subsection 4.5.1). The sample consists of round 1 data only.

**Sample Round 2:** We use identical sample restrictions as for round 1 data. The only difference is that the sample consists of round 2 data only.

**Quantities of Interest:** Everybody in a treatment group is exposed to a treatment presented on the website. We, hence, are confident that we are able to identify average treatment effects (ATEs) for the population of MTurk workers. However, as described previously, our experimental design includes manipulation checks. After the experiment, individuals will be asked about the advisors' race ($FS9$ and $FS12$ defined in subsection 4.5.1). In case we find evidence for imperfect manipulation, we will, instead, estimate local average treatment effects (LATEs).

**Main Specification Round 1:** Our main regression for first-round data is:

$$Y_i = \beta_0 + \beta_1 \cdot B_i + X_i \cdot \gamma + \varepsilon_i, \tag{1}$$

where $Y_i$ reflects the dependent variable for individual $i$ in round 1. The dependent variables are specified as defined previously. $B_i$ indicates whether individual $i$ is part of the black-hand ($B_i = 1$) or white-hand ($B_i = 0$) treatment in round 1. As it is common in the literature evaluating randomized controlled trials, we also include the vector of strata variables as controls ($X_i$).[22]

Example: Consider the case in which $Y_i$ reflects the willingness to pay for advise from the first advisor. Given that (a) we randomly assign the skin-color treatment and (b) the variation of the hand types and voices is orthogonal to the main treatment, $\hat{\beta}_1 < 0$ would indicate that individuals have a lower willing to pay for advice from black advisors.

**Main Specification Round 2:** Our main regression to analyze second-round data is:

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 \cdot BW_i + \beta_2 \cdot WW_i + \beta_3 \cdot WB_i \\ & + I_i \times (\beta_4 + \beta_5 \cdot BW_i + \beta_6 \cdot WW_i + \beta_7 \cdot WB_i) \\ & + X_i \cdot \gamma + \varepsilon_i, \end{aligned} \tag{2}$$

[22] In our main specification, we pool over our piece rate treatments.

where $Y_i$ refers to one of the round-two outcomes, $WW_i = 1$ indicates the treatment in which the first and second advisors are white (otherwise $WW_i = 0$), $BW_i = 1$ refers to the treatment in which only the second advisor is white, and $WB_i = 1$ reflects the treatment in which only the first advisor is white. $I_i = 1$ indicates that a participant $i$ receives the information treatment, and $X_i$ again indicates the vector of strata variables.[23]

**Further Specifications:** We also estimate the following variants of model (1) and (2):

1. Models with different sets of control variables (round 1):
   - The first variant will not include any control variables.
   - The second variant will account for an extended set of controls including the strata variables and in addition dummies obtained from participants' responses to questions in the final survey (see Subsection 4.5.1 for details). In particular, we include a gender dummy $FS1$, age dummies constructed from $FS2$ (defined by predefined age categories), a dummy that measures whether individuals already knew how to solve the sliding puzzle before participating in the HIT $FS6$, and a dummy for whether or not a participant is born in the US constructed from $FS3$. On top of that, instead of including the binary location dummy, it includes state dummies.
2. Fixed-effects specifications: When studying the overall effect on advice utilization, we can further increase statistical power by estimating fixed-effects specifications. In particular, we will estimate the following regressions (round 1):

$$Y_{i,p} = \beta_0 + \beta_1 \cdot B_i + X_i \cdot \gamma + \tau_p + \varepsilon_{i,p},$$

and (round 2):

$$\begin{aligned} Y_{i,p} = {} & \beta_0 + \beta_1 \cdot BW_i + \beta_2 \cdot WW_i + \beta_3 \cdot WB_i \\ & + I_i \times (\beta_4 + \beta_5 \cdot BW_i + \beta_6 \cdot WW_i + \beta_7 \cdot WB_i) \\ & + \tau_p + \varepsilon_{i,p}, \end{aligned}$$

where $Y_{i,p}$ is a dummy indicating whether individual $i$ solved puzzle number $p$ using the strategy proposed by the first (or second) advisor. $\tau_p$ represents a puzzle fixed effect.

[23] When studying performance outcomes, we will also consider regressions which analyze pooled skin color treatments ($WB$ and $BB$ vs. $WW$ and $BW$).

### 3.2.2 Topic 2: Advisor's Race and Advisee's Performance

Our second goal is to estimate how the advisor's race impacts the participants' performance (Research question 2, see subsection 1.3). We, again, consider both rounds separately.

**Sample Round 1:** We use the same sample restrictions as in Subsection 3.2.1. Again, we focus on individuals for whom the willingness to pay mechanism randomly draws a zero price. This design element ensures that there is no self-selection of participants into watching the full tutorial. Put differently, conditional on a price of zero, the allocation of participants into the black-hand and white-hand treatment is still random. We can, hence, study the causal effect of how the advisor's race impacts the participants' performance, independent of her willingness to pay for the full tutorial. We again focus on first-round observations and individuals who complete both rounds.

**Sample Round 2:** We use the same sample restrictions as in Subsection 3.2.1.

**Outcomes** We consider several performance measures (see Subsection 4.2 for details):

Primary outcomes:

- To measure the participants' overall performance, we consider the number of solved puzzles as an outcome variable ($NP$).

Secondary outcomes:

- To measure the participants' productivity, we use the completion time per puzzle ($CP$).
- To measure the participants' puzzle-solving efficiency, we count the number of moves used to solve the puzzle ($NM$).

**Quantities of Interest:** Similar to Subsection 3.2.1.

**Main Specifications Round 1:** We will present evidence from two types of main specifications. First, to analyze effects on participants' overall performance $NP$ in round 1, we will estimate regressions in the spirit of equation (1). Second, to estimate impacts on participants' first-round productivity ($CP$) and efficiency ($NM$), we consider the following specification:

$$Y_{i,p} = \beta_0 + \beta_1 \cdot B_i + X_i \cdot \gamma + \tau_p + \varepsilon_{i,p}, \tag{3}$$

where $X_i$ again stands for strata controls, and $\tau_p$ is a puzzle-specific fixed effect.

**Main Specifications Round 2:** We will present evidence from two types of main specifications. First, to analyze effects on participants' overall performance $NP$ in round 2, we will estimate regressions in the spirit of equation (2). Second, to estimate impacts on participants' second-round productivity ($CP$) and efficiency ($NM$), we consider the following specification:

$$\begin{aligned} Y_{i,p} = {} & \beta_0 + \beta_1 \cdot BW_i + \beta_2 \cdot WW_i + \beta_3 \cdot WB_i \\ & + I_i \times (\beta_4 + \beta_5 \cdot BW_i + \beta_6 \cdot WW_i + \beta_7 \cdot WB_i) \\ & + X_i \cdot \gamma + \tau_p + \varepsilon_{i,p}. \end{aligned} \tag{4}$$

**Further Specification:** We will also consider two types of further specifications. Models with different sets of control variables: For models (3) and (4), we will present specifications with and without control variables as described in Subsection 3.2.1.

### 3.2.3 Topic 3: Channel Through which Advisor's Race Impacts Selection of Advisor

We also study through which channels the advisor's race impacts advice-seeking behavior (secondary research question). We focus on two types of explanations. First, participants might dislike black advisors, meaning that they have a taste for discrimination in advice seeking. Along these lines, individuals may prefer to seek advice from white advisors to avoid interactions with black advisors. Second, individuals might expect to perform worse under black advisors, resulting in a lower willingness to pay for advice from black advisors and/or a lower preference for watching a tutorial given by a black advisor. We label this type of behavior as statistical discrimination in seeking advice.[24] To separate both types of effects, we exploit an experimental design that allows us to identify causal mechanisms along the lines of Imai *et al.* [2013]. We draw on standard methods of mediation analysis [see, e.g., Imai *et al.*, 2011].

**Sample Round 2:** We analyze the channels using second-round behavior (participants decide between advisors only in the second round). The sample restrictions are as described in Subsection 3.2.1.

**Outcomes:** To explore this secondary research question, we consider two outcomes.

- $DI2$: Dummy indicating preference for second advisor.

[24] One explanation for this type of discrimination is that participants might use the advisor's race as a proxy for her unobserved ability of giving useful advice. Another possible explanation is that participants may believe in certain barriers limiting the effectiveness of cross-racial giving and/or receiving advice.

- $WTP2$: Transformed willingness to pay for being advised by preferred advisor. See Subsection 3.2.2 for further details.

**Mediator:** Our mediator of interest is the expected number of solved puzzles in the second round, conditional on being instructed by the second advisor ($B2$). See Subsection 4.3 for further details.

**Quantities of Interest:** To identify the channels, we estimate three types of quantities [Imai *et al.*, 2013]:

1. The average treatment effect (ATE) of the skin-color treatments.
2. The average direct effect (ADE) of the skin-color treatments.
3. The average causal mediation effect (ACME) of the skin-color treatments.

To see why these quantities are of interest, we highlight that statistical discrimination can be understood as a causal process: An advisor's race in round 2 (treatment) causally impacts participant's ranking of the first-round and new advisor (outcome) through a mediating variable (belief about own performance under both advisors).[25] The analysis of statistical discrimination is one that aims at identifying the average causal mediation effect (i.e., the effect of a treatment that runs through a mediating variable). In a similar vein, taste-based discrimination in advice seeking can be understood as a direct effect of the skin-color treatment on the outcome, which is not transmitted through beliefs.[26] Thus, our inferential goal is to decompose the total causal effect of our race treatment (ATE) into the indirect effect (ACME), representing statistical discrimination, and the direct effect (ADE), representing taste-based discrimination.

**Main Specifications Round 2:** The following estimation strategy assumes that the information treatment equalizes beliefs (along the lines of our tests outlined in Subsection 3.1). In this case, our estimation strategy proceeds in several steps: First, we use specification (2) to estimate the ATEs on $DI2$ and $WTP2$. Second, focusing on observations in the information treatment, we follow Imai *et al.* [2013] and identify the average direct effect (ADE) of

[25] As an example, consider two of our treatments: The one in which the first and second advisors are white ($WW$) and the one in which only the first advisor is white ($WB$). In this case, being confronted with a black advisor in the second period may negatively impact a participant's belief about her performance under the second advisor (compared to the scenario in which the second advisor is white) which, in turn, may result in a lower ranking of the second advisor in the $WB$ treatment in round 2.

[26] A taste-based discriminator prefers white advisors, even if she expects to perform equally under white and black advisors.

the $BW$ treatment (relative to the $BB$ treatment) as:

$$\begin{aligned} ADE_{BW} = \int \{&\mathbb{E}(Y_i|BB_i = 0, BW_i = 1, WW_i = 0, WB_i = 0, B2_i = b, I_i = 1) \\ &-\mathbb{E}(Y_i|BB_i = 1, BW_i = 0, WW_i = 0, WB_i = 0, B2_i = b, I_i = 1)\} \mathrm{d}F_{B2_i|I_i=1}(b), \end{aligned} \tag{5}$$

where $B2_i$ refers to a participant's belief about her performance under the *second* advisor and $b$ is one particular level of beliefs. We, hence, estimate the $ADE_{BW}$ by computing the differences in the mean outcomes between the $BW$ and $BB$ treatments for each value of the mediator, and then average these values over the observed distribution of the mediator.[27]

The basic idea behind this estimation strategy is as follows.[28] First, given that individuals in both treatments face similar treatment histories in the first period, the expected second-round performance under the first advisor should be, on average, the same across both treatments. Second, if our information treatment successfully eliminates belief differences between the first and second advisor, the expected second-round performance under the second advisor should also be identical across treatments. Third, our estimator of the ADE then exploits participants' beliefs about their performance under the second advisor in order to "match" individuals in the $BW$ treatment to individuals in the $WW$ treatment. Specifically, it only compares individuals in the $BW$ treatment who expect to solve $b$ puzzles under the second (white) advisor to individuals in the $BB$ treatment who expect to solve $b$ puzzles under the second (black) advisor. We then say that there is, on average, taste-based discrimination if, conditional on similar beliefs, individuals more likely prefer the second advisor in the $BW$ than in the $BB$ treatment.

In a similar vein, we estimate the average direct effect of the $WW$ treatment (relative to the $WB$ treatment) as:

$$\begin{aligned} ADE_{WW} = \int \{&\mathbb{E}(Y_i|BB_i = 0, BW_i = 0, WW_i = 1, WB_i = 0, B2_i = b, I_i = 1) \\ &-\mathbb{E}(Y_i|BB_i = 0, BW_i = 0, WW_i = 0, WB_i = 1, B2_i = b, I_i = 1)\} \mathrm{d}F_{B2_i|I_i=1}(b). \end{aligned} \tag{6}$$

In this case, we say that there is, on average, taste-based discrimination if, conditional on similar beliefs, individuals more likely prefer the second advisor in the $WW$ than in the $WB$ treatment.

[27] Note that

$$ATE_{BW,I=1} = \mathbb{E}(Y_i|BB_i = 0, BW_i = 1, WW_i = 0, WB_i = 0, I_i = 1) - \mathbb{E}(Y_i|BB_i = 1, BW_i = 0, WW_i = 0, WB_i = 0, I_i = 1)$$

is not necessarily equal to $ADE_{BW}$. Only if $M_i|I_i$ and $BW_i$ are statistically independent, then $ATE_{BW} = ADE_{BW}$.

[28] The identifying assumptions are as follows: (a) Randomization of the treatment; (b) Consistency (experimental subject reveals same value of the outcome if the treatment and the mediator take a particular set of values, whether or not the value of the mediator is chosen by the subject or assigned by the experimenter); (c) Randomization of the mediator; (d) No causal interaction between the treatment and the mediator; (e) Information treatment eliminates belief differences.

Exploiting equations (5) and (6), we estimate the average causal mediation effects of the $BW$ and $WW$ treatments as:

$$ACME_{BW} = ATE_{BW} - ADE_{BW}, \tag{7}$$

$$ACME_{WW} = ATE_{WW} - ADE_{WW}, \tag{8}$$

with

$$\begin{aligned} ATE_{BW} &= \mathbb{E}(Y_i|BB_i = 0, BW_i = 1, WW_i = 0, WB_i = 0, I_i = 0) \\ &\quad - \mathbb{E}(Y_i|BB_i = 1, BW_i = 0, WW_i = 0, WB_i = 0, I_i = 0) \\ &= \hat{\beta}_1, \end{aligned}$$

and

$$\begin{aligned} ATE_{WW} &= \mathbb{E}(Y_i|BB_i = 0, BW_i = 0, WW_i = 1, WB_i = 0, I_i = 0) \\ &\quad - \mathbb{E}(Y_i|BB_i = 0, BW_i = 0, WW_i = 0, WB_i = 1, I_i = 0) \\ &= \hat{\beta}_2 - \hat{\beta}_3. \end{aligned}$$

Intuitively, $ACME_{BW}$ and $ACME_{WW}$ reflect effects of the treatments on outcomes running through beliefs.[29]

**Further Specifications:** We will also consider two types of further specifications.

1. By definition, the strategy to estimate the $ADEs$ and $ACMEs$ only utilizes observations in the region of common mediator support, which enables us to obtain "matched observations." In the absence of common support, we may also employ regression models to estimate the conditional distribution of $Y_i$ given $T_i$, $M_i$, and $I_i$ [Imai *et al.* 2011]
2. If the manipulation of the mediator turns out to be imperfect, we follow Imai *et al.* [2011] and use the 2SLS estimator to estimate direct and indirect effects. A simple interpretation of the IV approach is as follows: The encouragement (information treatment) is used as an excluded instrument to identify how a change in the belief affects the ranking. Contemporaneously, one measures how a change in the second-period skin-color treatment affects beliefs. By putting both pieces together, one can identify how the treatment affects the outcome through the mediator. One can also separately identify the direct effect of the treatment. Imai *et al.* [2013] also propose

[29] In general, we could perform a similar mediator analysis by comparing the $WB$ with the $BB$ treatment and the $BW$ with the $WW$ treatment. However, in those comparisons, advisors face advisors of different race in round 1. This might affect their beliefs and, hence, confound the estimation of the corresponding mediator effects.

a non-parametric estimation strategy that allows to bound the ACME and ADE in this case, which, if feasible, we will also employ under imperfect manipulation (See subsection 3.1.4).

## 3.3 Heterogeneous Treatment Effects

Our design allows us to study treatment effect heterogeneity in various dimensions:

1. *Piece Rate:* Given that we randomly manipulate the piece rate for each puzzle orthogonally to changing the advisor's skin color, we can analyze how and to what extent the effects of our skin-color treatments depend on the piece rate (a measure for the costs of discrimination). For that purpose, we will re-run the whole analysis described in Subsection 3.2 such that it allows for piece rate heterogeneity. In practice, we achieve this by interacting our treatment dummies with a high-piece rate dummy. In case of no statistical interaction effects, we will report pooled results in our paper.
2. *Strata Variables:* All analyses will also be done separately for the following categories of our strata variables: (a) north vs other ($SR$), (b) low vs high education ($LS$), (c) white vs black vs others ($R$). This analysis is rather exploratory, as we may lack sufficient statistical power. Again, we will use interaction models for this purpose.
3. *Further Variables:* Furthermore, the analyses will be done separately by gender ($FS1$) and political attitudes ($FS4$ and $FS5$). We will also consider heterogeneity in the IAT D-score ($IAT$), using a non-parametric estimation approach along the lines of Cagala *et al.* [2019].

## 3.4 Further Analysis

1. *Evaluation of Advisor:* To connect our work to Mengel *et al.* [2019], we will also evaluate whether our skin-color treatments impact the participants' evaluations of the advisor. This can be done separately for round 1 and round 2.

2. *Tab-switching behavior:* To test the idea that individuals may be less attentive once they see a black advisor, we analyze whether our skin-color treatments affect tab-switching behavior. We consider switching behavior during the trailer and main video. This can be done separately for round 1 and round 2. To identify times of low puzzle-solving effort, we can also study tab switching behavior during the two puzzle-solving periods.

3. *Validity of beliefs:* To descriptively test whether participants correctly assess the value of the tutorial, we study the relationship between individuals' performance (after having watched the first video), their beliefs ($B1F$), and their willingness to pay for the first tutorial ($WTP1$). This can be done separately by skin-color treatments.

# 4 Variables

## 4.1 Strata Variables

We collect three variables to stratify our sample.

### 4.1.1 Race ($R$)

- *Type:* survey
- *Time of Measurement:* Beginning of experiment.
- *Question:* What is your race or origin? (select the one that best describes you) [US Census question]
- *Answers:* Answers are clustered into three categories
  - White
  - Black or African American
  - Others: Hispanic, Latino, or Spanish; American Indian, or Alaska Native; Asian Indian; Chinese; Filipino; Japanese; Korean; Vietnamese; Other Asian; Native Hawaiian; Guamanian or Chamorro; Samoan; Other Pacific Islander; Some other race
- *Transformation of data to create variable:* Construction of three categories (black, white, others) as indicated above.

### 4.1.2 Level of Schooling ($LS$)

- *Type:* survey
- *Time of Measurement:* Beginning of experiment.
- *Question:* What is the highest degree or level of school you have COMPLETED? (if currently enrolled, select the previous grade or highest degree received) [US Census question]
- *Answers:* Answers are clustered into two categories
  - Low Education: No schooling completed; Nursery school; Kindergarten; Grade 1 through 11; 12th grade, no diploma; Regular high school diploma; GED or alternative credential; Some college credit, but less than 1 year of college credit; 1 or more years of college credit, no degree; Associate's degree (for example: AA, AS)
  - High Education: Bachelor's degree (for example: BA, BS); Master's degree (for example: MA, MS, MEng, MEd, MSW, MBA); Professional degree beyond Bach-

elor's degree (for example: MD, DDS, DVM, LLB, JD); Doctoral degree (for example: PhD, EdD)

- *Transformation of data to create variable:* Construction of two categories (low education, high education).

#### 4.1.3 State of Residence (*SR*)

- *Type:* survey
- *Time of Measurement:* Beginning of experiment.
- *Question:* In which U.S. state is your usual residence (the place where you live most of the time)?
- *Answers:* Answers are clustered into two categories
  - South [As defined by US Census Burreau]: Alabama; Arkansas; Delaware; Florida; Georgia; Kentucky; Louisiana; Maryland; Mississippi; North Carolina; Oklahoma; South Carolina; Tennessee; Texas; Virginia; West Virginia; District of Columbia
  - North: Alaska; Arizona; California; Colorado; Connecticut; Hawaii; Idaho; Illinois; Indiana; Iowa; Kansas; Maine; Massachusetts; Michigan; Minnesota; Missouri; Montana; Nebraska; Nevada; New Hampshire; New Jersey; New Mexico; New York; North Dakota; Ohio; Oregon; Pennsylvania; Rhode Island; South Dakota; Utah; Vermont; Washington; Wisconsin; Wyoming
- *Transformation of data to create variable:* Construction of two categories (South, North).

### 4.2 Main Outcome Variables

#### 4.2.1 Willingness to Pay for First Advisor (*WTP*1)

- *Type:* survey
- *Time of Measurement:* Round 1. After Trailer.
- *Description and Question:* We will now determine whether you will access the full e-learning tutorial as follows.
  1. We have added the amount of $1 to your payoff account. You can use all or part of this amount to access the tutorial.
  2. Please use the slider below to indicate the highest price you are willing to pay to watch the e-learning tutorial.
  3. Next, the computer will randomly draw a price for the tutorial. The price is a number between $0 and $1.
  4. If your stated willingness to pay is equal to or above the price drawn, you will

buy the tutorial. If your stated willingness to pay is lower than the price drawn, you will not buy the tutorial. Instead, you will watch the entertainment video that provides no instructions on how to solve the puzzle. Note that the price you pay will be the price drawn by the computer, not your stated willingness to pay.

It is in your interests to state the highest price that you are willing to pay for the tutorial:

- If you state a lower amount than your true willingness to pay, you may miss the chance to watch the tutorial at a price which is lower than what you think is the value of the tutorial for you.
- If you state a higher amount than your true willingness to pay, you may end up buying the tutorial at a price which is higher than what you think is acceptable.

Further notes:

- You will have five minutes to solve as many puzzles as possible.
- The entertainment video and the e-learning tutorial are of equal length.
- You will earn a bonus of $[piece rate cost] for every puzzle you solve.
- The money you do not spend for the tutorial is added to your payoff.
- The money you spend is not distributed to the instructor.

- *Answer:* Individual's stated WTP using a fine-grained slider (101 steps), ranging from 0 to 100 cents.
- *Transformation of data to create variable:* We use the raw WTP data.

#### 4.2.2 Ranking of First and Second Advisor in Round 2 (*R*)

- *Type:* survey
- *Time of Measurement:* Round 2. After Trailer.
- *Description and Question:* In contrast to the first round, you will watch the tutorial for sure. Instead of choosing between the entertainment video and the tutorial, you will now select one of two potential instructors. The tutorial will explain way to solve the puzzle that is faster than the one presented before.
  - The instructor will either be the one from the first period (first instructor) or the one you have just seen in the trailer (second instructor).
  - Regardless of which instructor is selected, the length of the tutorial will be the same. [information treatment only: Furthermore, when recording the tutorials, both instructors followed the same script. Therefore, the contents of the two tutorials are identical, including the layout of the puzzle, the steps taken to solve it, and the wording used to explain the strategy.]

– The selection of instructors works as follows: You first rank the two instructors. Then, the computer randomly draws one of the two situations described in the table below:

| Your ranking matters | Your ranking does not matter |
|---|---|
| – If you indicate that you prefer one of the instructors, you will get the preferred instructor with a 70% chance.<br>– With a 30% chance, you will get the less preferred one.<br>– If you indicate that you are indifferent between the two instructors, the chances to get the first or the second instructor will be equal. | – You will watch the second instructor, irrespective of how you ranked the instructors. |

By selecting the instructor you like most, you increase the chance that you will end up seeing this instructor. Hence, it is in your interest to tell us which instructor you really prefer.

- *Answer:* Now, we ask you to rank the two instructors. Please select one option:
  - I prefer the first instructor to the second instructor.
  - I prefer the first to the second instructor.
  - I am indifferent between the two instructors.
- *Transformation of data to create variable:* First, using the subjects' willingness to pay in the second round (see Subsection 4.2.3), we calculate an incentivized ranking variable ($IR$) as follows. (a) Set $IR = R$. (b) Replace $IR =$ "I am indifferent between the two instructors" if participants stated that they are not indifferent but indicate a willingness for being able to watch the preferred advisor of zero. Second, using $IR$, we also define a dummy variable, $DI2$, that takes a value of one if individuals indicate a preference for the second advisor (and zero otherwise).

### 4.2.3 Willingness to Pay for Preferred Advisor ($WTP2$)

- *Type:* survey
- *Time of Measurement:* Round 2. After Ranking.

- *Description and Question:* Now suppose the following situation occurs:

  1. The computer-based random draw determines that your ranking matters. Hence, your ranking affects the chances that either the first instructor or the second instructor is selected.
  2. Ultimately, the [non-preferred instructor] is selected.

  If this situation indeed occurs, would you be willing to pay a small fee to get the [preferred instructor] for sure, although the [non-preferred instructor] was initially selected?
  Please state your willingness to pay as follows:

  1. We have added the amount of $1 to your payoff account. You can use all or part of this amount to pay for being able to watch the [preferred instructor] for sure.
  2. Please use the slider below to indicate the highest price you are willing to pay to watch the [preferred instructor]. Once you have stated your willingness to pay, the computer will randomly draw a price for watching the [preferred instructor]. The price is a number between $0 and $1.
  3. If your stated willingness to pay is equal to or above the price drawn, you will get the [preferred instructor]. If your stated willingness to pay is lower than the price drawn, you will not watch the [preferred instructor]. Instead, the [non-preferred instructor] will present the tutorial. Note that the price you pay will be the price drawn by the computer, not your stated willingness to pay.

  With the following choice, you can influence which instructor will be selected in the aforementioned situation. Hence, it is in your interest to state the highest price that you are willing to pay for being able to watch the [preferred instructor]. After having watched the tutorial, you will have 5 minutes to solve as many puzzles as possible. Again, you will earn a bonus of $[piece rate cost] for every puzzle you solve.
- *Answer:* Individual's state WTP using a fine-grained slider (101 steps), ranging from 0 to 100 cents.
- *Transformation of data to create variable:* We define the $WTP2$ variable as a subjects willingness to pay for the second advisor. If the subject stated that she prefers the second advisor, $WTP2$ simply corresponds to her stated willingness to pay for the second advisor. If she, instead, indicated that she prefers the first advisor, the variable $WTP2$ corresponds to the negative value of her willingness to pay for the first advisor. A willingness to pay of zero indicates that the subject is indifferent between both advisors. In this case, $WTP2$ also takes a value of zero.

#### 4.2.4 Completion Time per Puzzle ($CP$)

- *Type:* performance measure based on game data
- *Time of Measurement:* Round 1 and 2
- *Measurement:* Individuals get 5 minutes to solve a $3 \times 3$ sliding puzzle. We measure the seconds used to solve each puzzle.
- *Transformation of data to create variable:* This variable is part of the raw data.

#### 4.2.5 Number of Solved Puzzles ($NP$)

- *Type:* performance measure based on game data
- *Time of Measurement:* Round 1 and 2
- *Measurement:* We count the number of puzzles solved in 5 minutes.
- *Transformation of data to create variable:* This variable is part of the raw data.

#### 4.2.6 Number of Moves to Solve Puzzles ($NM$)

- *Type:* performance measure based on game data
- *Time of Measurement:* Round 1 and 2
- *Measurement:* We count the number of moves used to solve each puzzle.
- *Transformation of data to create variable:* This variable needs to be created by counting the number of moves in the raw data.

#### 4.2.7 Used Strategy to Solve Puzzle ($US$)

- *Type:* strategy measure based on game data
- *Time of Measurement:* Round 1 and round 2
- *Measurement:* For each puzzle, we identify whether the first or second strategy was used to solve puzzle.
  - Dummy first strategy ($US1$): The variable takes a value of 1, if the following pattern occurs in the data (otherwise 0):

| | | |
|---|---|---|
| 2 | | 3 |
| 1 | x | x |
| x | x | x |

    followed by:

| | | |
|---|---|---|
| | 2 | 3 |
| 1 | x | x |
| x | x | x |

and:

| | | |
|---|---|---|
| 1 | 2 | 3 |
| | x | x |
| x | x | x |

– Dummy second strategy ($US2$): The variable takes a value of 1, if the following pattern occurs in the data (otherwise 0):

| | | |
|---|---|---|
| 1 | 3 | |
| x | 2 | x |
| x | x | x |

followed by:

| | | |
|---|---|---|
| 1 | | 3 |
| x | 2 | x |
| x | x | x |

and:

| | | |
|---|---|---|
| 1 | 2 | 3 |
| x | | x |
| x | x | x |

- *Transformation of data to create variable:* Create panel data that documents every move in panel. Identify above-mentioned data patterns in game move data.

#### 4.2.8 Number of Puzzles Solved with Each Strategy ($NS$)

- *Type:* aggregate measure for preferred strategy based on game data
- *Time of Measurement:* Round 1 and round 2
- *Measurement:* For each individual, we calculate the number of completed puzzled solved with the first ($NS1$) and second ($NS2$) strategy.
- *Transformation of data to create variable:* Construct measure based on strategy dummies.

#### 4.2.9 Fraction of Puzzles Solved with Each Strategy ($FS$)

- *Type:* aggregate measure for preferred strategy based on game data
- *Time of Measurement:* Round 1 and round 2
- *Measurement:* For each individual, we calculate the fraction of all completed puzzled solved with the first ($FS1$) and second ($FS2$) strategy.
- *Transformation of data to create variable:* Construct measure based on strategy dummies.

## 4.3 Mediators

### 4.3.1 Belief Measure in Round 1 ($B1$)

- *Type:* survey
- *Time of Measurement:* After WTP.
- *Question:* In the following, you will either watch the full e-learning tutorial or an entertainment video that provides no instructions on how to solve the puzzle. Then you'll be solving 3x3 sliding puzzles that are similar to the ones you saw in the trailer.
  - How many puzzles do you expect to solve in 5 minutes after having watched the full tutorial? [Use numbers only] ($B1F$).
  - How many puzzles do you expect to solve in 5 minutes after having watched the entertainment video? [Use numbers only] ($B1E$).
- *Answers:* Number in textbox.
- *Transformation of data to create variable:* Use raw data.

### 4.3.2 Belief Measure in Round 2 ($B2$)

- *Type:* survey
- *Time of Measurement:* After ranking.
- *Question:* In the following, you will watch a full e-learning tutorial describing a simpler strategy to solve the puzzle. The person explaining the new strategy will either be the instructor from the first period (first instructor) or the instructor you have just seen in the trailer (second instructor).
  - Using the new strategy, how many puzzles do you expect to solve in 5 minutes after having watched a tutorial presented by the first instructor? [Use numbers only] ($B2FI$).
  - Using the new strategy, how many puzzles do you expect to solve in 5 minutes after having watched a tutorial presented by the second instructor? [Use numbers only] ($B2SI$).
- *Answers:* Number in textbox.
- *Transformation of data to create variable:* Use raw data.

### 4.3.3 Binary Belief Measures ($DB$)

- *Type:* binary mediator shift indicator based on belief data
- *Time of Measurement:* Round 1 and round 2
- *Measurement:* Create two dummy variables:

- $DB1$: takes a value of one if $B1F = B1E$ (otherwise zero).
- $DB2$: takes a value of one if $B2FI = B2SI$ (otherwise zero).

#### 4.3.4 Belief Differences ($BD2$)

- *Type:* differences in beliefs between a participant's expected performance under the first-round and the second advisor
- *Time of Measurement:* round 2
- *Measurement:* Create differences between beliefs as follows: $BD2 = B2SI - B2FI$

#### 4.3.5 Tab-switching Behavior as Attention Measure ($TS$)

- *Type:* tab-tracking data
- *Time of Measurement:* permanently
- *Measurement:* We record (a) whether and (b) for how long individuals open new browser tabs (tab-tracking)
- *Transformation of data to create variable:* We construct several variables from the data.
  - $TSNT1$ : Number of instances in which an individual opens and closes a new tab during the first trailer
  - $TSNT2$ : Number of instances in which an individual opens and closes a new tab during the second trailer
  - $TSNF1$ : Number of instances in which an individual opens and closes a new tab during the first full tutorial
  - $TSNF2$ : Number of instances in which an individual opens and closes a new tab during the second full tutorial
  - $TSNS1$ : Number of instances in which an individual opens and closes a new tab during the first puzzle-solving period
  - $TSNS2$ : Number of instances in which an individual opens and closes a new tab during the second puzzle-solving period
  - $TSTT1$ : Time span during the first trailer (in seconds) in which an individual is not focusing on the tab with our website
  - $TSTT2$ : Time span during the second trailer (in seconds) in which an individual is not focusing on the tab with our website
  - $TSTF1$ : Time span during the first full tutorial (in seconds) in which an individual is not focusing on the tab with our website
  - $TSTF2$ : Time span during the second full tutorial (in seconds) in which an individual is not focusing on the tab with our website

- $TSTS1$ : Time span during the first puzzle-solving period (in seconds) in which an individual is not focusing on the tab with our website
- $TSTS2$ : Time span during the second puzzle-solving period (in seconds) in which an individual is not focusing on the tab with our website

The variables $TSNT1$, $TSNT2$, $TSNF1$, $TSNF2$, $TSTT1$, $TSTT2$, $TSTF1$, and $TSTF2$ serve as attention measures. The variables $TSNS2$, $TSTT1$, $TSTS1$, and $TSTS2$ approximate times during which individuals do not invest effort into solving the sliding puzzle.

## 4.4 Further Outcome Variables

### 4.4.1 Evaluation of Lecturer ($EL$)

- *Type:* survey
- *Time of Measurement:* Round 1 and 2. After full tutorial. Before puzzle solving.
- *Questions:* Please evaluate the instructor and the tutorial you have just seen:
  *[Recall that answering all questions is mandatory. There are 6 questions in total.]*
  1. The tutorial provides useful instructions for solving the sliding puzzle ($EL1$).
  2. What overall grade do you assign the tutorial ($EL2$)?
  3. The instructor does a good job explaining how to solve the puzzle ($EL3$).
  4. I benefitted from the instructor's explanations ($EL4$).
  5. What overall grade do you assign the instructor ($EL5$)?
  6. Would you recommend this instructor to people who want to learn how to solve the sliding puzzle ($EL6$)?
- *Answers:* Individuals indicate whether they completely agree, agree, neither disagree nor agree, disagree, completely disagree with the statements 1, 3, and 4. Individuals rate the lecturer's performance in the questions 2 and 5 on the scale A-F. Questions 6 is answered with yes or no. Participants can also indicate that they did not watch the tutorial.
- *Transformation of data to create variable:* Use raw data.

## 4.5 Further Variables

### 4.5.1 Final Survey ($FS$)

- *Type:* survey
- *Time of Measurement:* After experiment.

- *Questions:* Before concluding, we would like to ask you a few final questions. *[Recall that answering all questions is mandatory. There are 14 questions in total. Question 14 is optional but we would greatly appreciate your input.]*
    1. What is your gender ($FS1$)? *[Male / Female]*
    2. How old are you ($FS2$)? *[20 and less / 21-25 / 26-30 / 31-35 / 36-40 / 41-50 / 51-60 / 61 and above]*
    3. In which country were you born ($FS3$)? *[List of countries]*
    4. In politics today, do you consider yourself a Republican, Democrat or Independent ($FS4$)? *[Republican / Democrat / Independent / Don't know]*
    5. Who did you vote for in the 2016 Presidential Election ($FS5$)? *[Donald Trump / Hillary Clinton / Other or Don't know / Didn't vote]*
    6. Did you already know how to solve the sliding puzzle before participating in this HIT ($FS6$)? *[Yes / No]*
    7. The first instructor was experienced in giving advice to others. ($FS7$). *[Yes / No / I did not see the video or I don't remember]*
    8. The first instructor's gender was: ($FS8$). *[Male / Female / I did not see the video or I don't remember]*
    9. The first instructor's race was: ($FS9$). *[White / Black or African American / I did not see the video or I don't remember]*
    10. The second instructor was experienced in giving advice to others. ($FS10$). *[Yes / No / I did not see the video or I don't remember]*
    11. The second instructor's gender was: ($FS11$). *[Male / Female / I did not see the video or I don't remember]*
    12. The second instructor's race was: ($FS12$). *[White / Black or African American / I did not see the video or I don't remember]*
    13. When working on the puzzle for the second time, did you use the strategy that was presented in the second tutorial ($FS13$)? *[Yes, always / Yes, sometimes / No, never / (I did not watch both tutorials)]*
    14. Please, write down any comments you might have regarding the HIT that would help us to improve it in the future (Was everything comprehensible? Did you have enough time to finish? Did you face any technical issues?) ($FS14$)
- *Transformation of data to create variable:* We define binary variables *FAC* and *SAC*

that are equal to one if participant correctly classified first and second advisor's race in questions $FS9$ and $FS12$, respectively. We further define binary variables *FAB* and *SAB* equal to one if participant responded that the race of the first and second advisor's was black in questions $FS9$ and $FS12$, respectively. Otherwise, we use raw data. Don't know or did not watch video coded as missing.

### 4.5.2 Implicit Association Test (*IAT*)

- *Type:* survey
- *Time of Measurement:* Several weeks after completing the HIT, participants are invited to take part in another HIT. In this HIT, we ask participants to complete a version of a race implicit association test.
- *Measurement:* We use the single-target version of the implicit association test of Greenwald *et al.* [1998]. We employ the same instructions and the same structure as in Lowes *et al.* [2015].
- *Transformation of data to create variable:* We follow the following procedure to calculate IAT D-score (*IAT*). We apply the method suggested in Table 3 in Lane *et al.* [2007]:
  1. Delete trials greater than 10,000 msec
  2. Delete subjects for whom more than 10% of trials have latency less than 300 msec
  3. Replace each error latency with an error penalty computed as Stage mean + 600 msec
  4. Compute the "inclusive" standard deviation for all trials in Stages 3 and 6 and likewise for all trials in Stages 4 and 7
  5. Compute the mean latency for responses for each of Stages 3, 4, 6, and 7
  6. Compute the two mean differences ($\text{Mean}_{\text{Stage 6}}$ - $\text{Mean}_{\text{Stage 3}}$) and ($\text{Mean}_{\text{Stage 7}}$ - $\text{Mean}_{\text{Stage 4}}$)
  7. Divide each difference score by its associated "inclusive" standard deviation
  8. $IAT$ = the equal-weight average of the two resulting ratios

# Appendix

## A Timeline

The precise timing of our experimental design is as follows:

**Introduction**

---

**1. Welcome Screen**

**Text:**

- introduction of our team and project (study on perceptions of elearning)
- info on completion time
- further notes (voluntary, anonymity, data can be withdrawn, contact mail, all questions relevant)

**Input:**

- worker ID

---

**2. Strata Survey**

**Input:**

- Strata variables (race, level of schooling, state). See Subsection 4.1

**Round 1**

---

**3. Instructions**

**Text:**

- timing
- important details (consent statement, back button)
- sequence
- calculation of payoff
- further notes

**Inputs:**

- instruction attention check
- consent statement

---

**4. Trailer: First (Complex) Strategy**

**Treatments:**

- randomization into skin-color treatments as described in Table 1

**Inputs:**

- tab switching; also recorded in other rounds. See Subsection 4.3.5

**Parameters:**

- equal probability of receiving each of the treatments
- stratification

---

**5. Willingness to Pay to Switch to Tutorial**

**Text:**

- detailed explanation of WTP elicitation method
- choice: switch from entertainment video to full video

**Treatments:**

- high versus low piece rate treatment

**Parameters:**

- base pay: \$4 and piece rate: \$0.5 and \$1
- WTP: \$1 extra for "buying" video
- probability for real video: 95%
- equal probability of receiving piece-rate treatments
- stratification

**Input:**

- WTP for video in \$0.01 steps. See Subsection 4.2.1

---

**6. Survey: Belief about Performance**

**Text:**

- description of belief questions

**Inputs:**

- belief about number of solved puzzle under tutorial and entertainment video. See Subsection 4.3.1

---

**7. Tutorial: First (Complex) Strategy**

**Treatments:**

- randomization into skin-color treatments as described in Table 1

**Inputs:**

- tab switching. See Subsection 4.3.5

---

**8. Advisor Evaluation Survey**

**Inputs:**

- evaluation survey. See Subsection 4.4.1

---

**9. Sliding Puzzle**

**Parameters:**

- $3 \times 3$ sliding puzzle
- 5 minutes

**Inputs:**

- performance measures and strategy measures. See Subsection 4.2
- tab switching. See Subsection 3.1.2

**Round 2**

---

**10. Instructions**

**Text:**

- announcement of second video
- reminder of own performance in round 1 sliding puzzle

---

**11. Trailer: Second (Simple) Strategy**

**Treatments:**

- randomization into skin-color treatments as described in Table 1

**Inputs:**

- tab switching; also recorded in other round. See Subsection 4.3.5

---

**12. Information Treatment (same page as next step)**

**Treatment:**

- randomly selected subset of individuals receives information treatment

**Parameters:**

- probability of receiving the information treatment: 50%
- stratification

**Text:**

- Both advisors use similar scripts

---

**13. Ranking of First-round and New Advisor**

**Text:**

- detailed description of ranking-elicitation method
- choice: choose between first-round and new advisor

**Input:**

- ranking: preference for first advisor; preference for new advisor; indifference. See Subsection 4.2.2

---

**14. Willingness to Pay to Get Preferred Instructor**

**Text:**

- detailed explanation of WTP elicitation method
- choice: get preferred advisor for sure

**Treatments:**

- high versus low piece rate treatment as in step 6.

**Parameters:**

- base pay: \$4 and piece rate: \$0.5 and \$1
- WTP: \$1 extra for getting preferred advisor for sure
- probability for new advisor: 95%
- equal probability of receiving piece-rate treatments
- stratification

**Input:**

- WTP for getting preferred advisor in \$0.01 steps. See Subsection 4.2.3

---

**15. Survey: Belief about Performance**

**Text:**

- description of belief questions

**Inputs:**

- belief about number of solved puzzle under first-round and new advisor. See Subsection 4.3.2

---

**16. Tutorial: Second (Simple) Strategy**

**Treatments:**

- randomization into skin-color treatments as described in Table 1

**Inputs:**

- tab switching; also recorded in other rounds. See Subsection 4.3.5

---

**17. Advisor Evaluation Survey**

**Inputs:**

- evaluation survey. See Subsection 4.4.1

---

**18. Sliding Puzzle**

**Parameters:**

- $3 \times 3$ sliding puzzle
- 5 minutes

**Inputs:**

- performance measures and strategy measures. See Subsection 4.2
- tab switching. See Subsection 3.1.2

---

**19. Final Survey**

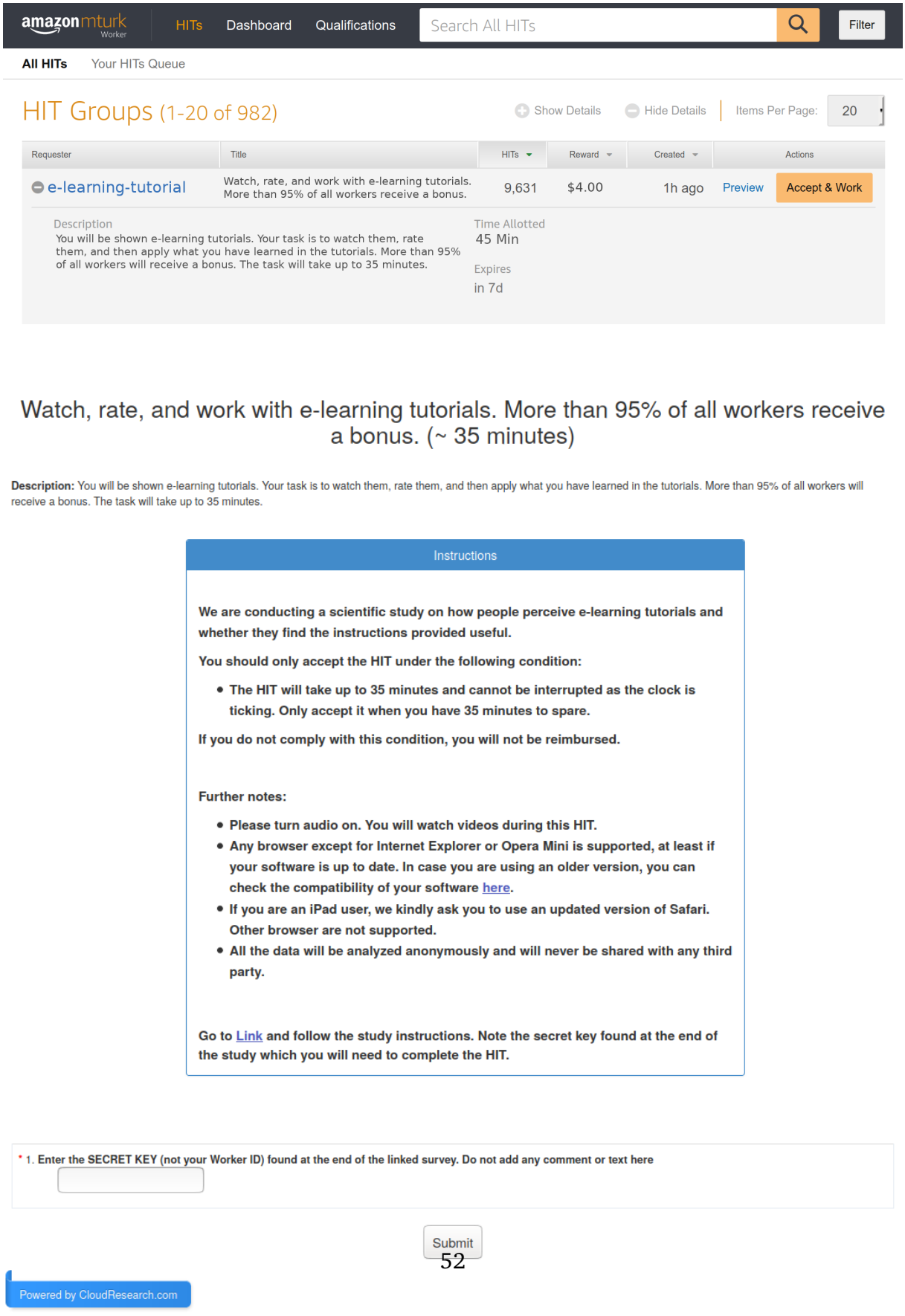
**Inputs:**

- Final survey questions. See Subsection 4.5.1

---

# B Description of HITs

**Figure B.1:** Appearance of HIT on MTurk



# C Video Production

In the following, we summarize how the videos were produced. The following description was part of the contract.

**Scope of Work**

- The project consists of pre-production, filming, and post-production work of different videos.
- First, two videos (labeled “long complicated” and “long simple”) showing a hand with an “intermediate” skin color will be filmed. The videos are different with respect to the used choreography. The duration of each clip depends on the provided choreography.
- Second, these videos will be digitally altered, resulting in four additional videos. Two videos (hereinafter called “long complicated black” and “long simple black”) display the exact same motion as the original video, but will be digitally altered to appear as to be from an “African” person with dark skin tone. The other two videos (hereinafter called “long complicated white” and “long simple white”) will also show the same motion as the original and will appear to be from a “Caucasian” person with a light skin tone.
- Furthermore, both choreographies will be done with two different set of hands, resulting in a total of 12 videos.
- Out of the 4 videos “long complicated black”, “long simple black”, “long complicated white”, and “long simple white”, 4 more videos will be cut being a short version of each individual video. The same will be done for the second set of hands.

**Further details**

- Each video (long complicated and long simple) will be filmed with 2 different hands, resulting in 4 original videos
- Each video will be digitally altered to change the intermediate hands to a “white” and an African “black” hand, resulting in 8 additional videos
- Hence, in total, there will be 12 videos:

| *Set of hands* | *choreography* | *Skin tone* |
|---|---|---|
| hand 1 | long complicated | original |
| | | white |
| | | black |
| | long simple | original |
| | | white |
| | | black |
| hand 2 | long complicated | original |
| | | white |
| | | black |
| | long simple | original |
| | | white |
| | | black |

- Out of each long video a short version will be cut
  - Short complicated: cut from final clips of long complicated
  - Short simple: cut from final clip of simple long
- Hence, in total, there will be 8 additional videos

| *Set of hands* | *Choreography* | *Skin tone* |
|---|---|---|
| hand 1 | short complicated | white |
| | | black |
| | short simple | white |
| | | black |
| hand 2 | short complicated | white |
| | | black |
| | short simple | white |
| | | black |

- Production details for producers and provided materials by client:
  - Each video will be shot on bluescreen
  - The bluescreen will be digitally replaced with a computer generated background displaying the template provided by the client.
  - The detailed choreography for the two initial types of videos (long complicated / long simple) will be provided by the client. This includes example video of both complicated long and simple long as well as a detailed description of the choreography and time codes. The producer makes sure that the hands' movements match the provided choreographies and fit exactly to the background (when hand/finger points to elements on the screen etc.).
  - The detailed choreography for the short videos (short complicated / short simple) will also be provided by the client.
  - Final grading and technical approval.
- Delivery specifics:
  - length: determined by provided choreography of each video
  - Format: 720p exr, 720p mov

**Milestones**

1. pre-production, proof of concept, background template selection, and hand model selection

2. principal photography of four clips as defined in the "scope of work" (complicated long and simple long for each hand)

3. final delivery of finished visual effects work as defined under "scope of work" for first hand (for complicated long and simple long; total of six videos). To avoid complica-

tions, the version with the black hand will be approved by the client before work on the white hand can start

4. final delivery of finished visual effects work as defined under "scope of work" for second hand (for complicated long and simple long; total of six videos). To avoid complications, the version with the black hand will be approved by the client before work on the white hand can start

5. final delivery of remaining eight videos (complicated short and simple short).

# D Performance of Complicated and Simple Strategy

We test if the second-round strategy ("simple") is indeed faster than the first-round strategy ("complicated"). To that end, we extended the "$A^*$ pathfinding algorithm" (proposed here). This algorithm allows us to count the minimum number of moves needed to solve the puzzle when using the simple or complicated strategy. Therefore, the algorithm allows us to compare the best possible performance under both strategies. The Python file is available upon request.

Before presenting the results of the pathfinding algorithm, let us introduce the precise puzzles that individuals solve in the experiment. Table 3 presents the used starting position. In each round, participants work on 15 different puzzles in a fixed and randomly chosen order. If they solve more than 15 puzzles, the puzzles are repeated in a similar order. To understand how to read Table 3, note that the starting position

$$\begin{array}{ccc} 1 & 6 & 4 \\ 7 & 3 & \\ 5 & 8 & 2 \end{array}$$

translates into the array [1, 6, 4, 7, 3, 0, 5, 8, 2].

Table 4 presents the results of the pathfinding algorithm for the puzzles presented in Table 3. Two observations stand out. First, the simple strategy is always faster than the complicated strategy. Considering the 30 puzzles in Table 3, on average, the algorithm executes 38.6 moves to solve the puzzle with the complicated strategy and 28.1 moves with the simple strategy. Second, by comparing Columns (4) and (1), one can immediately see that the $n$-th puzzle in round 2 can be solved faster (when using the simple strategy) than the $n$-th puzzle in round 1 (when using the complicated strategy).

# E Experimental instructions

**Table 3:** Starting Positions of Puzzles

| Puzzle $n$ | Round 1 | Round 2 |
|---|---|---|
| 1 | [1, 6, 4, 7, 3, 0, 5, 8, 2] | [4, 1, 5, 7, 2, 6, 0, 8, 3] |
| 2 | [6, 2, 8, 0, 1, 3, 7, 5, 4] | [8, 1, 7, 3, 2, 0, 6, 5, 4] |
| 3 | [6, 5, 2, 4, 1, 3, 7, 0, 8] | [8, 4, 5, 1, 2, 3, 0, 7, 6] |
| 4 | [0, 1, 6, 4, 7, 3, 8, 2, 5] | [8, 1, 4, 5, 3, 2, 0, 6, 7] |
| 5 | [5, 0, 8, 1, 2, 7, 6, 4, 3] | [0, 1, 5, 8, 2, 7, 6, 4, 3] |
| 6 | [3, 2, 4, 1, 0, 8, 6, 5, 7] | [7, 1, 2, 4, 0, 6, 5, 3, 8] |
| 7 | [7, 0, 1, 8, 3, 2, 5, 6, 4] | [7, 1, 8, 2, 3, 5, 4, 0, 6] |
| 8 | [6, 3, 8, 5, 4, 0, 7, 2, 1] | [6, 3, 1, 4, 8, 0, 5, 2, 7] |
| 9 | [5, 4, 6, 3, 7, 0, 2, 8, 1] | [1, 6, 2, 4, 3, 0, 5, 8, 7] |
| 10 | [6, 2, 7, 4, 8, 1, 5, 0, 3] | [5, 1, 4, 0, 3, 8, 6, 7, 2] |
| 11 | [5, 0, 2, 3, 7, 1, 8, 6, 4] | [8, 7, 4, 2, 5, 3, 1, 6, 0] |
| 12 | [1, 5, 0, 6, 3, 4, 7, 8, 2] | [1, 3, 0, 2, 7, 5, 8, 6, 4] |
| 13 | [5, 1, 8, 4, 6, 0, 2, 7, 3] | [4, 6, 8, 7, 5, 3, 2, 0, 1] |
| 14 | [3, 7, 6, 5, 2, 0, 4, 1, 8] | [1, 3, 6, 0, 4, 8, 5, 7, 2] |
| 15 | [0, 8, 1, 3, 6, 5, 4, 2, 7] | [8, 1, 4, 5, 0, 3, 7, 2, 6] |

**Table 4:** Performance of Complicated and Simple Strategy

| Puzzle $n$ | Round 1 | | Round 2 | |
|---|---|---|---|---|
| | Complicated (1) | Simple (2) | Complicated (3) | Simple (4) |
| 1 | 25 | 17 | 32 | 24 |
| 2 | 33 | 23 | 33 | 23 |
| 3 | 35 | 25 | 34 | 26 |
| 4 | 42 | 28 | 42 | 28 |
| 5 | 39 | 35 | 40 | 32 |
| 6 | 40 | 26 | 40 | 22 |
| 7 | 33 | 27 | 33 | 25 |
| 8 | 43 | 31 | 43 | 35 |
| 9 | 41 | 29 | 41 | 25 |
| 10 | 47 | 37 | 45 | 31 |
| 11 | 35 | 27 | 36 | 32 |
| 12 | 38 | 26 | 38 | 26 |
| 13 | 49 | 43 | 47 | 29 |
| 14 | 39 | 19 | 39 | 27 |
| 15 | 38 | 24 | 38 | 30 |

**Figure E.2:** Login page

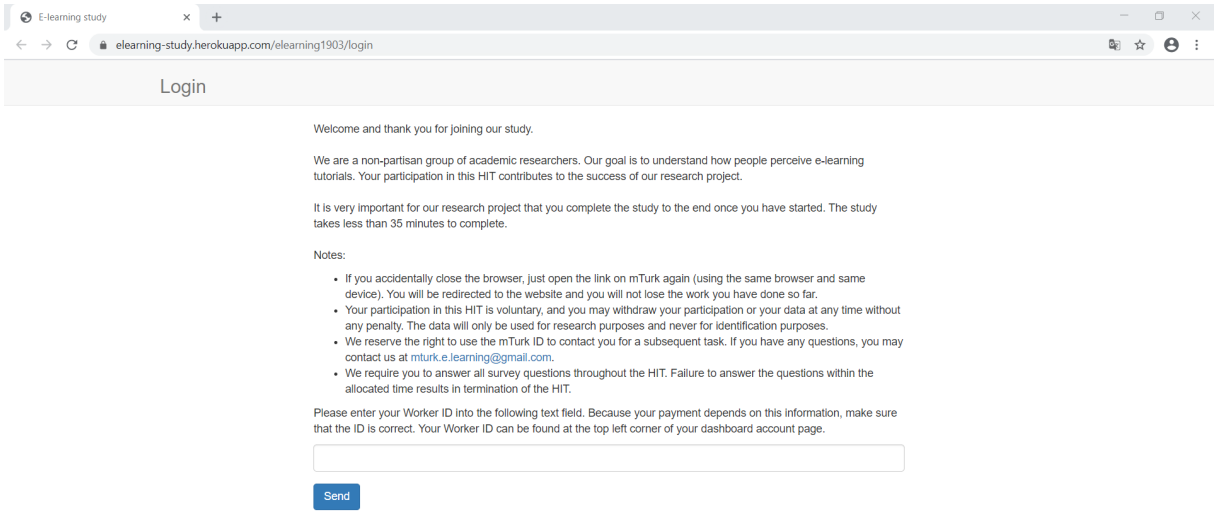



**Figure E.3:** Stratification questions

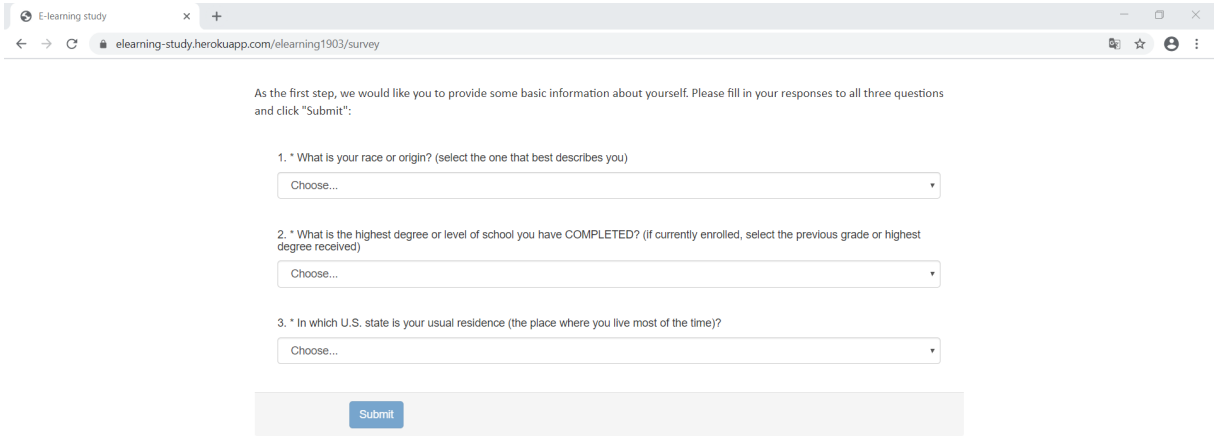


**Figure E.4:** Instructions page

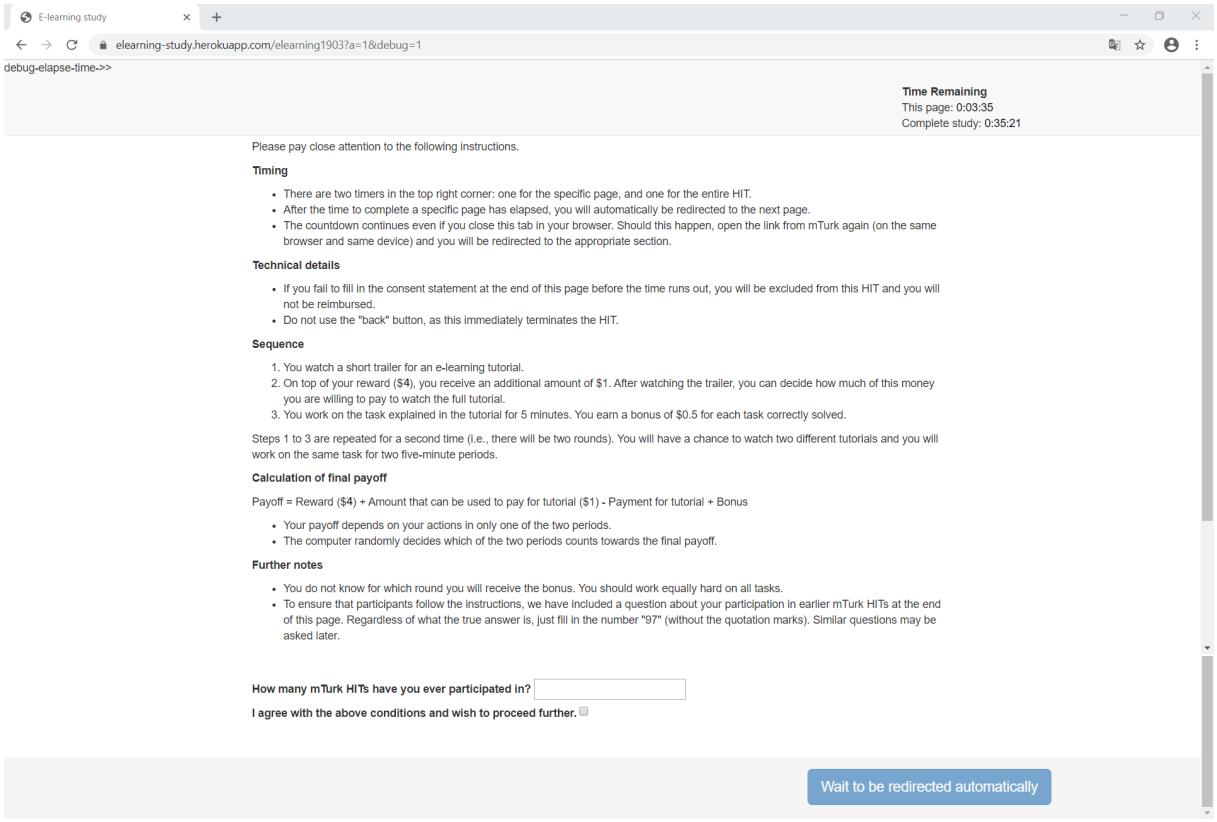



**Figure E.5:** Round 1: Tutorial: hands

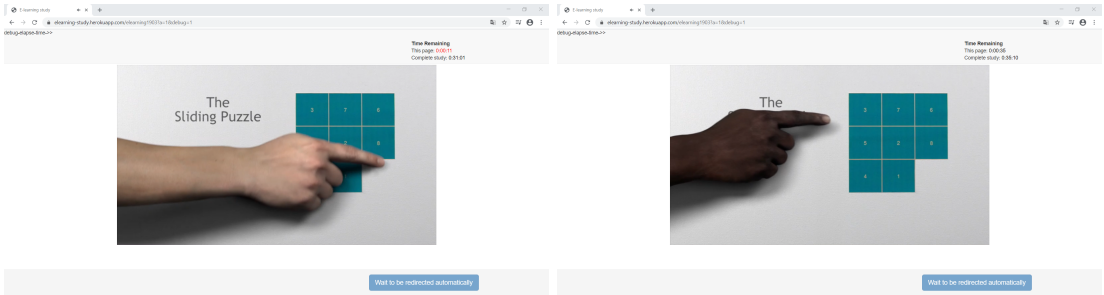


**Figure E.6:** Round 1: Willingness to pay

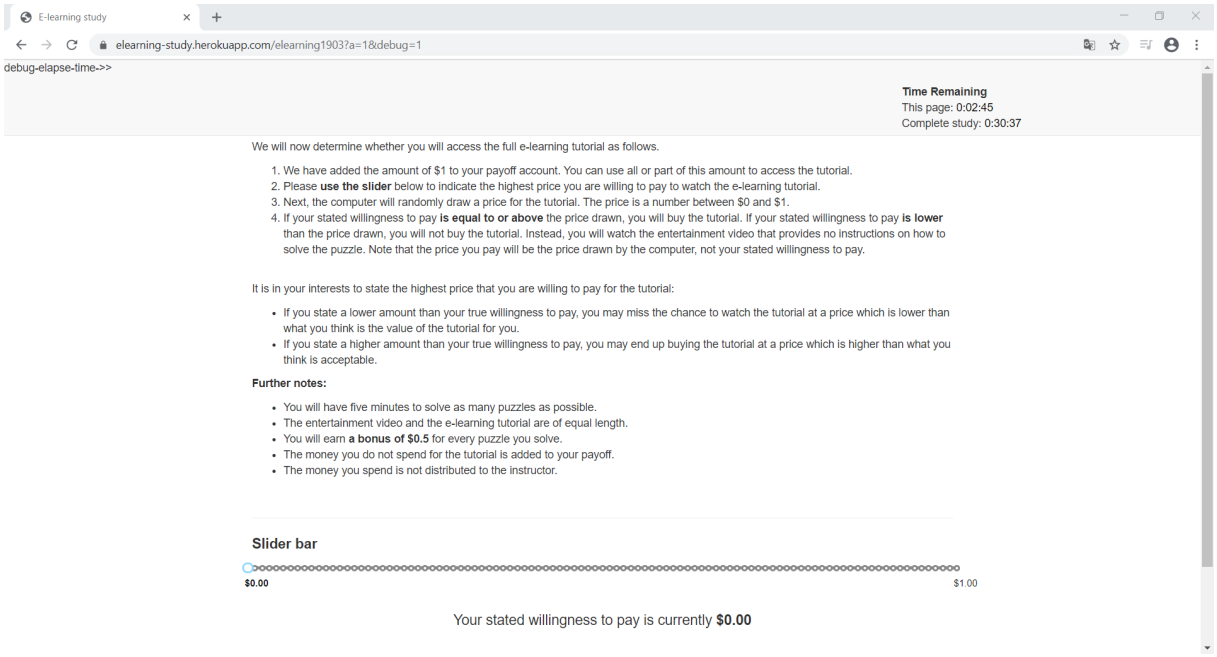



**Figure E.7:** Round 1: Beliefs

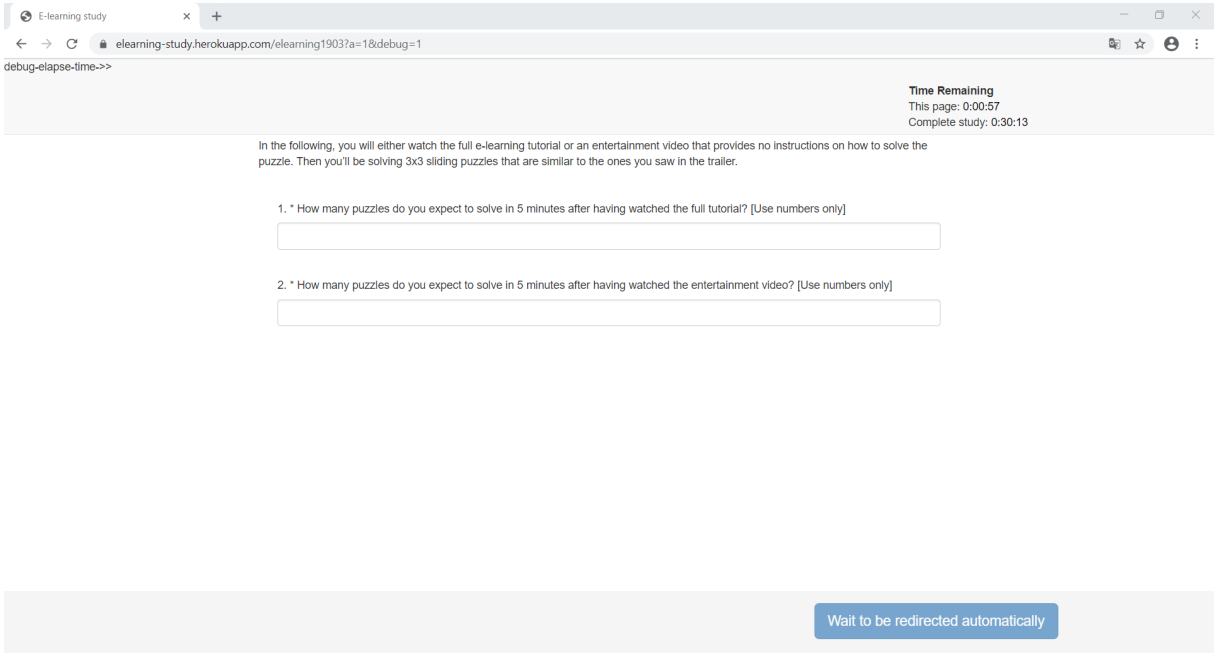


**Figure E.8:** Round 1: Main tutorial

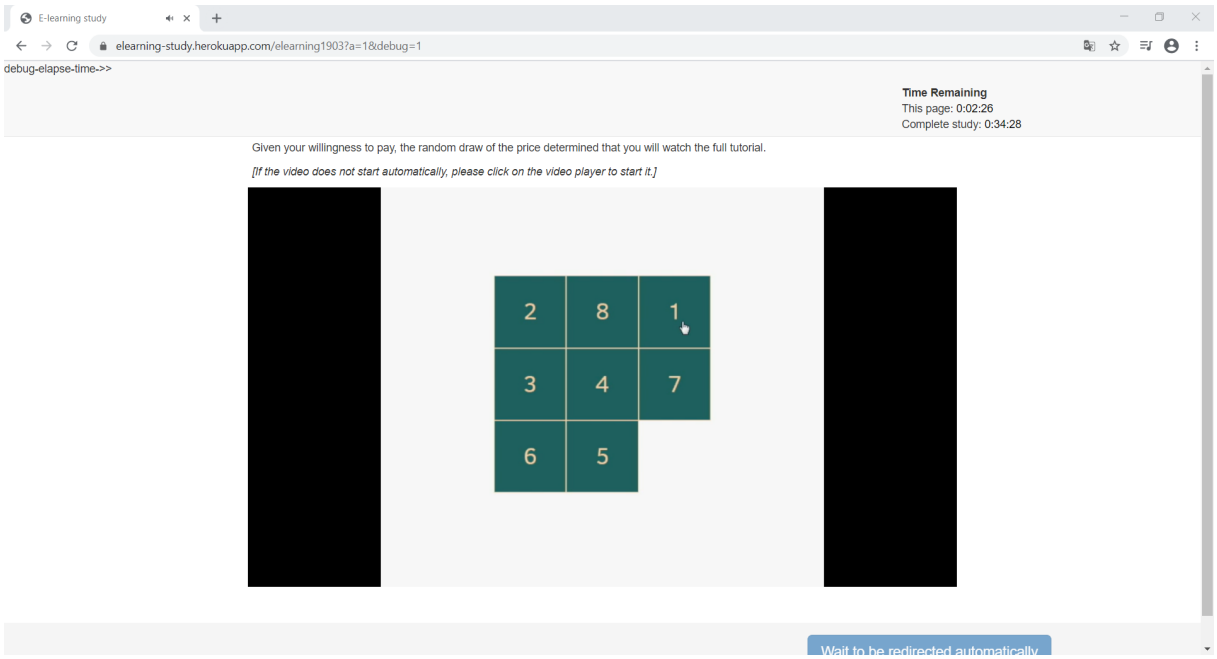


**Figure E.9:** Round 1: Instructor evaluation

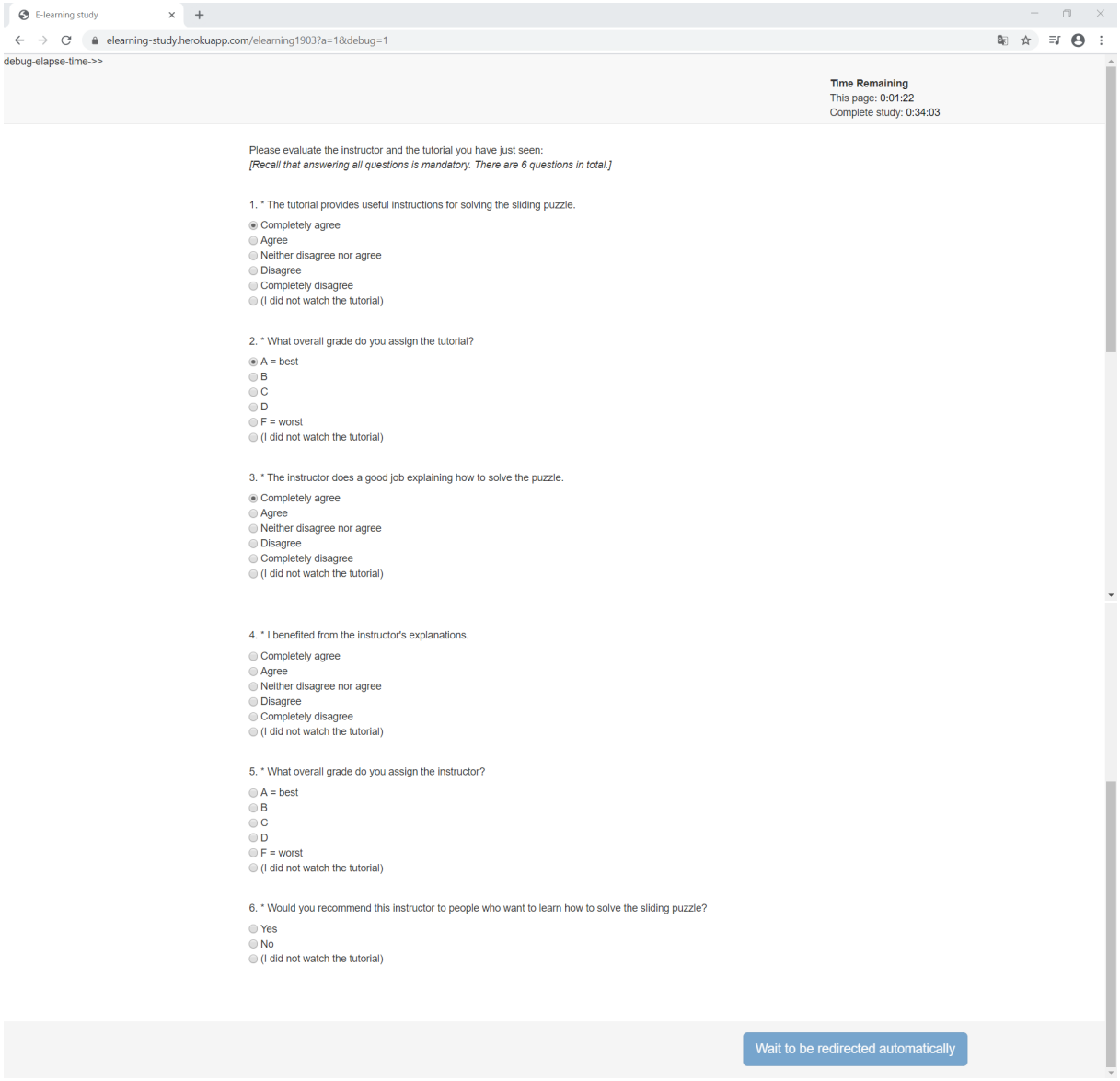


**Figure E.10:** Round 2: Trailer

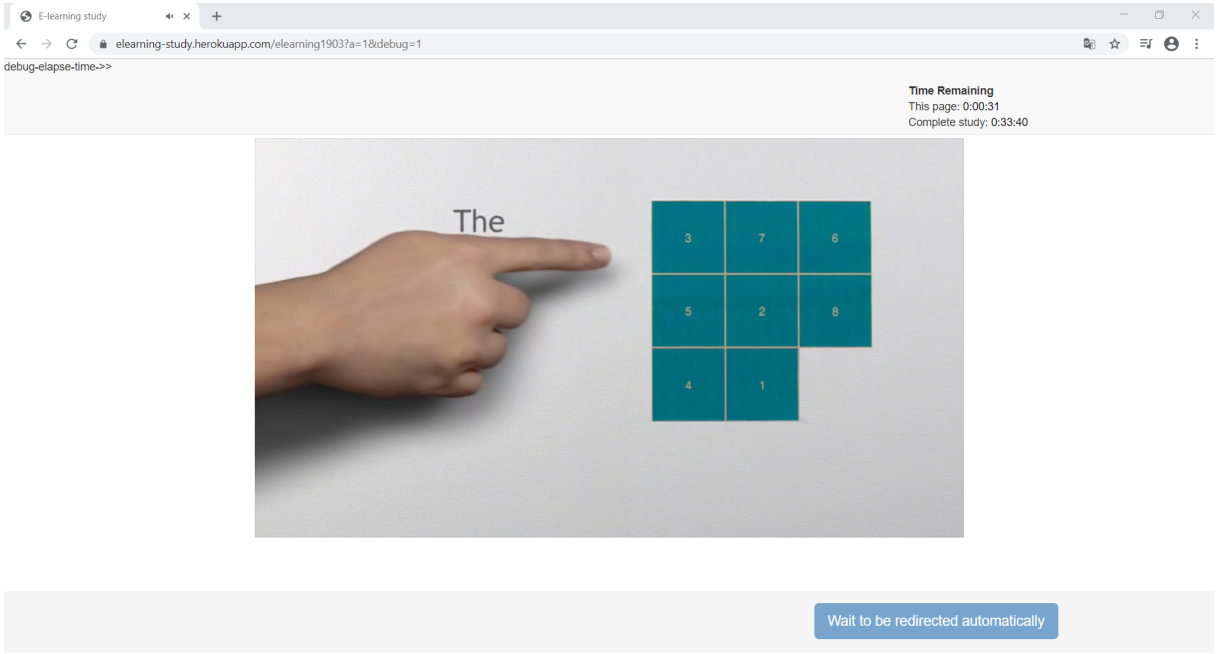
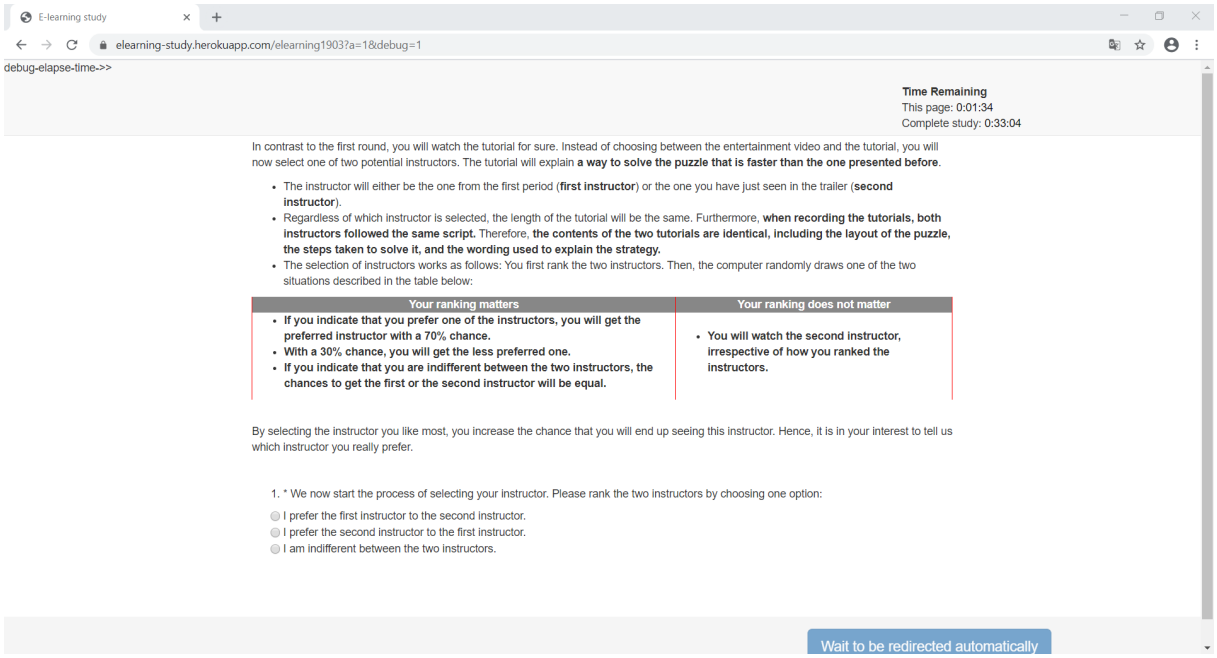



**Figure E.11:** Round 2: Preference ranking




*Note:* Information treatment in the second bullet point in bold. No information treatment leaves this part out.

**Figure E.12:** Round 2: Willingness to pay

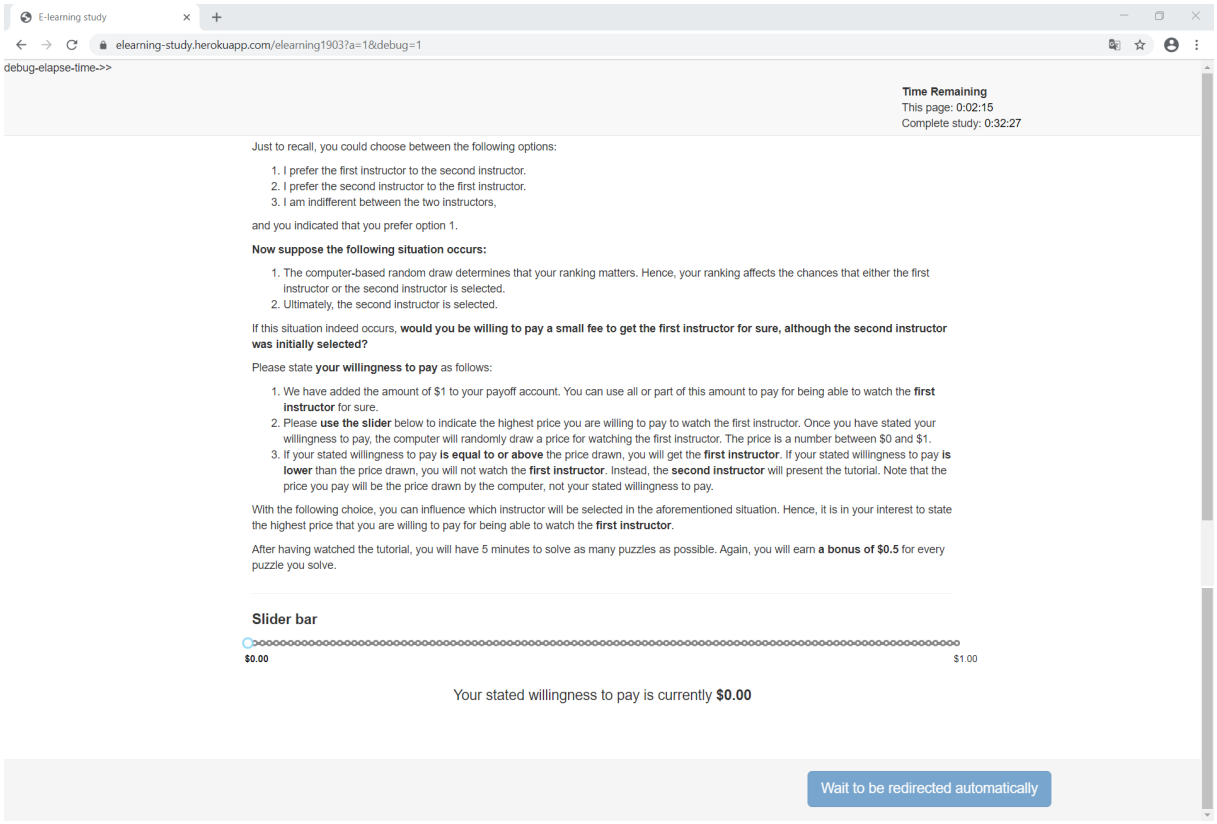



**Figure E.13:** Round 2: Beliefs

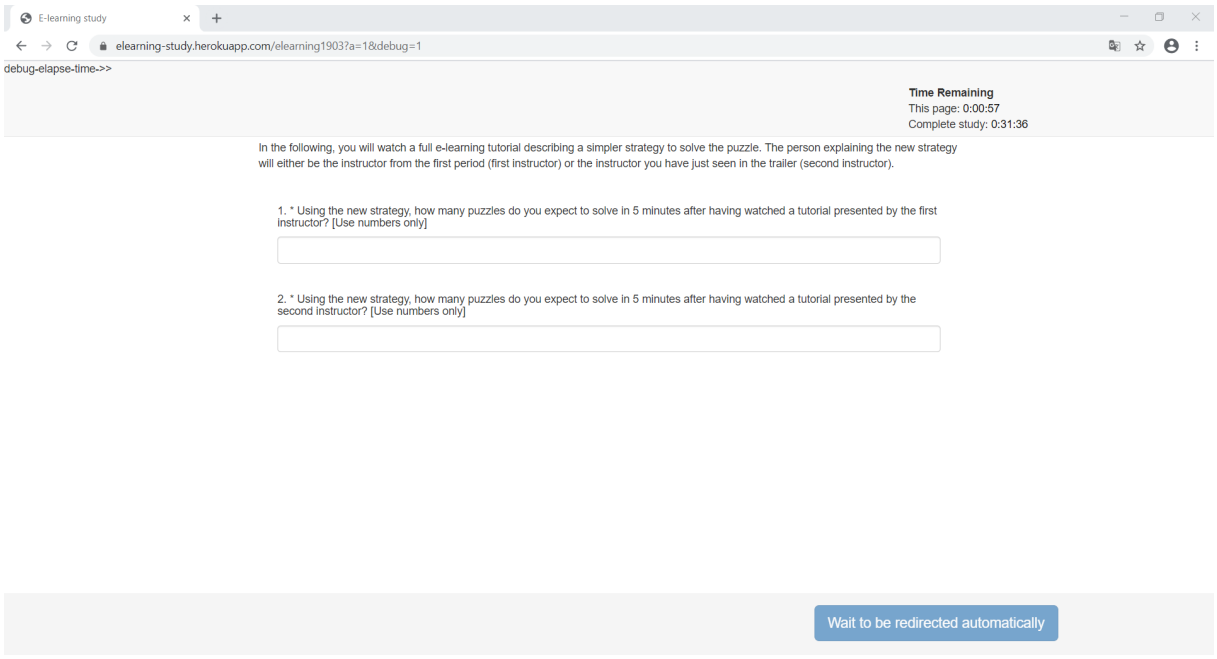


**Figure E.14:** Round 2: Main tutorial

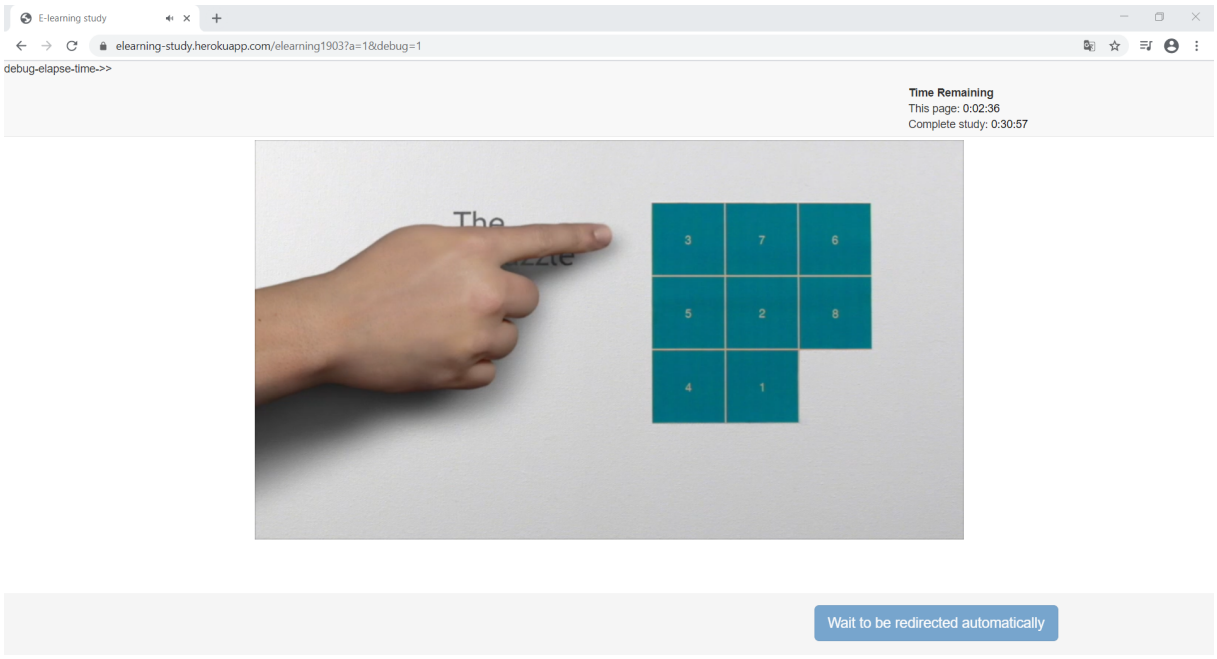


**Figure E.15:** Round 2: Instructor evaluation

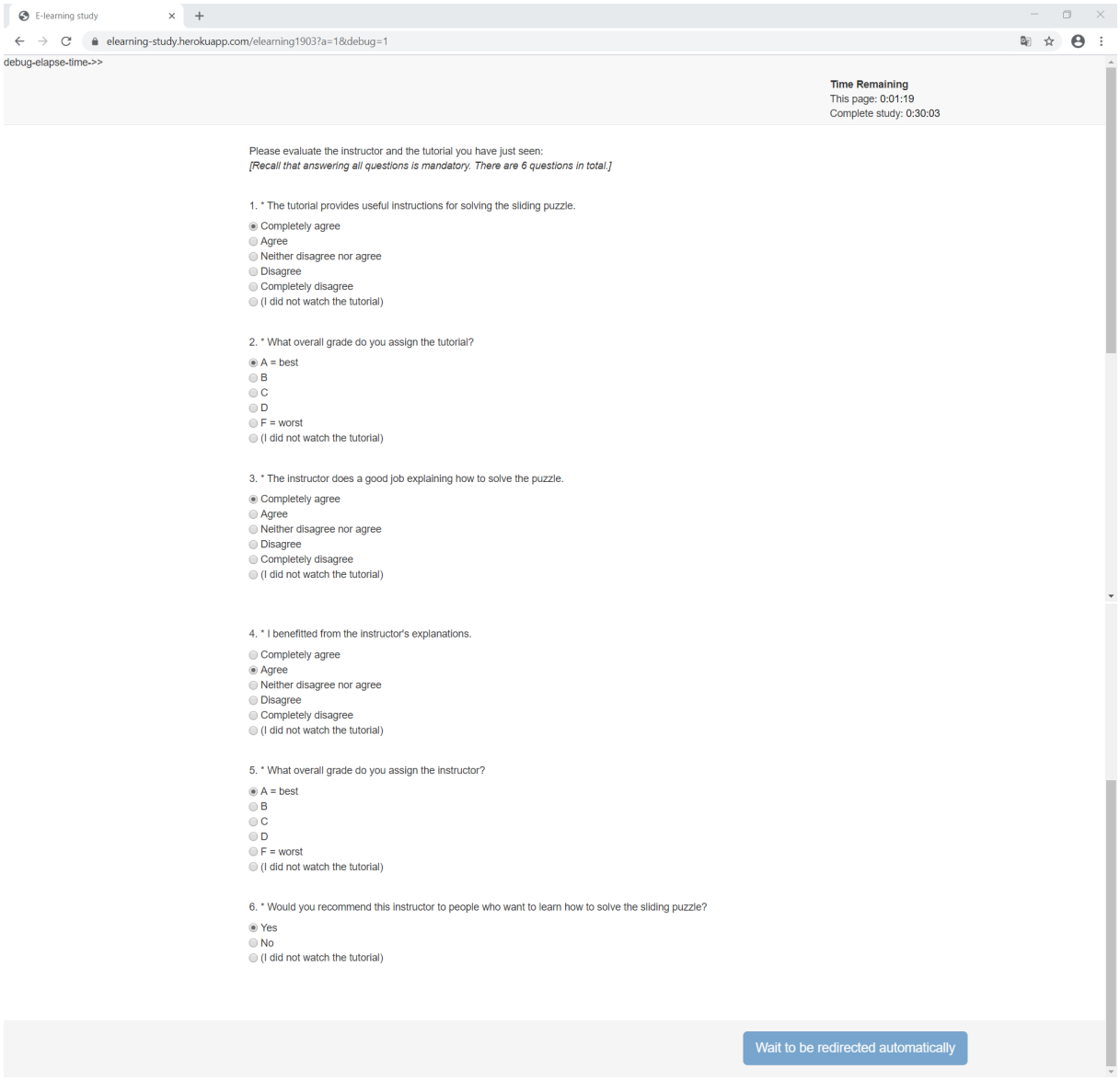
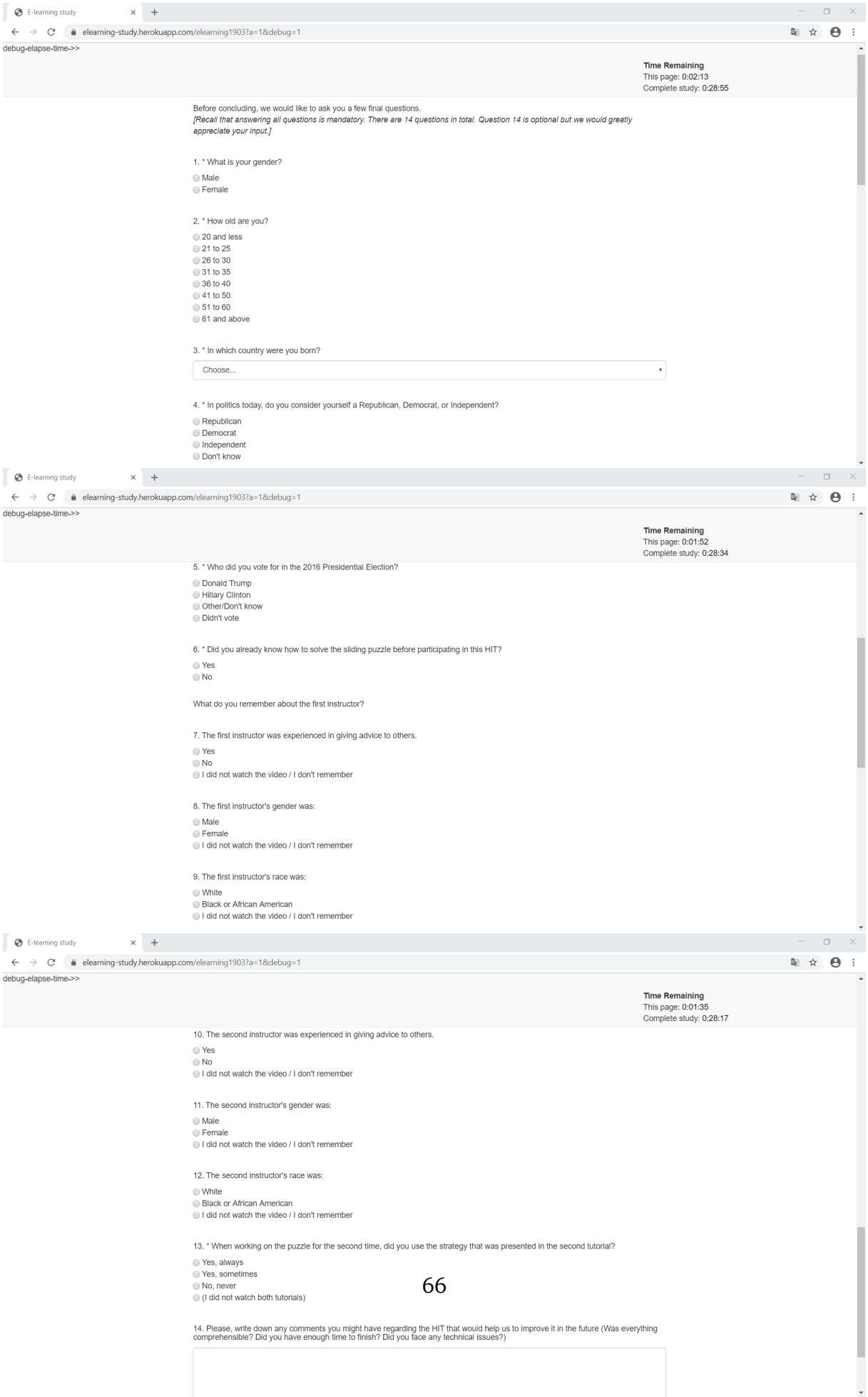


**Figure E.16:** Final survey



# References


ALLISON, P. D. and CHRISTAKIS., N. (1994). Logit models for sets of ranked items. In P. V. Marsden (ed.), *Sociological Methodology*, vol. 24, Oxford: Blackwell, pp. 123–126.

ALSAN, M., GARRICK, O. and GRAZIANI, G. C. (2019). Does Diversity Matter for Health? Experimental Evidence from Oakland. *American Economic Review*, **109** (12), 4071–4111.

AMERICAN SOCIETY OF NEWS EDITORS (2018). The ASNE Newsroom Employment Diversity Survey, available at: https://www.asne.org/diversity-survey-2018.

ARROW, K. (1973). The theory of discrimination. In O. Ashenfelter and A. Rees (eds.), *Discrimination in Labor Markets*, Princeton University Press, pp. 3–33.

ASSOCIATION OF AMERICAN MEDICAL COLLEGES (2014). Diversity in the Physician Workforce: Facts & Figures 2014, available at: http://www.aamcdiversityfactsandfigures.org/.

ATHEY, S. and IMBENS, G. W. (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, vol. 1, Elsevier, pp. 73–140.

AYRES, I. and SIEGELMAN, P. (1995). Race and Gender Discrimination in Bargaining for a New Car. *American Economic Review*, **85** (3), 304–321.

BAKER, R., DEE, T., EVANS, B. and JOHN, J. (2018). Bias in Online Classes: Evidence from a Field Experiment, SIEPR Working Paper No. 18-055.

BARTOS, V., BAUER, M., CHYTILOVA, J. and MATEJKA, F. (2016). Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition. *American Economic Review*, **106** (6), 1437–1475.

BECKER, G. M. and DEGROOT, M. H. (1974). *Measuring Utility by a Single-Response Sequential Method (1964)*, Dordrecht: Springer Netherlands, pp. 317–328.

—, — and MARSCHAK, J. (1964). Measuring Utility by a Single-response Sequential Method. *Behav Sci*, **9** (3), 226–232.

BECKER, G. S. *et al.* (1957). *Economics of discrimination*. University of Chicago Press.

BERTRAND, M. and DUFLO, E. (2017). Field Experiments on Discrimination. In A. Banerjee and E. Duflo (eds.), *Handbook of Field Experiments*, vol. 1, Amsterdam and Oxford: Elsevier, North-Holland, pp. 309–393.

— and MULLAINATHAN, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, **94** (4), 991–1013.

BETTINGER, E. P. and BAKER, R. B. (2014). The Effects of Student Coaching: An Evaluation of a Randomized Experiment in Student Advising. *Educational Evaluation and Policy Analysis*, **36**, 3–19.

BLAU, F. D., CURRIE, J. M., CROSON, R. T. A. and GINTHER, D. K. (2010). Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial. *American Economic Review: Papers & Proceedings*, **100** (2), 348–352.

BOHANNON, J. (2011). Social Science for Pennies. *Science*, **334** (6054), 307.

BRANDTS, J., GROENERT, V. and ROTT, C. (2015). The Impact of Advice on Women's and Men's Selection into Competition. *Management Science*, **61** (5), 1018–1035.

CAGALA, T., GLOGOWSKY, U. and RINCKE, J. (2019). *Who to Target in Fundraising? A Field Experiment on Gift Exchange*. Working Paper mimeo.

CASEY, K., GLENNERSTER, R. and MIGUEL, E. (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *Quarterly Journal of Economics*, **127** (4), 1755–1812.

CELEN, B., KARIV, S. and SCHOTTER, A. (2010). An Experimental Test of Advice and Social Learning. *Management Science*, **56** (10), 1687–1701.

CENTER FOR FINANCIAL PLANNING (2018). Removing Barriers to Racial and Ethnic Diversity in the Financial Planning Profession, available at: https://centerforfinancialplanning.org/initiatives/2018-diversity-summit/racial-and-ethnic-diversity-research/.

DEE, T. S. (2004). Teachers, Race, and Student Achievement in a Randomized Experiment. *Review of Economics and Statistics*, **86** (1), 195–210.

DOLEAC, J. L. and STEIN, L. C. (2013). The Visible Hand: Race and Online Market Outcomes. *Economic Journal*, **123** (572), F469–F492.

EGALITE, A. J., KISIDA, B. and WINTERS, M. A. (2015). Representation in the Classroom: The Effect of Own-Race Teachers on Student Achievement. *Economics of Education Review*, **45**, 44–52.

FISHER, R. A. (1937). *The design of experiments*. Oliver And Boyd; Edinburgh; London.

GREENWALD, A. G., MCGHEE, D. E. and SCHWARTZ, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, **74** (6), 1464.

—, NOSEK, B. A. and BANAJI, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of Personality and Social Psychology*, **85** (2), 197.

HEDEGAARD, M. S. and TYRAN, J.-R. (2018). The Price of Prejudice. *American Economic Journal: Applied Economics*, **10** (1), 40–63.

HEIKENSTEN, E. and ISAKSSON, S. (2018). Simon Says: Examining Gender Differences in Advice Seeking and Influence in the Lab, mimeo, Stockholm School of Economics.

HOFF, K. and PANDEY, P. (2006). Discrimination, Social Identity, and Durable Inequalities. *American Economic Review*, **96** (2), 206–211.

HORTON, J., RAND, D. and ZECKHAUSER, R. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics*, **14** (3), 399–425.

HORTON, J. J. (2011). The condition of the Turking class: Are online employers fair and honest? *Economics Letters*, **111** (1), 10–12.

IMAI, K., KEELE, L., TINGLEY, D. and YAMAMOTO, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, **105** (4), 765–789.

—, KING, G. and STUART, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **171** (2), 481–502.

—, TINGLEY, D. and YAMAMOTO, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **176** (1), 5–51.

IMBENS, G. W. and WOOLDRIDGE, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, **47** (1), 5–86.

KLING, J. R., LIEBMAN, J. B., KATZ, L. F. and SANBONMATSU, L. (2004). *Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health From a Randomized Housing Voucher Experiment*. Working Papers 5, Princeton University, Department of Economics, Industrial Relations Section.

KRAMER, M. M. (2016). Financial Literacy, Confidence and Financial Advice Seeking. *Journal of Economic Behavior & Organization*, **131**, 198–217.

KUZIEMKO, I., NORTON, M. I., SAEZ, E. and STANTCHEVA, S. (2015). How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments. *American Economic Review*, **105** (4), 1478–1508.

LANE, K. A., BANAJI, M. R., NOSEK, B. A. and GREENWALD, A. G. (2007). Understanding and using the implicit association test: Iv. *Implicit measures of attitudes*, pp. 59–102.

LEE, F. (1997). When the Going Gets Tough, Do the Tough Ask for Help? Help Seeking and Power Motivation in Organizations. *Organizational Behavior and Human Decision Processes*, **72** (3), 336–363.

LIST, J. A., SHAIKH, A. M. and XU, Y. (2016). Multiple hypothesis testing in experimental economics. *Experimental Economics*, pp. 1–21.

LOWES, S., NUNN, N., ROBINSON, J. A. and WEIGEL, J. (2015). Understanding ethnic identity in africa: Evidence from the implicit association test (iat). *American Economic Review*, **105** (5), 340–45.

LUSHER, L., CAMPBELL, D. and CARRELL, S. (2016). TAs Like Me: Racial Interactions Between Graduate Teaching Assistants and Undergraduates. *Journal of Public Economics*, **159**, 203–224.

MENGEL, F., SAUERMANN, J. and ZÖLITZ, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, **17** (2), 535–566.

MOBIUS, M. M. and ROSENBLAT, T. S. (2006). Why Beauty Matters. *American Economic Review*, **96** (1), 222–235.

NEUMARK, D. (2018). Experimental Research on Labor Market Discrimination. *Journal of Economic Literature*, **56** (3), 799–866.

—, BANK, R. J. and VAN NORT, K. D. (1996). Sex Discrimination in Restaurant Hiring: An Audit Study. *Quarterly Journal of Economics*, **111** (3), 915–941.

OPPENHEIMER, D. M., MEYVIS, T. and DAVIDENKO, N. (2009). Reports. *Journal of Experimental Social Psychology*, **45** (4), 867–872.

OSTER, E., SHOULSON, I. and DORSEY, E. R. (2016). Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease: Corrigendum. *American Economic Review*, **106** (6), 1562–1565.

PAPKE, L. E. and WOOLDRIDGE, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, **11** (6), 619–632.

PEW RESEARCH CENTER (2016). Research in the crowdsourcing age, a case study.

PHELPS, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, **62** (4), 659–661.

RAO, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*, **109** (3), 774–809.

ROSENBAUM, P. R. and RUBIN, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, **39** (1), 33–38.

SCHOTTER, A. (2003). Decision Making with Naive Advice. *American Economic Review*, **93** (2), 196–201.

STEWART, N., UNGEMACH, C., HARRIS, A. J., BARTELS, D. M., NEWELL, B. R., PAOLACCI, G. and CHANDLER, J. (2015). The average laboratory samples a population of 7,300 amazon mechanical turk workers. *Judgment and Decision Making*, **10** (5), 479–492.

TOUSSAERT, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica*, **86** (3), 859–889.

WHITEHEAD, J. (1993). Sample size calculations for ordered categorical data. *Statistics in Medicine*, **12** (24), 2257–2271.