The Color of Knowledge: Impacts of Tutor Race on Learning and Performance

Vojtěch Bartoš and Ulrich Glogowsky and Johannes Rincke*

May 21, 2025

Abstract

We demonstrate that racial biases against tutors hinder learning. In e-learning experiments, U.S. conservatives are more likely to disregard advice from Black tutors, resulting in reduced performance compared to learners taught by white tutors. We show that the bias is unconscious and, consequently, does not skew tutor selection. In line with our theory, the bias disappears when the stakes are high. In contrast, liberals favor Black tutors without experiencing learning disparities. Methodologically, we contribute by using video post-production techniques to manipulate tutor race without introducing typical confounds. Additionally, we develop a novel two-stage design that simultaneously measures tutor selection, learning, and productivity.

Keywords: Discrimination; racial bias; advice-seeking; online experiment

*Bartoš: University of Milan, Department of Economics, Management, and Quantitative Methods, Via Conservatorio 7, 20122 Milano, Italy, vojtech.bartos@unimi.it. Glogowsky: Johannes Kepler University Linz, Department of Economics, Altenbergerstrasse 69, 4040 Linz, Austria, ulrich.glogowsky@jku.at. Rincke: University of Erlangen-Nuremberg, School of Business, Economics, and Society, Lange Gasse 20, 90403 Nuremberg, Germany, johannes.rincke@fau.de. The authors would like to thank Michal Bauer, Stefan Bauernschuster, Tobias Cagala, Alexander Cappelen, Julie Chytilová, Hajdu Gergely, Matthias Fahn, Susann Fiedler, Alex Imas, Yves Le Yaouanq, Jakob Möller, Elizabeth Levy Paluck, Alexander Todorov, Ricardo Perez-Truglia, and Bertil Tungodden for helpful comments. Seminar participants at various places provided valuable feedback. We gratefully acknowledge funding from the German Research Foundation through a research grant (project 444754857) and through CRC TRR 190 (project 280092119). Bartoš acknowledges financial support within the "Fund for Departments of Excellence academic funding" provided by the Ministero dell'Universita e della Ricerca (MUR), established by Stability Law, namely "Legge di Stabilita n.232/2016, 2017" - Project of the Department of Economics, Management, and Quantitative Methods, University of Milan. Glogowsky gratefully acknowledges financial support from the Linz Institute for Transformative Change (C-2022-1-118). This experiment was pre-registered at the AEA Social Science Registry as AEARCTR-0005812. Updates: AEARCTR-0007737, AEARCTR-0008563, AEARCTR-0009715. We obtained IRB approval from the Ethics Commission of the School of Business and Economics at the University of Erlangen-Nuremberg. Ülkü Bıçakçı, Irina Fux, and David Gstrein provided valuable research assistance.

1 Introduction

Learning from others is essential for individual performance and success. It enhances human capital (Acemoglu and Pischke, 1998; Arrow, 1962; Ben-Porath, 1967), drives innovation (Bell *et al.*, 2019; Bloom *et al.*, 2012; Mokyr, 2011), and creates economic opportunity (Chetty *et al.*, 2011, 2014, 2020). Yet, the learning process frequently involves many complex choices. Learners must decide (a) whom to select as an advisor, (b) whether to trust and acquire the knowledge offered, and (c) whether to utilize the gained knowledge in practice. In today's digital age, the abundance of available information complicates these decisions, especially when the advisor's quality is unclear. Learners may then rely on visible advisor characteristics, such as race or gender, to guide their decisions (Arrow, 1973; Becker, 1957; Nosek *et al.*, 2007, 2009; Phelps, 1972). Hereby, they invite (conscious or unconscious) biases into the learning process. As a result, learners may overlook skilled minority advisors or disregard sound advice when engaging with them. However, we know little about whether discrimination against advisors exists or how it shapes learners' success.[1]

This paper leverages a large-scale e-learning experiment to provide the first systematic evidence of racial discrimination by learners against advisers. Our key contribution is to uncover how discrimination shapes core decisions and outcomes throughout the whole learning process. We study whether learners select white over Black tutors to occupy advisory roles, whether they are less likely to acquire and utilize the knowledge provided by Black tutors, and how discrimination translates into performance gaps. Our study also breaks new ground by identifying the drivers and dimensions of discriminatory behavior. First, we disentangle whether learners discriminate against Black tutors due to the belief that they are less effective (statistical discrimination), conscious aversion (taste-based discrimination), or unconscious bias in learning (unconscious discrimination). Second, motivated by recent public debates and evidence suggesting that racism and bias are more prevalent among people with conservative political views,[2] we examine if discrimination varies by political orientation.

Together, our results provide a comprehensive understanding of how discrimination

[1] There is some suggestive evidence in line with discrimination against advisors. For example, white boys perform better in tests when taught by white rather than Black teachers (e.g. Dee, 2004; Egalite *et al.*, 2015; Lusher *et al.*, 2016). However, this effect may be (partly) due to white teachers treating white boys preferentially rather than discrimination in learning. Studying a different domain, DiBartolomeo *et al.* (2023) suggests that people demand services from Black financial advisers less often.

[2] Conservative social media users have been shown to share more misinformation, including racially insensitive content (Mosleh *et al.*, 2024). There is also evidence suggesting systematic associations between conservative political views and higher levels of implicit and explicit racial bias (Nosek *et al.*, 2007, 2009; Pew Research Center, 2021; Coutts, 2023; Haaland and Roth, 2023).

manifests and impacts learning. One of our main findings is that participants engage in unconscious discriminatory behavior. They more often reject knowledge from Black than white tutors and perform worse with Black tutors without being aware of these behaviors. Notably, this tendency is limited to conservatives. By contrast, liberals do not discriminate in learning. If anything, they show Black-favoring behavior when choosing between advisors. We offer a theoretical framework explaining these findings.

Design: To study discrimination in learning, we designed our study to achieve three objectives: (a) tracking core decisions and choices throughout the learning process, (b) isolating the causal effects of race, and (c) identifying types of discrimination.

We achieve the *first goal* with a two-stage experiment that tracks the entire learning process. Specifically, we use data on more than 2,400 U.S. residents recruited via CloudResearch for a study on "perceptions of e-learning materials." In the first stage, participants watch a trailer of an e-learning tutorial introducing the central task of our study: the *sliding tile puzzle* (Isaksson, 2018; Kinnl *et al.*, 2023).[3] The trailer features a tutor's hand, randomized to appear as Black or white. Participants then decide whether to purchase access to the full tutorial explaining a distinct solution strategy. A truthful incentive-compatible mechanism ensures that most participants receive access regardless of their purchasing decision.[4] After watching the full tutorial, they work on solving puzzles for five minutes, earning a piece rate. The second stage introduces an explicit selection decision. Participants choose between a Black and white tutor to learn a faster strategy for solving the puzzle and complete another puzzle-solving session. In sum, our design measures tutor-selection decisions (e.g., purchasing tutorials and choosing between tutors) and performance (e.g., solving puzzles). Moreover, the sliding tile puzzle allows us to monitor knowledge acquisition and utilization. This is because specific sequences of tile moves reveal the strategies participants use to solve the task. Thus, we can track whether participants solve tasks using the strategies taught in the tutorials, capturing both their ability to acquire knowledge and their willingness to apply it.[5]

To achieve the *second goal of isolating the causal effects* of race, our design needs to

[3] The puzzle requires participants to arrange tiles numbered one to eight in ascending order on a three-by-three grid, with one empty space to allow movement.

[4] We use a truthful BDM procedure to model the purchasing decision. Participants state their willingness to pay (*WTP*) for the tutorial. They access the tutorial if their *WTP* equals or exceeds a random price. To ensure most participants get access, we set the price to zero with a 95% probability.

[5] In our setting, selecting one tutor over another or performing better with one instructor does not harm the tutors involved. We, nevertheless, refer to such differences as *discriminatory* behaviors on the basis that classic definitions of discrimination do not necessarily imply discriminated individuals to be harmed. Taste-based discrimination arises, for example, from personal aversions against minorities, leading individuals to avoid certain groups even at their own disadvantage. Statistical discrimination, instead, relies on group-based beliefs. Both types do not require individuals to act on malice.

address two challenges. The first is manipulating the tutor's race without introducing confounders (e.g., differences in teaching style or quality). We hired Hispanic actors to record the tutorials. A Hollywood Visual Effects Artist with experience in major films then adjusted their skin tones to Black or white using industry-standard post-production techniques. The second challenge is measuring causal effects at multiple margins simultaneously. Participants' tutor choices can, in principle, create self-selection biases that complicate causal inference for subsequent outcomes (performance and strategy use). For example, if participants consistently chose white tutors, we could not observe their performance under Black tutors. We address self-selection by focusing on the subset of participants that our incentive-compatible mechanism randomly assigns to Black or white tutors (regardless of their purchasing decision or tutor preference).

To address the *third goal of identifying types of discrimination*, we propose a simple conceptual framework to guide our analysis. The model predicts how statistical, taste-based, and unconscious discriminators behave in our experiment. The predictions for unconscious discrimination are unique: Unlike the other types, they, for example, do not expect to learn less from Black tutors and, thus, do not choose white over Black ones. Differentiating statistical from taste-based discrimination is more challenging. For that purpose, we introduce an information treatment clarifying that both second-stage tutors follow the same script and, thus, provide identical instructions. This step aims to eliminate perceived differences in tutor quality, allowing us to separate statistical from taste-based discriminators.

Results: Our first set of findings focuses on participants with conservative political views.[6] Conservatives show *unconscious discriminatory behavior* in the learning process against Black tutors. They are less likely to learn from and utilize the knowledge provided by Black (rather than white) tutors. For example, the joint probability of learning and adopting a puzzle-solving strategy explained by the tutor is around 13.4% (or 8.4 percentage points) lower when the tutor is Black instead of white. This behavior translates into adverse performance effects: The number of puzzles conservatives solve drops by 18.8% with Black tutors. Consistent with the theory of unconscious discrimination, conservatives do not anticipate learning and performing worse under Black tutors. Acting on this belief, they also do not choose white over Black tutors. In additional analyses, we examine how discriminators respond to changes in the cost of discrimination. For this purpose, we introduce a high-stakes (cross-cutting)

[6]Studying discrimination by political orientation is challenging. Eliciting political views can, for example, prompt participants to alter their behavior to appear socially acceptable. To mitigate such challenges, we collaborated with a survey provider that pre-screens participants by political orientation. This method allows us to target conservatives and liberals explicitly. We pre-registered this procedure.

treatment that increases the piece rate for solved puzzles by 400%.[7] Thus, this treatment substantially raises the cost of rejecting knowledge from tutors, making it more costly for participants to act on unconscious biases. As predicted by our model, unconscious discrimination vanishes in such a high-stakes setting. This finding aligns with the notion that higher stakes encourage reflective, deliberate thinking and, hereby, help people overcome unconscious biases (Kahneman, 2003; Bertrand *et al.*, 2005; Kahneman, 2011).

The second set of results focuses on participants with liberal political views. Their behavior contrasts sharply with that of conservatives: Liberals do not discriminate when learning from Black tutors but rather exhibit *Black-favoring behavior* (independently of the piece rate). Specifically, liberals learn and apply strategies equally well with Black and white tutors, and their performance does consequently not depend on the tutor's race. They also do not anticipate performance differences across tutors. However, in the second stage, learners with very liberal political views choose Black tutors over white ones. This behavior aligns with a version of our model where learners (a) do not expect Black tutors to be better teachers, (b) are indifferent between actually learning from Black or white tutors, but (c) derive non-instrumental utility from selecting Black tutors. Motives such as a desire for inclusivity or virtue signaling may explain why liberals derive satisfaction from this choice.

Lastly, the information treatment does not affect the behavior of either conservatives or liberals. This finding is consistent with the fact that neither group expects performance differences between Black and white tutors.

Literature: Our work contributes to several literature strands. The first strand examines the role of race in advisor-learner relationships. Previous research shows that Black students benefit from Black teachers (Dee, 2004; Egalite *et al.*, 2015), Black mentees from Black mentors (Kofoed and mcGovney, 2019), and Black patients from Black doctors (Alsan *et al.*, 2019). These studies, therefore, focus on the positive aspects of homophily. For example, they highlight that shared racial identity can foster trust, better communication, and improved outcomes. Our study, instead, highlights that minority advisors may face challenges due to learner biases and the consequences of these biases for learners themselves. Hence, we take a new perspective by showing that biases shape learning and advisor-learner relationships.

The second related strand is experimental work on racial discrimination, highlighting the barriers racial minorities face across many domains. In the following, we highlight the most relevant studies in this extensive literature and direct readers to existing reviews for further details (Lang and Lehmann, 2012; Bertrand and Duflo, 2017;

[7] Here, participants' hourly earnings were about 90% higher than their usual rate on the platform.

Neumark, 2018; Lang and Spitzer, 2020; Onuchic, 2023). A study particularly close to ours is Bartos *et al.* (2016) who examine attention discrimination. They use field experiments with fictitious résumés and personal websites to track how employers and landlords acquire information about applicants. Like them, we show that biases can influence the demand for information. However, we extend the scope of the analysis by highlighting how people process and utilize information available to them. This extended focus pinpoints a critical consequence of discrimination: it leads to higher-level productivity losses as discriminators reject useful information. Additionally, our work highlights unconscious biases as an essential part of discriminatory behaviors in information demand. Another closely related paper is Evsyukova *et al.* (2025). They employ an innovative design based on A.I.-morphed profile pictures on LinkedIn to examine discrimination against advice seekers. Specifically, they examine if Black or white men (advisees) are more likely to receive advice from users (advisors) within their LinkedIn network. Instead, we examine if advisees discriminate against advisors, thereby reversing the direction of discrimination. We also build on Hedegaard and Tyran (2018), who show that ethnic workplace discrimination depends on the price of prejudice. We extend this concept to different forms of discrimination in the learning context. Moreover, we use a design that avoids confounding effects from production complementarities between discriminators and minorities. Finally, Doleac and Stein (2013) use hand images to distinguish Black and white sellers in online markets. We build on their design and manipulate the skin color with post-production techniques. From a broader perspective, no single study explored discrimination in learning. We address this gap with a comprehensive experiment that analyzes all stages of the learning process and disentangles different types of discriminators.

The third strand we contribute to is the methodological literature on designing experiments to study racial discrimination. Most existing work relies on the correspondence methodology (Bertrand and Mullainathan, 2004; Bertrand and Duflo, 2017; Neumark, 2018). Correspondence studies typically measure discrimination by sending fictitious applications with randomized characteristics (such as race or gender) to real-world employers. While the approach has been instrumental in measuring the extent of discrimination, it also has limitations (Bertrand and Duflo, 2017). One of the drawbacks is that typical outcomes, such as callback rates, offer only coarse discrimination measures. Typical correspondence studies also do not allow researchers to study longer-term interactions or productivity. Moreover, they often signal race through names, which can inadvertently evoke stereotypes unrelated to race itself (Kreisman and Smith, 2023). Our study introduces new tools to address these issues and limitations. We signal race in video tutorials and, hereby, avoid unwanted confounders. Our study also proposes methods for simultaneous causal estimation of tutor selection and learning and measures an exceptionally rich set of outcomes (including productivity). Finally, we propose new ways to disentangle different types of discrimination, a notoriously challenging task in correspondence studies.[8] Together, these methodological innovations allow us to demonstrate that biases against Black tutors undermine learning and productivity.

We organize the paper as follows. Section 2 describes the experimental design. Section 3 presents conceptual considerations. Sections 4 and 5 discuss results for conservatives and liberals, respectively, and Section 6 concludes.

2 Experimental Design

2.1 Overview

This subsection summarizes our design, while subsequent subsections provide details. Appendix C contains the complete set of instructions, and Appendix D provides additional background on the preregistration.

Recruitment: We conducted the experiment through CloudResearch, a widely used platform for online recruitment known for its high data quality compared to other providers (Hauser *et al.*, 2023; Stagnaro *et al.*, 2024). This platform offers access to diverse participant pools, customizable survey tools, and advanced data collection features. It also pre-collects and provides data on many variables (such as political orientation). This feature allows us to target our study to conservative and liberal Americans, without revealing our interest in this variable.

Summary of design: Our goal is to study how discrimination affects learners' (a) choices of tutors, (b) decisions to acquire and use knowledge, (c) performance, and (d) beliefs. To that end, we invited CloudResearch users to a study on "how people perceive e-learning materials,"[9] without mentioning our interest in discriminatory behavior. In the experiment, learners buy and watch e-learning tutorials offered by

[8] Correspondence studies, for example, attempt to separate taste-based and statistical discrimination by adding numerous productivity-related characteristics to résumés. The aim is to control for all factors influencing perceptions of applicant quality. However, designing résumés that capture every relevant characteristic is highly challenging. Moreover, even when qualifications are held constant, decision-makers may rely on name-related stereotypes to fill perceived gaps in the applicant's profile. Our approach takes a more direct route: Instead of attempting to manage expectations about the tutors' quality by providing information about their characteristics, we explicitly inform people in the second stage that both tutors provide identical instructions and even use the same wording.

[9] The invitation text read: "This study is about how people perceive e-learning tutorials. You watch tutorials, play games, and answer a few questions. In the past, 95% of our participants earned a bonus."

Figure 1: Timeline of the First Stage

Step 1: Learners watch the trailer of an e-learning tutorial. The trailer reveals the tutor's skin color and introduces the sliding tile puzzle

↓

Step 2: Learners state belief about their performance

↓

Step 3: Learners make a purchasing decision about the full tutorial

↓

Step 4: Most learners watch the full e-learning tutorial (self-selection prevented by design)

↓

Step 5: Learners work on the task (piece rate)

randomly selected tutors that are either Black or white. These tutorials teach strategies for solving the sliding tile puzzle. Learners then solve puzzles for a randomly chosen piece rate, incentivizing them to acquire and apply knowledge. The design also features an information treatment to separate taste-based and statistical discrimination.

Summary of timeline: The timeline of our experiment is as follows: After logging in and completing a brief demographic survey for stratification purposes, learners receive information about the general procedures and the timing of the experiment, which consists of two stages.

Figure 1 outlines the structure of the first stage. In *Step 1*, learners watch a trailer of the full e-learning tutorial, which they can purchase later. The trailer has two purposes: it introduces the sliding tile puzzle and reveals the tutor's race by showing his hand during explanations. Specifically, a *first-stage tutor* presents the puzzle, states its objective, and mentions that the full tutorial provides an effective solving strategy *(first strategy)*.[10] Importantly, the trailer does not include any solution strategies. *Step 2* elicits learners' beliefs about their performance both (a) after having watched the tutorial and (b) without tutorial exposure. Next, *Step 3* implements a purchasing decision to measure how strongly learners value the tutorial. Specifically, learners state their willingness to pay (*WTP*) for the full tutorial, which influences their chance of accessing it in *Step 4*. Those with a *WTP* above a randomly drawn price gain access. To ensure most learners watch the tutorial regardless of their valuation, 95% randomly receive a price of zero. Therefore, among this group, we can causally identify how the tutor's race affects learning (as tutor assignment is random). In *Step 5*, learners solve sliding tile puzzles for five minutes.

[10] He states that the full tutorial presents one of the "easiest and fastest" ways to solve the puzzle.

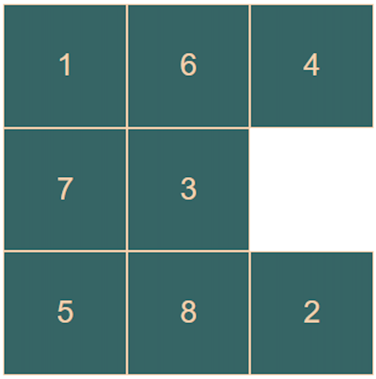
The structure of the second stage is similar to that of the first one. Again, learners first watch a trailer (*Step 1*) and later a full tutorial (*Step 4*). They also get a second chance to work on a series of puzzles (*Step 5*). However, there are six key differences to the first stage. First, a new *second-stage tutor*, who differs from the first one, presents the second-stage trailer. Second, the full tutorial in the second stage presents a (truthfully) more efficient (i.e., faster) strategy to solve the puzzle (*second strategy*). Participants also learn in the trailer that the second strategy is faster than the first one, giving them an incentive to learn it. Third, instead of choosing whether to buy the full tutorial (*Step 3*), learners now select a tutor who presents the second strategy in the full tutorial. They can choose between the first-stage or second-stage tutor. Learners, hence, watch the second tutorial for sure but choose who presents it. Fourth, using a 2×2 experimental design, we randomize the tutors' skin colors across both stages. Consequently, about half of the learners face tutors of different skin colors in both stages. This feature allows us to determine if learners deliberately choose Black over white tutors when they have a choice (while preserving plausible deniability). Fifth, we elicit learners' beliefs about their performance under the first-stage and second-stage tutors in stage 2. Sixth, we implement an orthogonal information treatment. Before selecting a tutor, participants in this treatment learn that both tutors present precisely the same strategy using the same script. This treatment shuts down residual information asymmetries and allows us to identify the nature of discrimination (belief-based or not).

Design goals: Our two-stage design achieves the goals outlined in the introduction. Most importantly, the e-learning context facilitates causal identification through the use of video. Specifically, we (a) cleanly manipulate the tutor's race with post-production techniques, (b) randomly assign learners to tutors, and (c) let learners naturally observe the tutor's skin color in tutorials. We never mention the tutor's race to minimize experiment demand. The design also allows us to achieve our second goal: capturing all stages of the learning process. As detailed in Subsection 2.4, this includes measuring tutor selection, advice acquisition and utilization, and performance. Finally, the setting enables us to study different forms of discrimination, our third goal. To that end, we assess multiple outcomes (e.g., beliefs) and incorporate additional treatments (e.g., the information treatment).

2.2 Sliding Tile Puzzle

The task: The sliding tile puzzle tests the participant's ability to apply spatial reasoning and plan moves strategically (Isaksson, 2018; Kinnl *et al.*, 2023). It consists of eight tiles labeled one to eight and one empty space (see Figure 2). Learners slide

Figure 2: Sliding Tile Puzzle



adjacent tiles into the empty space to arrange them numerically (first row: 1, 2, 3; second row: 4, 5, 6; third row: 7, 8, blank). In each stage, learners solve as many puzzles as possible in five minutes. They earn a piece rate for each puzzle completed (see Subsection 2.3).

Benefits of this task: The central benefits of the sliding tile puzzle are that it (a) enables measurement of learners' performance and (b) allows us to assess whether they acquire and utilize advice (see Subsection 2.4 for details). Another important feature is that learners benefit from the instructions provided (they solve more puzzles after watching the tutorials).[11] Resultingly, rejecting knowledge due to discriminatory behavior should affect their performance.

Beyond these core advantages, the task offers additional benefits. First, it is simple in appearance, its goal is easy to understand, and it has a unique solution. Second, we can harmonize the task difficulty across participants by standardizing the starting positions. This feature reduces variance in our data. Third, we can control the strategies' complexity and effectiveness. Our design ensures that the second strategy is simpler and faster than the first one.[12] Moreover, the second trailer informs learners about this fact, and they act upon it.[13] Fourth, we can incentivize performance by granting a piece rate for each puzzle solved. Fifth, the task allows us to measure participants' beliefs about the tutorials' usefulness (see Subsection 2.4).

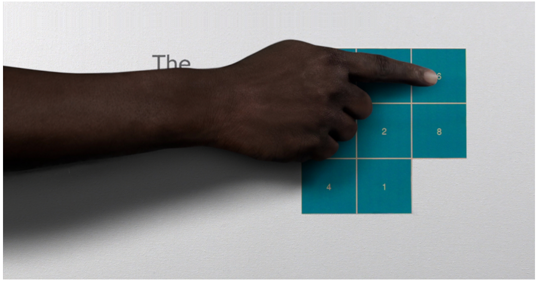
[11] Learners who did not watch the full tutorial solved on average 0.92 fewer puzzles than those who saw it ($p = 0.01$). On average, the latter group of participants solved 2.9 puzzles in five minutes.

[12] We use the A^* pathfinding algorithm to validate this statement. This algorithm presents a way to find the shortest path from the source (starting position) to the solution (solved puzzle). In our case, tile moves define a path, and we can count the minimal number of moves needed to solve the puzzle with a given strategy. Appendix E.2 demonstrates that the second strategy outperforms the first one.

[13] In the first stage, 66.2% of participants use the strategy presented in the tutorial. In the second stage, 74% of all learners employ the second-stage strategy.

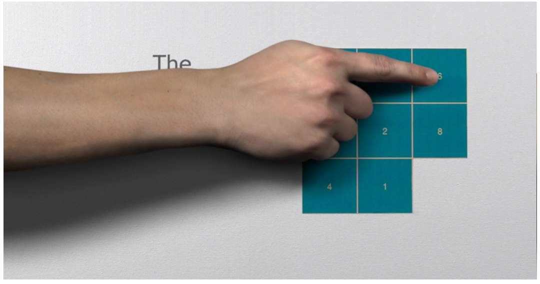
Figure 3: Skin-Color Treatments

(a) First actor: Black



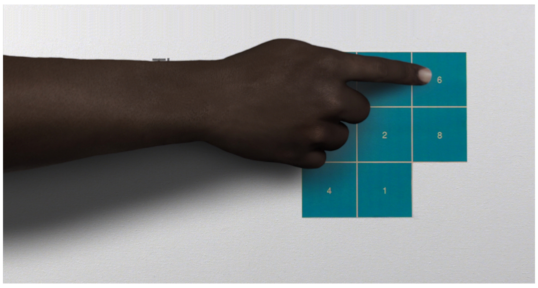
First strategy: preview, full tutorial
Second strategy: preview, full tutorial

(b) First actor: White



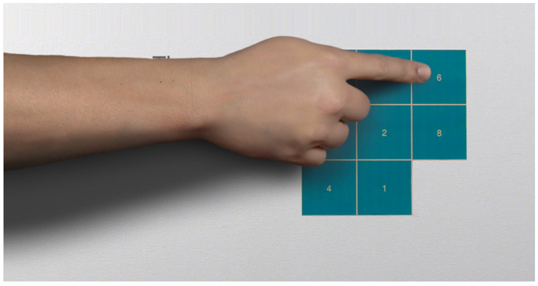
First strategy: preview, full tutorial
Second strategy: preview, full tutorial

(c) Second actor: Black



First strategy: preview, full tutorial
Second strategy: preview, full tutorial

(d) Second actor: White



First strategy: preview, full tutorial
Second strategy: preview, full tutorial

Notes: Screenshots of skin-color treatments. Each figure provides links to the preview videos and the full tutorials for both stages.

2.3 Treatments

We implemented orthogonal treatments that vary the tutor's skin color (*skin-color treatments*), the available information (*information treatments*), and the piece rate for each solved puzzle (*piece-rate treatments*).

Skin-color treatments: Building on Doleac and Stein (2013) and Evsyukova *et al.* (2025), our skin-color treatments modify the tutor's race without changing any other aspect of the tutorials (e.g., the tutor's quality or his characteristics). We achieve this by using a highly standardized video-production process (see Appendix E.1 for details). Specifically, we recruited two male Hispanic actors as tutors. A professional recorded each actor presenting the puzzle in front of a greenscreen, which he later rendered to display the sliding tile puzzle in front of a whiteboard. Specifically, each actor recorded four videos: a 30-second trailer (firstly revealing the tutor's race) and a 170-second e-learning tutorial (detailing puzzle-solving strategies) for each of the two stages. At several instances in the videos, the actor's hand appears to point at and highlight certain aspects of the puzzle. A Visual Effects Artist—who has worked on blockbuster movies such as Star Wars, Terminator, or Aladdin—then used post-production techniques to alter the skin color of the Hispanic actors to "Black" or "white" (resulting in 16 different videos). Figure 3 displays screenshots of the final videos for both actors. It also provides links to the corresponding trailers and tutorials. Moreover, to standardize the tutors' voices, we recorded two distinct voice-overs following a detailed script. One was provided by a white and one by a Black US native. Lastly, we combine all videos with these voice-overs.[14] Our procedure ensures that the tutor's race is orthogonal to the actor and the voice.

We then randomly assign learners to one of 16 treatment arms that vary by actor, voice, and skin color across the two stages. Specifically, in the first stage, each learner receives one of the two actors with a randomly assigned voice and skin color (Black or white). The same tutor appears in the trailer and full tutorial. In the second stage, learners are assigned to (a) the other actor and (b) the other voice not used in the first stage. Instead, the skin color is independently re-randomized. The reason is that learners choose between both tutors in the second stage, so they must be different in terms of hand model and voice (but not race). Appendix Table A1 visualizes the characteristics of all 16 treatment arms. We conducted a survey to confirm that the videos are of high quality and appear natural.[15]

Information treatments: Our next intervention aims to test if learners engage in statistical (belief-based) discrimination[16] when selecting a tutor *in the second stage*. The correspondence design pioneered by Bertrand and Mullainathan (2004) is the standard method for identifying this type of behavior. It informs participants that members of two groups (e.g., Black and white tutors) share similar characteristics (e.g., educational backgrounds). Statistical discriminators should update their beliefs and, in turn, adjust

[14] Earlier research documents no differences between aerodynamic and acoustic characteristics of African American and White speakers (Sapienza, 1997). Thus, presenting race-incongruent voices should not lead to an unnatural experience (see Subsection 4.3 and footnote [41] for supporting evidence).

[15] We recruited 100 U.S. residents via CloudResearch to elicit perceptions of the hands. Each participant viewed two randomly selected screenshots featuring either the (a) original Hispanic, (b) the manipulated white, or (c) the manipulated Black hand. When asked, "Does the screenshot look natural to you?", 97% approved the Hispanic hand, 95% the white hand, and 93% the Black hand. Those who found the images unnatural primarily mentioned concerns about the background. Only four participants mentioned the skin color of one of the two actors. No one mentioned the skin color of the second actor. In sum, most respondents did not notice any manipulation.

[16] Statistical discrimination occurs when people judge individuals based on beliefs about group statistics instead of assessing their individual characteristics. It arises from a lack of individual-level information, not prejudice, and disappears when such information becomes available. For example, learners may have limited knowledge of a tutor's teaching style and ability, leading them to infer their type based on group statistics (unless they obtain more information).

their choices. However, correspondence studies can fail to provide information on all relevant characteristics, leading to persisting belief differences.

Responding to this challenge, we implement a more direct *information treatment* in the second stage. Immediately before learners choose the second-stage tutor, we randomly select 50% of all learners and inform them truthfully 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."*

The treatment aims to equalize learners' beliefs about the quality of instruction directly, instead of indirectly by providing details on tutor characteristics. Participants in the *no-information treatment* only learn that both tutorials have the same length.

Piece-rate treatments: In additional analyses, we build on Hedegaard and Tyran (2018) and examine how discrimination reacts to changes in its costs. For that, we randomly assign 50% of learners to a cross-cutting treatment that substantially increases the piece rate from $0.2 (*standard piece-rate*) to $1 (*high-stakes piece rate*).[17] The key idea is that under a higher piece rate, disregarding advice leads to greater financial losses, making discrimination more costly.

Stratification: We stratify the treatment allocation by race (Black, white, other), education (no college degree, some college, or higher), and state (South, other), following the U.S. Census classification. Within each stratum, we randomly order all available treatments into a list. Incoming learners in a given stratum are assigned sequentially: the first receives the first treatment in the list, the second receives the second, and so on. Once all treatments are assigned, we generate a new random order and repeat the process.

2.4 Measuring Outcomes Throughout the Learning Process

Our design measures (a) tutor selection, (b) knowledge acquisition and utilization, (c) performance, and (d) beliefs. Next, we detail these measures.

Tutor-selection decisions: Discrimination in the selection process may limit Black individuals' access to advisory roles. We include two measures to assess whether Black people are, in fact, less likely to be selected as tutors.

[17] In the high-stakes treatment, participants earned over $18 per hour on average ($\approx$ 90% more than they usually get on CloudResearch).

WTP: Our first measure is the participant's willingness to pay for the full tutorial, stated in their purchasing decision (*Step 3* of the first stage). The *WTP* serves as proxy for tutor selection as higher values reflect a stronger desire to get a given tutor.

We use the Becker-DeGroot-Marchak (BDM) procedure (Becker *et al.*, 1964; Becker and DeGroot, 1974) to model the purchasing decision and ensure an incentive-compatible *WTP* measurement. The procedure is as follows: After having watched the tutorial in *Step 1*, learners receive an extra \$1 they can use to buy the full tutorial. They can allocate any amount of this money, in \$0.01 increments, toward the purchase. This value represents participant i's WTP. The computer then randomly assigns each learner a \$ price $p_i \in \{0, 0.01, 0.02, ..., 1\}$ for the full tutorial. If $WTP_i \geq p_i$, learner i pays p_i and purchases the full tutorial. If $WTP_i < p_i$, learner i pays nothing and watches an uninformative video of equal length showing fish.[18] By comparing the *WTPs* between learners in the Black and white treatments, we can estimate the average willingness to discriminate.

We assign a 95% probability to the price \$0.00 and distribute the remaining 5% uniformly across all positive prices $p \in \{0.01, 0.02, ..., 1\}$. This design feature ensures that (a) most learners watch the full tutorial and (b) face Black or white tutor randomly (independent of their *WTP*). Random assignment eliminates self-selection bias, while assigning a 95% probability maximizes statistical power. As a result, for learners with $p_i = 0$, we can not only isolate the causal effect of race on performance but also study if this effect operates through differences in knowledge acquisition and utilization. Note that the instructions truthfully state that each price is drawn with a positive probability (without specifying the details).[19]

Explicit tutor choice: Our second measure captures the explicit tutor choice. In *Step 3* of the *second stage*, learners choose whether the first-stage or second-stage tutor delivers the full tutorial. To implement this tutor-selection process, we use an incentive-compatible mechanism inspired by Toussaert (2018). The mechanism has two properties:[20] One is that it incentivizes learners to reveal their tutor preference by ensuring that stating a preference for a tutor increases the probability of being assigned to him. The second is that it incorporates a random component in the final tutor assignment to enable causal identification of tutor effects.

The mechanisms works as follows: After the trailer, learners indicate if (a) they

[18] Learners know that this video (a) provides no puzzle-solving strategies and (b) has equal length.

[19] Reassuringly, the distribution of the *WTP* spans the entire domain of possible values and it is shifted to the right under the high piece rate, as predicted by the law of demand (Appendix Figure A2).

[20] The mechanism also offers a plausible deniability of learners' preferences for a tutor of a particular skin color. On an individual level, we cannot observe whether the learner prefers a tutor because of the skin color or because of the (non-)familiarity with him.

prefer the first-stage tutor, (b) the second-stage tutor, or (c) if they are indifferent between the two. This is our measure of interest. A lottery then determines if the stated preference is implemented or not. With a probability of 5%, learners can get their preferred tutor for sure. This feature guarantees the mechanism's incentive compatibility. With the residual probability (95%), the second-stage tutor presents the full tutorial regardless of the stated preference. By design, the skin color of this tutor is random, so most learners get a Black or white tutor by chance. The instructions mention that the probabilities are positive without disclosing their values (see Appendix E.4 for further information).

Beyond its core function, our selection procedure has two additional benefits. First, it neither explicitly juxtaposes Black and white tutors nor mentions their race explicitly. Instead, learners discover the tutors' skin color organically through the tutorials. This feature reduces experimenter demand effects and offers learners plausible deniability in tutor selection. Second, it provides the opportunity to integrate the information treatment. We can notify learners that the two available tutorials, despite being presented by different tutors, are identical in terms of content.

Knowledge acquisition and utilization: Each of the two full tutorials teaches a puzzle-solving strategy that prescribes a distinct sequence of tile moves (described in Appendix E.3). This feature allows us to measure advice acquisition and utilization by tracking whether learners adopt the strategies. Intuitively, strategy adoption requires both that learners internalize a strategy (knowledge acquisition) and apply it (utilization). To operationalize strategy adoption, we construct (a) a dummy indicating if a learner ever followed the prescribed sequence of moves in a stage and (b) a variable counting the number of puzzles solved using this strategy.

Performance: The sliding tile puzzle enables straightforward performance measurement. We track whether a learner solves at least one puzzle (extensive margin), count the number of puzzles solved in five minutes (intensive margin), and calculate the average number of moves needed to solve a puzzle in a stage.

Beliefs: We elicit beliefs in *Step 2* of both stages to differentiate belief-based discrimination from other forms and to assess whether learners expect the tutorials to be helpful. Specifically, in the first stage, we ask learners to estimate how many puzzles they will solve in five minutes after watching either the full tutorial or the uninstructive entertainment video. In the second stage, they predict their performance after watching the full tutorial presented by either the first-stage or the second-stage tutor. Screenshots 6 and 13 in Appendix C present the precise questions used to elicit beliefs.

One potential concern is that our belief measures do not reflect true expectations because we did not incentivize truthful reporting. However, several patterns suggest otherwise: Participants expect to solve more puzzles after watching the tutorial than without it (5.15 vs. 2.97, $p < 0.01$). Also, learners who reported knowing the puzzle were significantly more optimistic about their performance than those who did not, both when they watched the full tutorial (6.49 vs. 5.15, $p < 0.01$) and when they did not (5.19 vs. 2.97, $p < 0.01$). Lastly, learners adjust their beliefs after the first stage in line with their actual performance: Those who underestimated their performance in the first stage increased their expectations, while those who overestimated it decreased them ($p < 0.01$).

2.5 Further Details and Sampling

Payoffs: We compensate learners for participation. A learner i's payoff function is:

$$Payoff_i = \overbrace{\$5}^{\textit{Fixed pay}} + \{ \overbrace{\$r_i P_i}^{\textit{Performance-based pay}} + \overbrace{\$1}^{\textit{WTP endowment}} - \overbrace{q_i}^{\textit{Price for tutorial}} \}, \tag{1}$$

where the first term is a guaranteed fixed payment of \$5, $r_i \in \{0.2, 1\}$ reflects i's randomly assigned piece rate per completed puzzle, and P_i denotes how many puzzles i solved. Moreover, i receives a \$1 endowment for the WTP procedure[21] and q_i is the paid price for accessing the tutorial. This price equals the randomly drawn price p_i if $WTP_i \geq p_i$; instead, it is zero if $WTP_i < p_i$. We randomly select one of the two stages and pay learners based on their performance in this stage. The average payoff is \$8.5, translating to an hourly wage of over \$20.

Further details: Several further details are worth nothing. First, to standardize the opportunity cost of time across learners and treatments, we fix the duration of each step of the experiment. Upon expiration of this allotted time, the system automatically redirects learners to the subsequent page. Should a page necessitate an action and the learner does not respond in time, we discontinue their participation in the study. Second, our website incorporates a standard attention check at the start of the instructions and excludes learners who fail it.[22] Third, we inform learners that tutors do not benefit from participants purchasing the tutorials (i.e., learners cannot influence a tutors' earnings). Thus, differences in purchasing decisions between Black and white tutors cannot be driven by (a) race-specific altruism, (b) a desire to harm

[21] As detailed in Appendix E.4, in the second stage, learners indicate their WTP to get their preferred advisor for sure. This feature allows us to separate learners with a strict preference for a certain tutor from those who are truly indifferent.

[22] Our instructions include the following text: "Enter the following number into the text box to show that you pay attention: 65."

tutors of a given race, or (c) a wish to counteract existing racial income inequalities. Fourth, the experiment ends with a survey on learners' demographics, prior knowledge of the sliding tile puzzle, and recall of tutor attributes (such as race). It also asks about the puzzle-solving strategies learners used in both stages. Fifth, we take steps to prevent data quality from being compromised by issues commonly arising in online settings (see Appendix E.5).

Sampling: We applied three sample-selection criteria when recruiting participants through CloudResearch. First, we only invited "CloudResearch-approved participants" who passed the platform's own attention and engagement checks. Second, we sampled users who CloudResearch pre-profiled on their political views.[23] This step allows us to target our study to conservative and liberal Americans. Third, we drew from two distinct participant pools offered by CloudResearch: approved MTurk workers and users from CloudResearch Connect.[24] Neither pool alone provided a sufficiently large number of observations. In total, 4,396 participants started our study.

We then construct our final estimation sample in several steps. The first step is to drop participants who did not even complete the stratification procedure and did not open the initial instructions. Next, we filter out participants who (a) failed our attention check, (b) have been screened out due to inactivity, or (c) had missing data, likely due to connectivity issues. Next, we exclude participants with prior knowledge of the sliding tile puzzle. These participants benefit less from the tutorial, complicating the measurement of discrimination in learning.[25] Finally, we restrict the sample to learners for whom we randomly assigned tutors in both stages (90.25% of all participants). Put differently, our sample is the intersection of participants who were randomly assigned (a) a price of $p = 0$ in the first stage (95%) and (b) a second-stage tutor, regardless of their stated preference (95%). The resulting estimation sample consists of 2,406 participants: 1,231 liberals and 1,175 conservatives, as classified by CloudResearch (see footnote [23]). Appendix Table A2 reports summary statistics, Appendix Table A3 demonstrates balanced observables across treatments, and Appendix Table A4 confirms that attrition and sample restrictions are treatment-independent.

[23] CloudResearch constantly runs its own surveys, allowing for targeted study recruitment. The political view question reads: "How would you describe your political views?" The response categories are "very conservative," "conservative," "moderate," "liberal," and "very liberal." We classify conservatives as those who answered "conservative" or "very conservative" and liberals as those who chose "liberal" or "very liberal."

[24] Initially, CloudResearch served as a management tool for MTurk studies. However, following its professionalization, it now features its own participant panel "CloudResearch Connect."

[25] Participants familiar with the puzzle expect a smaller increase in own productivity—measured as the difference in first-stage beliefs between watching and not watching the tutorial—compared to those unfamiliar with the puzzle (1.30 vs. 2.18, $p < 0.01$). Consequently, they also state a lower WTP (0.40 vs. 0.58, $p < 0.01$).

PAP sample: The PAP specified that we would focus on learners with extreme political views who CloudResearch identified as either "very conservative" or "very liberal." We committed to collect about 1,000 observations in each category, plus the same number of observations on learners with moderate political views (classified by CloudResearch as "liberals," "moderates," and "conservatives"; about 330 observations in each category). However, we demonstrate that "very conservative" and "conservative" participants not only exhibit (a) highly similar observable characteristics (see Appendix Table A5) but (b) also very similar behavioral patterns (see Appendix Figures A3 and A4). The same is true for "very liberal" and "liberal" participants. To reflect these unanticipated similarities, our main analyses combine "very conservatives" and "conservatives" into a single group (labelled "conservative"), and does the same for "very liberals" and "liberals" (labelled "liberal"). Our results remain virtually unchanged when restricting the analyses to the pre-registered categories (see Sections 4 and 5). Note that the broader categorization increases the external validity of our findings, because the results now apply to a larger segment of the population. The second deviation from the PAP is that we exclude participants who already new the puzzle, as this share was higher than we expected when designing the study (11% of conservatives and 8% of liberals). Lastly, we did not pre-register the exclusion of participants with missing data due to connectivity issues, as we also did not anticipate this issue. This restriction is necessary to ensure identical samples sizes across all specifications.[26] Crucially, also these two additional restrictions do not impact our results: the point estimates and the confidence intervals are remarkably stable (see Appendix Figures A3 and A4). Due to the additional restrictions, our final estimation sample of 2,406 participants is slightly smaller than the sum of the planned sample sizes across the categories "very conservative", "conservative", "liberal", and "very liberal." Note that we preregistered our study in several steps. Appendix D provides background on the evolution of the design.

Discrimination in online samples: A key concern for our study is that people in online samples may differ from the general population in that they do not engage in discriminatory behavior. To test this idea, we conducted a pre-registered follow-up study. Its goal is to examine whether discriminatory behavior emerges in an experiment researchers have commonly used to demonstrate discrimination: the "help-or-harm task" (Bartos *et al.*, 2021).

The design was as follows: Several weeks after the main experiment, all participants who completed it received a standard invitation for a new study. The invitation did not mention the original experiment, rendering it unlikely that participants could

[26]Column 3 of Appendix Table A4 shows that the missing data is uncorrelated with treatment.

connect both studies. After accepting the invitation, participants' task was to increase or decrease payoffs of randomly selected Black or white people, at no monetary costs or benefits to themselves. The default payoff was $5, which they could increase ("prosocial behavior"), decrease ("hostile behavior"), or leave unchanged (within a $0-$10 range). We classify behavior as discriminatory when participants award higher payoffs to white than Black people or display greater hostility toward Black individuals.

In total, 2,057 (or 85%) of the main study's participants completed the follow-up study.[27] Among conservatives (liberals), 15.8% (3.0%) gave more to white than Black recipients. Instead, 34.2% (11.2%) of liberals (conservatives) allocated more money to Black recipients. Conservatives were also more likely to reduce the rewards of a Black than white recipients below the default (17.4% vs. 13.0%). Instead, only 7.3% of liberals acted hostile against Black recipients. These findings reveal meaningful levels of (hostile) discriminatory behavior in our sample. Moreover, they also confirm the presence of ideological differences in discrimination in our sample, consistent with findings from large, representative studies (Pew Research Center, 2021; Haaland and Roth, 2023). Appendix F provides additional details on the design and results of the follow-up study.

3 Conceptual Considerations

This section proposes a simple framework to structure the discussion of how discrimination (in theory) affects the learning process in our experiment. The framework guides our empirical analysis and distinguishes four types of individuals: (a) non-discriminators, (b) statistical discriminators, (c) taste-based discriminators, and (d) unconscious discriminators. Appendix H presents a more detailed discussion of the model, including all derivations.

Learning process: We view the learning process as a multifaceted process that goes beyond mere knowledge acquisition. It involves several interconnected steps: selecting whom to learn from, acquiring and internalizing new knowledge, and applying it to relevant contexts. We embed this view of the learning process into a model with two stages. The first stage, the *tutor-selection stage*, enables learner i to choose between a Black (B) and a white (W) tutor. Learner i selects the tutor that offers the highest expected utility. Forces such as statistical discrimination (based on beliefs about how effective each tutor is in teaching) and taste-based discrimination (based on conscious

[27] The treatments in the main experiment did not influence participants' decision to join the follow-up study. Moreover, the participation probability is uncorrelated with observable characteristics of conservatives and liberals (see Appendix Table F1).

biases against Black tutors) can influence this choice. The second stage is the *knowledge-acquisition-and-utilization stage*. For simplicity, learner i faces a binary decision on whether or not to acquire knowledge (i.e., to learn) from the tutor.[28] To make this decision, i compares the expected benefits of learning (i.e., higher monetary payoffs) with the associated expected learning costs (e.g., cognitive effort or discomfort). As clarified later, statistical, unconscious, and taste-based discrimination can affect the expected benefits and costs of learning. Learner i then applies the acquired knowledge in puzzle-solving.[29]

Expected utility: We next formulate a simple expected utility function that nests the four types of learners and allows us to predict behavior in our experiment. Consider learner i who (a) chooses between a Black or white tutor $j \in \{B, W\}$ and (b) decides whether ($L_i = 1$) or not ($L_i = 0$) to learn from j. Learner i's *expected utility* is:

$$U_i^j(L_i) = \overbrace{r_i \cdot \left[L_i \cdot \phi^j \cdot E[P_i^{Max}] + (1 - L_i) \cdot E[P_i^{L_i=0}]\right]}^{\text{expected earnings}} - \overbrace{L_i \cdot (\bar{c} + c^j)}^{\text{learning cost}} - \overbrace{\tau^j}^{\text{disutility of interacting w. } j} . \quad (2)$$

The *first term* reflects i's expected earnings. These depend on the piece rate (r_i) and vary based on whether i chooses to learn from j ($L_i = 1$) or not ($L_i = 0$). Without learning, i expects to solve $E[P_i^{L_i=0}]$ puzzles. Her expected earnings are $r_i \cdot E[P_i^{L_i=0}]$. If the learner, instead, decides to learn from j, expected earnings are $r_i \cdot \phi^j \cdot E[P_i^{Max}]$. Here, $E[P_i^{Max}]$ denotes the expected maximum number of puzzles solved under optimal instructions[30] and $\phi_j \in [0, 1]$ captures how effective i believes tutor j is in helping them reach that potential. A higher value implies that i anticipates to learn more from j. Importantly, this belief parameter $\phi_j \in [0, 1]$ determines the realized performance when learning (i.e., $\phi^j \cdot E[P_i^{\text{Max}}]$). This reflects the idea that learners who expect to learn less may engage less (e.g., by paying less attention or processing advice less deeply).

The *second term* reflects the (monetized) learning costs paid when i decides to learn from j ($L_i = 1$). Learning costs have two components: The first component $\bar{c}$ models the general (tutor-independent) cost associated with acquiring knowledge. It includes factors like time spent studying or mental effort expended. The second component c^j measures that some learners could experience an additional conscious cost of learning from a tutor of type j. Reasons include conscious aversions, biases, or

[28]In practice, whether learners acquire and utilize knowledge can reflect multiple factors. For example, they may either choose not to pay attention or listen actively but still disregard the information.

[29]Since our model would predict equivalent behavior for the decisions to acquire and to utilize knowledge, we do not model these outcomes separately. Accordingly, our empirical design does not distinguish between them.

[30]Thus, $E[P_i^{Max}]$ is the expected performance when i faces the best possible tutor and perfectly applies the advice.

Table 1: Types of Learners

Type	Beliefs about effectiveness	Distaste parameter	Conscious cost of learning	Unconscious cost of learning
Non-discriminators	$\phi^B = \phi^W = \phi$	$\tau^B = \tau^W = 0$	$c^B = c^W = 0$	$c^{B,u} = c^{W,u} = 0$
Statistical discriminators	$\boldsymbol{\phi^B < \phi^W = \phi}$	$\tau^B = \tau^W = 0$	$c^B = c^W = 0$	$c^{B,u} = c^{W,u} = 0$
Taste-based discriminators	$\boldsymbol{\phi^B \leq \phi^W = \phi}$	$\boldsymbol{\tau^B > \tau^W = 0}$	$\boldsymbol{c^B \geq c^W = 0}$	$c^{B,u} = c^{W,u} = 0$
Unconscious discriminators	$\phi^B = \phi^W = \phi$	$\tau^B = \tau^W = 0$	$c^B = c^W = 0$	$\boldsymbol{c^{B,u} > c^{W,u} = 0}$

negative preferences towards this tutor.

The *third term* (τ^j) represents a general taste-based (monetized) disutility learners may feel when interacting with tutor j. This term always reduces i's overall utility whenever facing j, whereas i incurs the cost c^j only conditional on choosing to learn from j.

Unconscious biases in learning: For some learners, unconscious costs and biases may influence the decision to acquire knowledge from j without their awareness. These costs may stem from past experiences, implicit stereotypes, or societal conditioning (Devine, 1989; Greenwald *et al.*, 1998). To capture this idea, we introduce the concept of the *actual (unconscious) expected decision utility*. While i expects and intends to base the learning decision on equation (2), the actual decision is based on:

$$U_i^{j,actual}(L_i) = U_i^j(L_i) - L_i \cdot c^{j,u}. \tag{3}$$

In equation (3), $c^{j,u}$ models an unconscious cost of learning from tutor j. For example, i may unconsciously associate race j with lower teaching effectiveness, leading to higher cognitive effort or discomfort when engaging with that tutor's material. This cost can diminish i's ability to absorb and apply knowledge.[31]

Next steps and baseline scenario: The following paragraphs define the four types of learners within our framework by setting restrictions on the model parameters. They also summarize predictions about each type's behavior in the experiment. We derive these predictions by applying backward induction to a *baseline scenario* that closely mirrors our setup (see Appendix H). In this scenario, (a) learners randomly access instructions from white or Black tutors (no selection bias), (b) both tutors provide equally effective instruction (though beliefs may differ due to factors outside the

[31]Our reduced-form approach to model unconscious biases abstracts from the specific underlying psychological mechanisms. The reason is that we aim to establish a general (simple) framework that guides our analysis, rather than capturing all nuances of unconscious discrimination. We, however, acknowledge that the cost $c^{u,B}$ can stem from various sources (such as implicit biases formed through socialization, stereotype threat, or reliance on automatic processing under cognitive load).

experiment), (c) instructions boost performance, and (d) learners receive no additional information about tutor effectiveness (no-information treatment). Table 1 summarizes the definitions of the types, and Table 2 shows type-specific predictions.

Non-discriminators: *Non-discriminators* believe Black and white tutors $j \in \{B, W\}$ provide equally effective instructions ($\phi^j = \phi$), experience no disutility when interacting with either group ($\tau^j = 0$), and incur no conscious or unconscious j-specific learning cost ($c^j = c^{j,u} = 0$). Given these parameters, they exhibit no discriminatory behavior (see Appendix H.1 for derivations): First, as they perceive no differences in both tutors' effectiveness, they are equally likely to learn from Black and white tutors whom they randomly face (knowledge-acquisition stage). Second, because both tutors offer equally effective instructions, non-discriminators perform the same under B and W. Third, they anticipate this behavior and, thus, expect similar performance with either tutor. Fourth, based on this belief and identical expected utilities, non-discriminators are indifferent between tutors (tutor-selection stage).

Statistical discriminators: Statistical discrimination occurs when people treat individuals differently based on beliefs about group statistics instead of assessing their qualifications or characteristics (Phelps, 1972; Arrow, 1973).[32] Statistical discriminators could, for example, believe that Black tutors (a) use didactic approaches that do not align with their learning style, (b) tailor instructions to the needs of other groups, or (c) offer lower overall quality. We operationalize these considerations by assuming that statistical discriminators expect Black tutors to be less effective teachers ($\phi^B < \phi^W$). Statistical discriminators feel no disutility interacting with Black tutors ($\tau^B = \tau^W = 0$) and have no conscious or unconscious motive to reject their advice ($c^B = c^W = c^{B,u} = c^{W,u} = 0$). Appendix H.2 derives predictions: Because statistical discriminators perceive Black tutors to be less effective, they are less likely to learn from them and, when they do, engage less with their content. Therefore, they also perform worse with Black tutors. Moreover, they rationally anticipate these performance differences and choose white over Black tutors in line with their expected performance. In short, statistical discriminators perform better with white tutors because they (wrongly) expect Black tutors to provide less effective instructions and, thus, reject their advice.

Taste-based discriminators: Due to prejudice or bias against Black tutors (Becker, 1957), taste-based discriminators experience disutility when interacting with them ($\tau^B > \tau^W = 0$). This distaste can extend to an aversion to learning from Black tutors, modeled as additional conscious learning costs ($c^B \geq c^W = 0$). Their bias may also

[32]These beliefs stem not from personal prejudice but from perceived (rational) expectations. Many papers on statistical discrimination assume beliefs are correct. Bohren *et al.* (2023) is a notable exception.

Table 2: Summary of Model Predictions for Different Types of Discriminators

Type	**Learns more likely from W**	**Performs better under W**	**Expects to perform better under W**	**Tutor Selection**
Non-discriminators	No	No	No	Indifferent
Statistical discriminators	Yes	Yes	Yes	Selects W
Taste-based discriminators	Yes	Yes	Yes	Selects W
Unconscious discriminators	Yes	Yes	No	Indifferent

Notes: Summary for (non-)discriminator type whether learners are more likely to learn from white (W) tutors, perform better under W, foresee own performance differences under W, and whom they choose as a tutor. Predictions for a baseline scenario with participants randomly assigned to a white or Black tutor, both tutors providing equally effective instructions, watching instructions boosts performance, and learners receiving no additional information on tutor quality.

lead them to believe that Black tutors are less effective ($\phi^B \leq \phi^W = \phi$).[33] Taste-based discriminators—who either perceive additional conscious learning costs and/or expect Black tutors to be less effective—act like statistical discriminators (see Appendix H.3): They are more likely to acquire knowledge when they randomly face white tutors, perform better under them, anticipate this performance difference (aware of their biases, Bertrand *et al.* 2005), and consciously select white over Black tutors. Unlike statistical discrimination, this behavior arises from bias against B.

Unconscious discriminators: *Unconscious discriminators* believe Black and white tutors are equally effective ($\phi^B = \phi^W = \phi$). They feel no disutility from interacting with either tutor ($\tau^B = \tau^W = 0$) and experience no B-specific learning cost ($c^B = c^W = 0$). However, they incur an unconscious learning cost with Black tutors ($c^{B,u} > c^{W,u} = 0$). This unconscious bias affects their behavior (see Appendix H.4): They are more likely to learn from white tutors, leading to better performance. Unaware of this bias, they, however, expect to perform equally well with both tutors. This belief keeps them indifferent when choosing between B and W. Empirically, such unconscious biases appear stronger under time pressure and in low-stakes situations (Correll *et al.*, 2002; Kahneman, 2003; Bertrand *et al.*, 2005; Kahneman, 2011). In these contexts, learners have fewer opportunities or weaker incentives to engage in slower, deliberate thinking (System 2 Thinking). Instead, they rely on fast, automatic, and intuitive thinking (System 1 Thinking). Then, automatic, heuristic-based processes can dominate

[33] For example, taste-based discriminators may assume, without evidence, that Black tutors lack skill or intelligence in certain areas. This belief reflects their distaste rather than objective judgment. It is *ex-ante* unclear whether, in addition to a general distaste toward Black tutors (τ^B), taste-based discrimination leads to higher conscious learning costs when learning from them (c^B), or lower perceived effectiveness of their instruction (ϕ^B). Different learners may exhibit any combination of these factors. By measuring various outcomes, our experiment can differentiate between these scenarios.

decision-making, allowing unconscious biases to influence behavior.

Unconscious discrimination in selection decisions: As apparent, our framework abstracts from explicitly modeling unconscious discrimination in tutor selection. The reason is that, in our experiment, participants are encouraged to make deliberate tutor choices and are given ample time to do so, without any external pressure. Thus, unconscious bias in selection is less likely. However, in other settings, people might discriminate in selection decisions. An extended model including an unconscious disutility of interacting with a tutor ($\tau^{j,u}$) would capture this possibility.

Black-favoring behavior: In principle, our general framework can also accommodate anti-discriminatory behavior (depending on parameter values). For example, learners might perceive Black tutors as more effective ($\phi^B > \phi^W$) or derive additional utility from interacting with or learning from them ($\tau^W < 0 = \tau^W$ or $c^B < 0 = c^W$). Then, the model's predictions reverse. In theory, all parameter combinations are possible. Some people, for example, may feel positive utility from interacting with Black tutors but might gain no extra benefit when learning from them.

Effect of high-stakes treatment: Appendix I discusses how increasing the piece rate r_i affects discriminatory behavior in our framework. The three main results are: First, as a higher piece rate raises the monetary gains from learning, it reduces discrimination in *knowledge acquisition* across all types. With greater potential rewards, learners become more likely to learn from Black tutors.[34] Second, a higher piece rate does not prevent discrimination in *tutor selection* for statistical and taste-based discriminators. Even when statistical discriminators start to acquire knowledge from Black tutors, they still expect (and realize) lower performance than when learning from white tutors (due to $\phi^W > \phi^B$). Thus, they continue to select W over B. Similarly, taste-based discriminators prefer white tutors because of their aversion against Black tutors ($\tau^B > 0$), potentially higher learning costs ($c^B \geq 0$), or potentially lower perceived effectiveness ($\phi^B \leq \phi^W$). Third, as they already expect equal performance under both tutors, unconscious discriminators *remain indifferent* in tutor selection. In sum, a higher piece rate may reduce discrimination in knowledge acquisition but does not necessarily eliminate it in tutor selection.

[34] The mechanisms vary by type: (a) For statistical discriminators, a higher r_i increases the financial benefit of learning from any tutor. Despite the lower perceived effectiveness ($\phi^W > \phi^B$), statistical discriminators might also find it beneficial to acquire knowledge from Black tutors. (b) In taste-based discrimination, the higher piece rate can additionally offset the extra learning costs (c^B), encouraging learning despite biases. (c) Among unconscious discriminators, a rise in r_i helps counterbalance the unconscious cost $c^{B,u}$. Thereby, it increases the attractiveness of learning from Black tutors. This mechanism aligns with the mentioned idea that higher stakes increase the likelihood learners overcome their unconscious biases by engaging in more reflective thinking.

Effect of information treatment: Statistical and taste-based discriminators behave identically in the baseline scenario. The information treatment—designed to equalize the perceived effectiveness between B and W—can nevertheless help us to distinguish these two types (see Appendix J for details). This is because, in theory, both types respond to information very differently. When statistical discriminators discover W and B provide identical (equally effective) instructions, they learn from both at the same rate. They, therefore, also perform equally well with either tutor and stop selecting white ones. By contrast, due to their biases, the treatment does not fully eliminate taste-based discrimination in tutor selection and knowledge acquisition.[35] When facing Black tutors, taste-based discriminators even accept lower performance (and thus lower payoffs) to avoid learning from B (due to the cost $c^B > 0$). As a result, they reject valuable, performance-enhancing knowledge. Note that the information treatment does not affect unconscious discriminators (as they expect no difference in tutor effectiveness).

4 Results for Conservatives

Our first set of findings focuses on learners with conservative political views. Conservatives engage in unconscious discrimination (see Subsection 4.1), but not if the stakes are high (see Subsection 4.2).

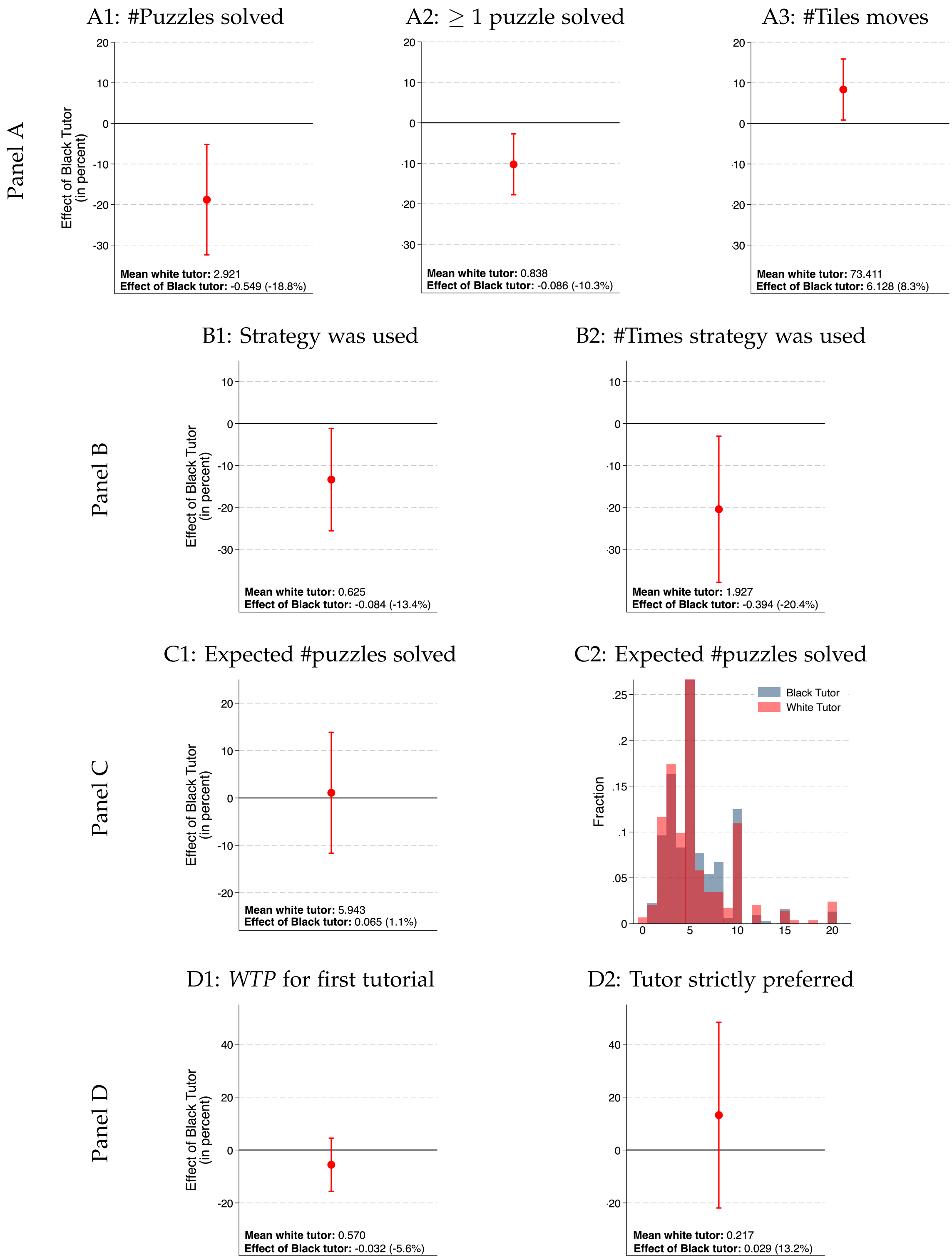
4.1 Behavior Under Standard Incentives

Performance: We first focus on the standard piece-rate treatment and demonstrate the effects of having a Black tutor, compared to a white one, on conservatives' *first-stage* performance. Panel A in Figure 4 summarizes the results using the performance measures introduced in Section 2.4. It shows all treatment effects and their 95% confidence intervals as percentages of the average performance among learners who had white tutors.[36] Columns (1) to (3) of Appendix Table A6 report the corresponding regression results. Our first finding is:

[35]Biases toward learning from white tutors persist in two cases. First, learners might reject the information and still perceive the instructions of white tutors to be more effective ($\phi^W > \phi^B$). Second, even if they adjust their perceived effectiveness, they still experience an additional B-specific learning cost (c^B). Biases in the tutor-selection decision exist due to $\tau^B > 0$ and $c^B \geq 0$.

[36]Except for Figure 4D2, all estimates in Figure 4 are based on the OLS regression: $Y_i = \beta_0 + \beta_1 Black_i + \beta_2 Black_i \times High_i + \beta_3 High_i + X_i'\gamma + \varepsilon_i$, where Y_i is an outcome for learner i, $Black_i$ indicates whether tutor is Black, $High_i$ is a dummy variable for the high-stakes piece rate, and X_i is a vector of strata controls. The figure then plots $\widehat{\beta}_1 / E[Y_i | Black_i = 0, High_i = 0]$, the effect of having a Black tutor under standard incentives relative to the average outcome of learners with white tutors under standard incentives.

Figure 4: Results for Conservatives Under Standard Piece Rate



Notes: Effects of tutor race on learners' behavior. Sample: Conservatives under standard piece rate. Dependent variables: Number of puzzles solved in first stage (4A1), dummy for learners who solved at least one puzzle in first stage (4A2), average number of moves needed to solve a puzzle in first stage (4A3), dummy for learners who solved at least one puzzle in first stage using first-stage strategy (4B1), number of puzzles solved in first stage using first-stage strategy (4B2), expected number of puzzles solved in first stage (4C1), WTP for full tutorial in first stage (4D1), dummy indicating a strict preference for a tutor in second stage (4D2). Confidence intervals based on robust standard errors (learner-level clusters in 4A3 and 4D2). See also Appendix Table A6.

Result 1: Conservatives perform worse with Black instead of white tutors.

Specifically, conservatives solve 0.55 fewer puzzles when they have Black instead of white tutors (Figure 4A1). They are also 8.6 percentage points less likely to solve at least one puzzle (Figure 4A2). When they do solve a puzzle, they require, on average, about 6.1 additional tile moves (Figure 4A3). These effects are substantial compared to the average performance of conservatives who had white tutors (reported at the bottom of each panel). If paired with Black tutors, conservatives solve 18.8% fewer puzzles, are 10.3% less likely to solve at least one puzzle, and use 8.3% more tile moves per puzzle.

Appendix Figure A5 examines heterogeneity in the effects by the stratification variables. The point estimates tend to be larger among Black learners, highly educated individuals, and those from northern states. Despite this heterogeneity, the sign of the effect is consistent across our sample splits, suggesting that the negative effects of tutor race on performance is robust across groups.

Knowledge acquisition and utilization: Next, we test whether the negative performance effects of having a Black tutor are caused by learners either failing to fully learn the strategies or not applying them effectively. Panel B of Figure 4 and Columns (4) and (5) of Appendix Table A6 summarize the results on strategy adoption. Figure 4B1 studies the extensive margin. Conservatives are 8.4 percentage points less likely to ever adopt a strategy presented by a Black tutor in the first stage (a reduction of about 13%). Figure 4B2, instead, examines the extent of knowledge acquisition and use. It shows that conservatives solve 0.39 fewer puzzles using the recommended strategy when advised by a Black tutor (a reduction of about 20%).[37] We conclude:

Result 2: Conservatives are less likely to learn and utilize knowledge if the advice is provided by a Black rather than a white tutor.

Put differently, conservatives perform worse because they are more likely to disregard advice from Black tutors. Moreover, similar to the performance results, the effects on strategy use are stronger for Black learners, highly educated individuals, and those from northern states (Appendix Figure A5). Again, the point estimates remain negative for most subgroups.

Beliefs: A key question when interpreting these results is whether learners anticipate differences in performance. Panel C of Figure 4 and Column (6) of Appendix Table

[37] A potential concern is that learners might naturally use the proposed first-round strategy even without advice, which would undermine our measure's validity. However, only about 20% of participants who did not receive advice (i.e., those who watched the entertainment video instead of the tutorial) adopted this strategy. Instead, 66% of the participants who watched the first tutorial used it.

A6 examine this topic. In particular, the figures study whether the tutor's race affects learners' expectations about how many puzzles they can solve in five minutes (after watching the first-stage tutorial). We find:

> **Result 3:** Conservatives' beliefs about their own performance are independent of the tutor's race.

Specifically, the tutor's skin color neither impacts learners' average beliefs (Figure 4C1) nor the distribution of their beliefs (Figure 4C2). The lack of distributional differences suggests that nonlinearities do not drive the aggregate null result. Given the reassuring patterns indicating that our belief measures reflect learners' true expectations (see Subsection 2.4), we interpret this result as evidence that learners' beliefs are genuinely independent of the tutor's race. The null result also holds across all subsamples defined by strata variables (Appendix Figure A5).

Tutor selection: We also study how tutor race influences tutor-selection decisions using two different empirical approaches. The first approach is a straightforward treatment comparison: it tests whether being randomly paired with a Black instead of a white tutor affects learners' willingness to pay for the full first-stage tutorial. The second approach analyzes how tutor race impacts learners' explicit tutor choices in the second stage. Specifically, it examines whether learners who could actively choose between Black and white tutors are less likely to indicate that they prefer the Black one.[38]

Panel D in Figure 4 and Columns (7) and (8) of Table A6 present the results. Figure 4D1 reveals that the tutor's race does not affect learners' willingness to pay for the full tutorial. Similarly, Figure 4D2 shows no significant effect of the tutor's race on the probability of preferring a tutor. The probability that learners strictly prefer a white over a Black tutor is about 22%, which is not statistically different from the probability that learners strictly prefer a Black over a white tutor.[39] This finding suggests that

[38]The details are as follows: We construct a panel dataset with two observations per participant—one for each stage's tutor—and restrict our analysis to participants who faced a Black tutor in the first and a white tutor in the second stage (or vice versa). We then estimate the parameters of the following model by OLS: $Y_{is} = \beta_0 + \beta_1 Black_{is} + \beta_2 Black_{is} \times High_i + \beta_3 High_i + X'_{is}\gamma + \varepsilon_{is}$, where Y_{is} is a dummy variable indicating if participant i strictly prefers the stage-s tutor, $Black_{is}$ indicates whether i's tutor in stage s is Black, $High_i$ is a dummy for the high-stakes treatment, and X_{is} are control variables (including indicators for the hand model, the tutor's voice, and the stage). The figure, once again, plots $\widehat{\beta}_1 / E[Y_{is}|Black_{is} = 0, High_{is} = 0]$. We prefer this model over the pre-registered rank-ordered logit model for the ease of interpretation. Both models, hence, rely on identical comparisons and yield very similar results (see Appendix Table K1). Furthermore, due to randomization, the sub-sample of participants exposed to tutors of different skin colors across stages is statistically indistinguishable from the full sample (see Appendix Table A7).

[39]In principle, there are many race-independent reasons why learners may prefer one or the other tutor. Examples include the hand, the voice, or a preference to keep a certain tutor.

conservative learners have no systematic racial bias in their tutor-selection decision. Our next result is:

> **Result 4:** Despite the performance differences, conservatives do not choose white over Black tutors.

For completeness, Appendix Figure A5 shows that there is not much heterogeneity by strata (though some estimates are noisy).

4.2 Behavior Under the High Piece Rate

Next, we test if conservatives' discriminatory behavior vanishes with higher stakes. To that end, we repeat our analysis for conservatives who receive \$1 (high-stakes piece rate) per solved puzzle instead of \$0.20 (standard piece rate). Appendix Figure A9 and Appendix Table A6 show no effect of tutor race on any of our outcomes when the incentives are high. We conclude:

> **Result 5:** Under sufficiently high stakes, conservatives stop discriminating in knowledge acquisition, utilization, and performance.

Appendix Figure A7 shows that there is also limited heterogeneity.

4.3 Further Analyses

This subsection contextualizes our main findings through additional analyses.

Information treatment and mediation analysis: The information treatment allows us to test for statistical (belief-based) discrimination. If learners initially believed tutor quality differed by race, revealing that Black and white tutors provide identical content should eliminate this perceived difference. Consequently, in the information treatment, statistical discriminators should become less likely to discriminate in their tutor choices in the second stage. However, we find that the information treatment does neither affect beliefs nor tutor choices (see Appendix Table G1).[40] This finding is not surprising given that learners' beliefs and tutor choices do not differ by race. Relatedly, note that we pre-specified a mediation analysis in our PAP that measures how much of the racial difference in tutor choices (second stage) operates through beliefs. However, because tutor race has no effect on tutor choices, performing this mediation analysis is irrelevant—there is no effect to mediate. Nevertheless, Appendix Table G1 presents the results.

[40] Second-stage beliefs about own performance do not differ by information treatment neither for Black ($p = 0.33$) nor for white tutors ($p = 0.14$).

Virtue signaling: One potential explanation for the lack of discrimination in tutor choices is that learners refrain from choosing white over Black tutors to signal virtue or avoid appearing discriminatory. However, two observations speak against this explanation. First, our design incentivizes tutor choices. Thus, selecting a less-preferred tutor due to signaling is costly. Second, and perhaps more convincingly, our follow-up study (described in Appendix F) provides clear evidence that conservatives openly engage in discriminatory behavior, even when such behavior is overtly hostile (which is not the case in our main study). This result strongly suggests that learners in our sample do not hesitate to express discriminatory preferences.

Tutor characteristics: Next, we perform two additional analyses regarding the role of the tutors' voice and hand. The first tests whether these factors act as potential confounders. Controlling explicitly for the hand model and tutor's voice, however, leaves all findings unchanged. Given that our treatments are randomized, this finding is no surprise. Our second check examines whether learners respond differently to tutors with matched characteristics (e.g., Black hand paired with Black voice) compared to tutors with mismatched characteristics (e.g., Black hand paired with white voice). Although research suggests that the voices of Black and white speakers do not differ in terms of aerodynamic and acoustic features (see footnote [14]), one might still be concerned that our treatment effects are primarily driven by mismatched tutor-learner combinations. However, we find no systematic evidence for this hypothesis.[41]

Political views: Our data also allows us to explore the heterogeneity across political views in more depth. The first step is to verify that our main results for the broader categories of conservatives closely match those obtained under the pre-registered specification, which restricts the analysis to "very-conservative" individuals (see Appendix Figures A3 and A4). The consistency of the point estimates across the samples suggests that our findings are not only driven by people at the ideological extremes, making our results more broadly applicable. As pre-specified, we also present results for a group that we initially labeled participants with moderate political views. This definition includes individuals who identify as "liberal," "moderate," or "conservative." The analysis predominantly yields null effects (Appendix Figures K1 and K2). Only one coefficient reaches statistical significance, but only due to the inclusion of conservatives and liberals in this group. This pattern supports our decision to group (a) conservatives with very conservatives and (b) liberals with very liberals in

[41]To probe the concern, we regress each first-stage outcome on an indicator for mismatched combinations (and strata controls), separately by political views and skin-color treatments (28 regressions). Only four of the mismatch indicators are significantly different from zero at the 10 percent level.

the main analyses.

Other second-stage outcomes: One might wonder how the race of the second-stage tutor affects strategy-adoption behavior and performance in this stage. We find predominantly imprecise and insignificant effects. In retrospect, this is not surprising as statistical power is much lower. As noted in the PAP, to identify second-stage effects in these dimensions, we must separately estimate treatment effects for two groups: participants who had a Black tutor in the first stage and those who had a white one.[42] This approach halves the sample size relative to the first-stage analysis, making it more difficult to detect meaningful effects.[43] Additionally, because learners already learned an effective strategy in the first stage, it is harder to detect differences in the second stage (particularly in performance outcomes). Discriminators can fall back on the strategy acquired earlier (first strategy), which reduces the contrast between treatment and control. As a result, any additional effect of the second tutor's race on performance is likely smaller and more challenging to isolate. Given these issues, we consider tutor selection to be the most meaningful second-stage outcome.

Multiple hypothesis testing: Because we estimate treatment effects separately for liberals and conservatives, and for each of two piece-rate levels, when considering a given outcome, we adjust for multiple hypothesis testing across the four resulting subgroups. To do so, we apply the method of Barsbai *et al.* (2020), an extension of List *et al.* (2019) for multivariate regression settings. Unlike traditional corrections such as Bonferroni or Holm, this approach accounts for dependence among hypotheses, preserving statistical power. After adjustment, all treatment effects that are significant under the standard piece rate remain so at the 10 percent level, except for the strategy use indicator ($p = 0.134$).

4.4 Discussion

Unconscious discrimination: Conservatives' behavior closely aligns with the key predictions for unconscious discrimination (see Section 3): First, in the standard piece-rate treatment, they perform worse and are less likely to adopt strategies from Black tutors, consistent with an unconscious learning cost ($c^{B,u} > 0$). Second, reflecting a lack of awareness of their bias, conservatives neither anticipate these performance differences nor explicitly prefer white tutors (as documented in *WTP* and explicit

[42] Learners' first-stage experiences may influence their second-stage behavior. Therefore, we conduct second-stage comparisons only within groups that share tutors of the same race in the first stage.

[43] For budgetary reasons, we conducted a pre-test that included only parts of the first stage. Accordingly, our power analysis focused on that stage.

choice). Third, the high stakes treatment eliminates conservatives' discriminatory behavior in strategy adoption and performance, while tutor choices remain unchanged.[44] At lower stakes, unconscious biases are typically stronger, as individuals are less likely to engage deliberately. Instead, the higher piece rate incentivizes more careful consideration, likely helping to overcome these automatic biases. Lastly, the information treatment leaves beliefs and behavior unchanged. This finding naturally follows from the fact that unconscious discriminators already perceive both tutors as equally effective. A treatment confirming content equivalence, thus, does not provide new information.

Alternative explanations: The results contrast sharply with predictions from statistical or taste-based discrimination. Both theories imply that conservatives (a) expect to perform worse with Black tutors and (b) prefer white over Black tutors both under high and low stakes.[45] Our data reject these predictions. The absence of belief updating and behavioral change in response to the information treatment provides further evidence against statistical discrimination.

Conclusion: Our paper identifies unconscious discrimination as the key driver of conservatives' behavior. Hereby, it highlights an often overlooked mechanism through which discrimination harms productivity. Because the bias is unconscious, its effects are likely underestimated and harder to address through conventional policy. Ironically, the documented performance costs may reinforce existing biases: conservatives may misattribute poor outcomes to tutor quality rather than their own bias.[46]

5 Results for Liberals

Our second set of results focuses on learners with liberal political views. Again, we first discuss behavior in the standard piece-rate treatment and focus on the high-stakes treatment in a second step.

5.1 Behavior Under Standard Incentives

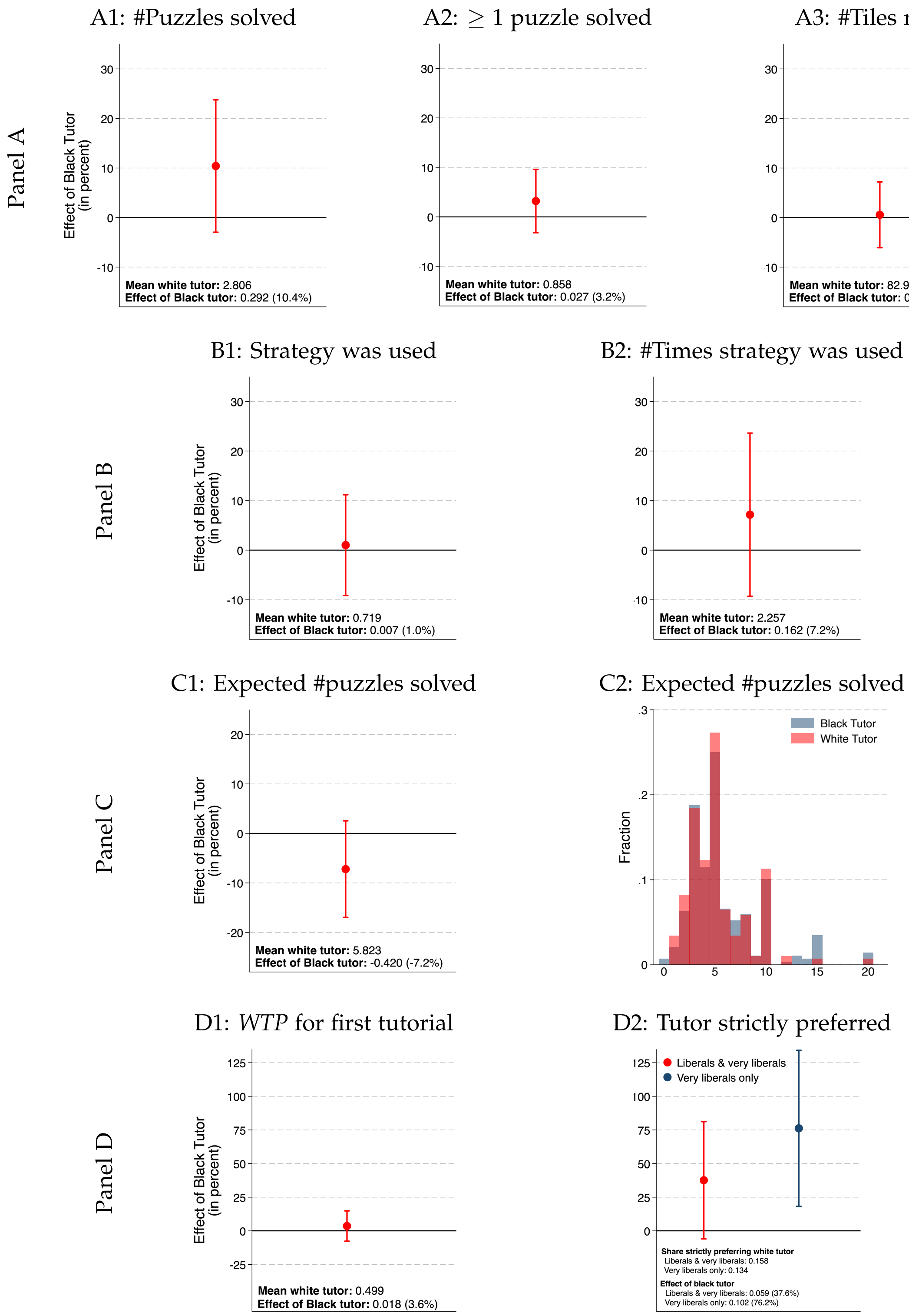
Performance, knowledge acquisition and utilization, and beliefs: Panels A to C in Figure 5 summarize the results for liberals' first-stage outcomes (standard piece-

[44] As unconscious discriminators already expect equal performance under both tutors, they remain indifferent in tutor selection.

[45] Statistical discriminators would still choose white tutors over Black when stakes are high because they expect higher performance. Taste-based discriminators would also favor white tutors due to Black-tutor aversion ($\tau^B > 0$), higher learning costs ($c^B \geq 0$), or perceived lower effectiveness ($\phi^B \leq \phi^W$).

[46] This mechanism is distinct from stereotype threat (Stone *et al.*, 1999; Hoff and Pandey, 2006), where performance declines arise from being discriminated against (not from holding unconscious biases).

Figure 5: Results for Liberals Under Standard Piece Rate



Notes: Effects of tutor race on learners' behavior. Sample: Liberals under standard piece rate. Dependent variables: Number of puzzles solved in first stage (5A1), dummy for learners who solved at least one puzzle in first stage (5A2), average number of moves needed to solve a puzzle in first stage (5A3), dummy for learners who solved at least one puzzle in first stage using first-stage strategy (5B1), number of puzzles solved in first stage using first-stage strategy (5B2), expected number of puzzles solved in first stage (5C1), WTP for full tutorial in first stage (5D1), dummy indicating a strict preference for a tutor in second stage (5D2). Confidence intervals based on robust standard errors (learner-level clusters in 5A3 and 5D2). See also Appendix Table A6.

rate treatment). Columns (1) to (6) in Appendix Table A6 present the corresponding regression results, and Appendix Figure A6 shows heterogeneity by strata variables. The key insight is that the tutor's race does not affect any of the outcomes related to (a) performance, (b) knowledge acquisition and utilization, or (c) beliefs. We conclude:

> **Result 6:** Liberals learn and apply strategies equally well with Black and white tutors, and their performance does consequently not depend on the tutor's race. They also do not expect performance differences.

This finding holds across all the considered samples (see Appendix Figure A3).

Tutor selection: Next, we use our two approaches to study how the tutor's race impacts tutor selection among liberals. Panel D in Figure 5 presents the key results, Columns (7) and (8) of Appendix Table A6 the corresponding regression tables, and Appendix Figure A6 results from heterogeneity analyses by strata.

The first insight from Panel D in Figure 5 concerns learners' explicit tutor choices in the second stage: Liberal learners tend to favor Black over white tutors. To demonstrate this insight, Figure 5D2 shows separate estimates for our baseline sample of liberals (those identifying as "liberal" or "very liberal") in red and for the subgroup of "very liberals" alone in blue. In the baseline sample, the probability that learners strictly prefer the white over the Black tutor is 15.8%. This contrasts with a a probability of 21.7% for strictly preferring the Black over the white tutor (an increase of 5.9 percentage points or approximately 37.6%). The estimate is a bit noisy and only significant at the 10% level (p-value$= 0.091$).[47] All the specifications in Appendix Figure A3 confirm this finding at a significance level of at least 10%. Notably, the separate estimate for very liberal learners (in red) suggests they have a stronger preference for Black tutors. Among them, the probability of strictly preferring the Black tutor is 76.2% higher than that of strictly preferring the white one (p-value$= 0.010$). Although the confidence interval is rather wide, we can confidently rule out effects smaller than about 20%. We conclude that tutor selection is the one dimension where we find substantial heterogeneity between liberal and very liberal participants.

The second insight from Panel D is that, while liberals tend to choose Black over white tutors, they are not willing to bet money on securing them. Specifically, we find no statistically significant difference in the average *WTP* for the first-stage tutorial between Black and white tutors (see Figure 5D1). This result holds even for very liberal learners (see Appendix Figure A3). Additionally, we cannot reject the null hypothesis of the standard Kolmogorov-Smirnov test that the *WTP*s for Black and white tutors come from the same distribution ($p = 0.395$). This result suggests that the effects are

[47]The effects are stronger for white learners and those with lower education (Appendix Figure A6).

also unlikely to be concentrated in specific parts of the *WTP* distribution (e.g., ceiling effects). In sum, our results indicate that:

> **Result 7:** Very liberal learners prefer Black over white tutors, yet are unwilling to pay to secure a Black tutor.

5.2 Behavior Under the High Piece Rate

Appendix Figure A10 and Appendix Table A6 show that across all studied outcomes, the behavior of liberals is insensitive to the piece-rate treatment. Specifically, also under higher stakes, liberals learn equally well from Black and white tutors, show no performance differences, and also do not expect such differences. However, they strictly prefer Black tutors in the second stage. This result is now consistently significant at the 5% level (see Appendix Figure A10 and all specifications in Appendix Figure A4). Lastly, the effect heterogeneity across strata closely mirrors that in the standard piece-rate treatment, though the estimates are somewhat noisy. We conclude:

> **Result 8:** The high-stakes treatment does not affect the behavior of liberal learners.

The results are qualitatively similar across subsamples defined by strata (see Appendix Figure A10).

5.3 Further Analyses

Again, we present the results of further analyses.

Attention to treatment manipulation: One concern about the null results in the first-stage outcomes is that they may reflect participants not paying attention to the treatment manipulation. However, this seems unlikely. In the final survey, 86% of liberal participants correctly recall the tutor's skin color. Moreover, liberals are somewhat more likely than conservatives to correctly recall the tutors' skin color: The difference is four percentage points for Black tutors (p-value $= 0.1$) and two points for white tutors, even if this difference does not reach statistical significance (p-value $= 0.32$). Lastly, the significant effects observed at the tutor-selection margin further challenge this hypothesis.

Information treatment and mediation analysis: Why do liberals prefer Black tutors when given a choice? One explanation works through beliefs. For example, liberals may perceive Black tutors to be positively selected and, therefore, expect better performance with them. Another possibility is that, for various reasons such as a desire for inclusivity or virtue signaling, liberals might derive utility from selecting Black tutors ($\phi_b > 0$). In this case, they would even select Black over white tutors if they expected them to be of similar quality and/or anticipated a similar performance under both types. Our information treatment allows us to separate these two explanations by isolating the role of beliefs. Specifically, as detailed in Appendix G, we follow Imai *et al.* (2011) and conduct a formal mediation analysis.[48] This approach allows us to decompose the effect of tutor race on tutor choices into (a) an indirect effect operating through learners' beliefs about their performance and (b) a direct effect capturing other factors. We find that the indirect effect is small and statistically insignificant (see Appendix Table G1). Put differently, liberals continue to prefer Black tutors even when they hold identical expectations. This finding suggests that preference-related factors beyond beliefs—such as a utility obtained from selecting Black tutors—drive choices.

Tutors characteristics and second-stage outcomes: As for conservatives, we analyze the role of tutor characteristics and study impacts on other second-stage outcomes. The take-away messages are very similar. First, our results remain unchanged if we control for tutor characteristics (such as voice and hand). Learners do also not respond differently to matched or mismatched tutor characteristics (see footnote [41]). Second, the second-stage outcomes show the same imprecise and statistically insignificant effects for liberals as for conservatives. Given the reduced sample size and limited scope for strategy adoption in this stage, these results are unsurprising.

Multiple hypothesis testing: The second-stage finding of a strict preference for Black tutors under the high piece-rate condition remains statistically significant ($p = 0.08$) after adjusting for multiple hypothesis testing (see Section 4.3 for details).

5.4 Discussion

The key finding for liberals is that they exhibit *Black-favoring behavior* in tutor selection (second stage), while showing no discrimination in learning. They acquire and apply knowledge equally well from Black and white tutors, experience no performance differences, and do not anticipate a different performance across tutors. These patterns are consistent with a version of our model where learners (a) do not expect Black tutors to be better teachers ($\phi^B = \phi^W$), (b) do not exhibit conscious or unconscious biases that affect strategy adoption ($c^B = c^W$ and $c^{B,u} = c^{W,u}$), and (c) derive greater utility

[48]We estimate the mediation effects using a two-stage least squares (2SLS) approach. In the first stage, we use the skin-color treatment to determine the effect of having a Black tutor on beliefs and the information treatment as an instrument to generate exogenous variation in beliefs. This exogenous variation is crucial for identifying how beliefs affect tutor choice in the second stage, where we regress tutor choices on both the predicted beliefs and the skin-color treatment.

from selecting Black tutors independent of performance expectations ($-\tau^B < -\tau^W$). Subsequently, we discuss why behavior is in line with this model.

Alternative explanations: We first examine whether mechanisms beyond preference-based explanations can account for our results. A natural first candidate is *statistical Black-favoring behavior*, meaning learners base their choice of Black tutors on expectations of superior teaching effectiveness ($\phi^B > \phi^W$). If this were the case, they should anticipate better learning outcomes with Black tutors, and we would expect to see an increase in learning engagement and performance when assigned to them. However, our findings contradict these predictions. Additionally, our mediation analysis confirms that tutor selection is not driven by belief-based mechanisms. These results indicate that their preference for Black tutors is not rooted in expected instructional quality. Another explanation is that liberals have a *conscious learning bias*, meaning they learn more easily or efficiently when taught by a Black tutor ($c^B < c^W$). Liberals who recognize this bias should also be more likely to select Black over white tutors to maximize their performance. If this were the case, we should also observe differences in beliefs, strategy adoption, and performance—which we do not. We can also reject the hypothesis of unconscious biases ($c^{B,u} < c^{W,u}$). Learners with such biases would not select Black over white tutors (as their bias is unconscious).

Symbolic preferences in tutor selection: There is no evidence that liberals select Black tutors based on expected performance or biases. Instead, their behavior is more consistent with another explanation within our model: liberals might gain utility from the act of selecting Black tutors itself, independent of performance expectations or biases. The nature of our setting provides further insight into this behavior, allowing us to better understand its underlying motivations. One important aspect of our setting is that tutor selection does not affect the tutors' outcomes—such as employment or financial benefits. Thus, motives such as a desire to support Black tutors cannot explain the Black-favoring behavior. Moreover, as previously discussed, liberals also do not seem to make this choice to improve their own outcomes. Together, these two findings indicate that they derive *non-instrumental utility* from tutor selection, meaning their preference is symbolic rather than outcome-driven. Several motives can explain this utility, including identity reinforcement, a desire to demonstrate one's moral or social values (virtue signaling), or ideological alignment with inclusivity.[49]

[49] Virtue signaling and ideological alignment are related but distinct concepts. Virtue signaling involves expressing moral or social values to gain approval, enhance status, or conform to norms, often in public settings. By contrast, ideological alignment reflects a genuine commitment to values driven by internal consistency rather than external recognition. For example, someone might choose a Black tutor publicly to signal their commitment to diversity (virtue signaling). Another person might make the same choice privately as it aligns with their beliefs, even if no one else is aware (ideological alignment).

Moreover, the fact that liberals do not exhibit a higher first-stage *WTP* for Black tutors suggests that this preference is context-dependent. While they actively choose Black tutors when selection carries no financial cost, this preference is not dominating their choice when monetary trade-offs are involved. This pattern indicates that the utility they gain from selecting Black tutors is not strong enough to persist in salient financially consequential decisions, reinforcing the idea that their behavior is driven by symbolic preferences.[50] The fact that liberals continue to select Black tutors even under higher stakes further aligns with this interpretation. Liberals do not expect worse performance with Black tutors, so they can choose them as a symbolic act without incurring financial costs.

6 Concluding Remarks

Using novel experimental evidence, we examine the role of racial discrimination against Black tutors throughout the whole learning process. Our first set of results focuses on learners with conservative political views. We find that these learners engage in unconscious discrimination against Black tutors. Specifically, conservatives are less likely to acquire and use knowledge from Black than white tutors, ultimately harming their performance. However, consistent with unconscious discrimination, they do not anticipate any performance difference. Acting on this belief, conservatives also do not discriminate in tutor selection. The adverse effects on learning, strategy adoption, and performance disappear under a treatment that substantially increases performance incentives, likely by encouraging more reflexive thinking. All the findings are in line with a simple model of unconscious discrimination. Our second set of results concerns liberal learners. These individuals tend to exhibit explicit Black-favoring behavior in tutor selection, regardless of the size of incentives. However, they learn and perform equally with Black and white tutors and do not anticipate performance differences. A version of our model in which liberals gain non-instrumental utility from the act of selecting Black tutors itself can explain this behavior.

From a broader perspective, our paper is the first to demonstrate that discrimination against information providers can shape knowledge acquisition, use, and productivity on the demand side of the information market. This finding stands in contrast to prior work on racial homophily, which highlights the benefits of shared racial identity (Dee, 2004; Egalite *et al.*, 2015; Lusher *et al.*, 2016; Alsan *et al.*, 2019; Kofoed and mcGovney, 2019; DiBartolomeo *et al.*, 2023). Importantly, the documented adverse

[50]Motives such as virtue signaling and social preferences often diminish when financial costs are introduced (Bénabou and Tirole, 2006; Hillman, 2010; Gneezy *et al.*, 2011). These findings suggest that individuals engage in symbolic actions when they are low-cost but abandon them when trade-offs exist.

productivity effects could unintentionally reinforce peoples' existing biases. For example, individuals experiencing poorer performance may wrongly attribute their outcomes to Black tutors rather than recognizing their own biases. Over time, these misattributions could create demand-side pressure, partially shaping who gets hired, promoted, or retained in advisory roles. Consistent with this idea, Black teachers are underrepresented in conservative U.S. school districts, where racial biases are typically stronger (see Appendix Figure A1). African Americans also represent less than 4% of financial advisors, 6% of medical doctors and news anchors, and 7% of teachers despite comprising 14% of the population (Center for Financial Planning, 2018; Association of American Medical Colleges, 2014; American Society of News Editors, 2018).

Conceptually, our study introduces several methodological innovations to advance the study of discrimination in information-based settings. First, we develop a two-stage experimental design that enables us to study discrimination across the entire learning process. A key innovation of our design is that it addresses several identification problems simultaneously. For example, it allows us to isolate the causal effect of tutor race on both selection and performance in a setting where learners choose tutors before receiving and applying any advice in a task. Second, we propose new ways to disentangle different forms of discrimination. To do so, we combine belief measures, an information treatment, and multi-dimensional behavioral outcomes within an incentivized experimental environment. Third, we employ post-production video editing techniques to manipulate the tutor's skin color exogenously. This approach allows us to signal tutor race naturally, avoiding potentially confounding signals such as names (Kreisman and Smith, 2023). Recent advances in AI have made such video-based manipulations more accessible and scalable. As a result, these methods offer a powerful complement to established approaches, such as image manipulations (Evsyukova *et al.*, 2025), body-weight manipulations (Macchi, 2023), or manipulations of facial expressions (Albohn *et al.*, 2022; Peterson *et al.*, 2022).

Future research: In conclusion, our study marks an initial step in unpacking how discrimination operates in learning contexts, but much work remains to explore its full implications. One key insight of our paper is that conservative learners remain unaware of how their biases shape learning outcomes. Importantly, we study a setting where learners engage with entirely new content, leaving them without prior knowledge to guide expectations. Future research should assess whether learners remain unaware of their biases in settings where they have prior knowledge and, thus, might find it easier to anticipate the productivity costs of discrimination. If awareness does emerge, researchers could also examine the broader conditions under which learners recognize their biases. Moreover, it would be important to investigate whether awareness of biases in learning leads individuals to discriminate at the selection stage. More broadly, it would be interesting to explore whether our effects persist in other domains of learning and among other populations.

Another important goal for future research should be exploring the full implications of discrimination for tutors. Our results suggest that learners do not always engage equally with Black an white tutors. While such decisions may seem inconsequential for minority tutors in isolation, they could cumulatively disadvantage them, particularly in early stages of their careers. If low-stakes advice-seeking decisions, for example, disadvantage outgroup tutors early in their careers, these biases may have longer-term consequences for the advisors' career trajectories (similar to Bohren *et al.*, 2019). Investigating whether and how these (early-career) frictions shape tutors' long-term outcomes, and identifying strategies to mitigate them remains a promising direction for future work.

Lastly, our study highlights that increasing stakes can eliminate discriminatory behavior, suggesting that unconscious biases are not fixed. This finding raises important questions for future research: What types of policies, incentives, or decision environments most effectively reduce bias? How robust are their effects across contexts? Our theoretical considerations suggest that we can and should foster reflective thinking to reduce unconscious bias (Bertrand *et al.*, 2005; Kahneman, 2011). However, this finding is only a starting point for the design of effective interventions. We conclude that developing and testing targeted interventions—across ideological groups and addressing both conscious and unconscious bias—remains essential for creating more inclusive learning environments and expanding opportunity for all.

References

Acemoglu, D. and Pischke, J.-S. (1998). Why Do Firms Train? Theory and Evidence. *Quarterly Journal of Economics*, **113** (1), 79–119.

Albohn, D. N., Uddenberg, S. and Todorov, A. (2022). A Data-Driven, Hyper-Realistic Method for Visualizing Individual Mental Representations of Faces. *Frontiers in Psychology*, **13**, 997498.

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.

Arrow, K. J. (1962). The Economic Implications of Learning by Doing. *Review of Economic Studies*, **29** (3), 155–173.

Association of American Medical Colleges (2014). Diversity in the Physician Workforce: Facts & Figures 2014, available at: http://www.aamcdiversityfactsandfigures.org/.

Barsbai, T., Licuanan, V., Steinmayr, A., Tiongson, E. and Yang, D. (2020). *Information and the Acquisition of Social Network Connections*. NBER Working Paper 27346, National Bureau of Economic Research.

Bartos, V., Bauer, M., Cahlikova, J. and Chytilova, J. (2021). Covid-19 Crisis and Hostility Against Foreigners. *European Economic Review*, **137**, 103818.

—, —, 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. *Behavioral Science*, **9** (3), 226–232.

Becker, G. S. (1957). *The Economics of Discrimination*. Chicago: University of Chicago Press.

Bell, A., Chetty, R., Jaravel, X., Petkova, N. and Van Reenen, J. (2019). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. *Quarterly Journal of Economics*, **134** (2), 647–713.

Ben-Porath, Y. (1967). The Production of Human Capital And the Life Cycle of Earnings. *Journal of Political Economy*, **75** (4), 352–365.

Bénabou, R. and Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, **96** (5), 1652–1678.

Bertrand, M., Chugh, D. and Mullainathan, S. (2005). Implicit Discrimination. *American Economic Review*, **95** (2), 94–98.

— and Duflo, E. (2017). Field Experiments on Discrimination. In A. Banerjee and E. Duflo (eds.), *Handbook of Economic 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.

Bloom, N., Sadun, R. and Reenen, J. V. (2012). Americans do IT Better: US Multinationals and the Productivity Miracle. *American Economic Review*, **102** (1), 167–201.

Bohren, J. A., Haggag, K., Imas, A. and Pope, D. G. (2023). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics*, pp. 1–45.

—, Imas, A. and Rosenberg, M. (2019). The Dynamics of Discrimination: Theory and Evidence. *American Economic Review*, **109** (10), 3395–3436.

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/.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. and Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, **126** (4), 1593–1660.

—, — and Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, **104** (9), 2633–2679.

—, —, Saez, E., Turner, N. and Yagan, D. (2020). Income Segregation and Intergenerational Mobility Across Colleges in the United States. *Quarterly Journal of Economics*, **135** (3), 1567–1633.

Correll, J., Park, B., Judd, C. M. and Wittenbrink, B. (2002). The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals. *Journal of Personality and Social Psychology*, **83** (6), 1314–1329.

Coutts, A. (2023). Racial Bias Around the World, Mimeo, Schulich School of Business, York University.

de Benedictis-Kessner, J., Lee, D. D. I., Velez, Y. R. and Warshaw, C. (2023). American Local Government Elections Database. *Scientific Data*, **10** (1), 912.

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

Devine, P. (1989). Stereotypes and Prejudice: Their Automatic and Controlled Components. *Journal of Personality and Social Psychology*, **56** (1), 5.

DiBartolomeo, J. A., Kothakota, M. G., Parks-Stamm, E. J. and Tharp, D. T. (2023). Racial Animosity and Black Financial Advisor Underrepresentation. *Financial Planning Review*, **6** (2), e1164.

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.

Evsyukova, Y., Rusche, F. and Mill, W. (2025). LinkedOut? A Field Experiment on Discrimination in Job Network Formation. *Quarterly Journal of Economics*, **140** (1), 283–334.

Gneezy, U., Meier, S. and Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, **25** (4), 191–210.

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), 1464–1480.

Haaland, I. and Roth, C. (2023). Beliefs About Racial Discrimination and Support for Pro-Black Policies. *Review of Economics and Statistics*, **105** (1), 40–53.

Hauser, D. J., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J. and Litman, L. (2023). Evaluating CloudResearch's Approved Group as a Solution for Problematic Data Quality on MTurk. *Behavior Research Methods*, **55** (8), 3953–3964.

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

Hillman, A. L. (2010). Expressive Behavior in Economics and Politics. *European Journal of Political Economy*, **26** (4), 403–418.

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

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.

Isaksson, S. (2018). It Takes Two: Gender Differences in Group Work, mimeo, November 20, 2018.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, **93** (5), 1449–1475.

— (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kinnl, K., Möller, J. and Walter, A. (2023). *Borrowed Plumes: The Gender Gap in Claiming Credit for Teamwork*. WU Working Paper Series 345.

Kofoed, M. S. and mcGovney, E. (2019). The Effect of Same-Gender or Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point. *Journal of Human Resources*, **54** (2), 430–467.

Kreisman, D. and Smith, J. (2023). Distinctively Black Names and Educational Outcomes. *Journal of Political Economy*, **131** (4), 877–897.

Lang, K. and Lehmann, J.-Y. K. (2012). Racial Discrimination in the Labor Market: Theory and Empirics. *Journal of Economic Literature*, **50** (4), 959–1006.

— and Spitzer, A. K.-L. (2020). Race Discrimination: An Economic Perspective. *Journal of Economic Perspectives*, **34** (2), 68–89.

List, J., Shaikh, A. M. and Xu, Y. (2019). Multiple Hypothesis Testing in Experimental Economics. *Experimental Economics*, **22**, 773–793.

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.

Macchi, E. (2023). Worth Your Weight: Experimental Evidence on the Benefits of Obesity in Low-Income Countries. *American Economic Review*, **113** (9), 2287–2322.

Meckler, L. and Rabinowitz, K. (2019). America's Schools Are More Diverse Than Ever. But the Teachers Are Still Mostly White, *Washington Post*, Dec. 27.

Mokyr, J. (2011). *The Gifts of Athena: Historical Origins of the Knowledge Eeconomy*. Princeton University Press.

Mosleh, M., Pennycook, G. and Rand, D. G. (2024). Differences in Misinformation Sharing Can Lead to Politically Asymmetric Sanctions. *Nature*, **615**, 1–6.

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

Nosek, B. A., Banaji, M. R. and Jost, J. T. (2009). The Politics of Intergroup Attitudes. In J. Jost, A. Kay and H. Thorisdottir (eds.), *Social and Psychological Bases of Ideology and System Justification*, Oxford University Press, pp. 480–506.

—, Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G. and Banaji, M. R. (2007). Pervasiveness and Correlates of Implicit Attitudes and Stereotypes. *European Review of Social Psychology*, **18** (1), 36–88.

Onuchic, P. (2023). *Recent Contributions to Theories of Discrimination*. arxiv:2205.05994.

Peterson, J. C., Uddenberg, S., Griffiths, T. L., Todorov, A. and Suchow, J. W. (2022). Deep Models of Superficial Face Judgments. *Proceedings of the National Academy of Sciences*, **119** (17), e2115228119.

Pew Research Center (2021). Deep Divisions in Americans' Views of the Nation's Racial History - and How To Address It, Research Report, August 2021.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review*, **62** (4), 659–661.

Sapienza, C. M. (1997). Aerodynamic and Acoustic Characteristics of the Adult AfricanAmerican Voice. *Journal of Voice*, **11** (4), 410–416.

Stagnaro, M. N., Druckman, J., Arechar, A. A., Willer, R., Rand, D. *et al.* (2024). Representativeness Versus Attentiveness: A Comparison Across Nine Online Survey Samples, PsyArXiv, 22.

Stone, J., Lynch, C. I., Sjomeling, M. and Darley, J. M. (1999). Stereotype Threat Effects on Black and White Athletic Performance. *Journal of Personality and Social Psychology*, **77** (6), 1213.

Toussaert, S. (2018). Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment. *Econometrica*, **86** (3), 859–889.

Online Appendix to

The Color of Knowledge: Impacts of Tutor Race on Learning and Performance

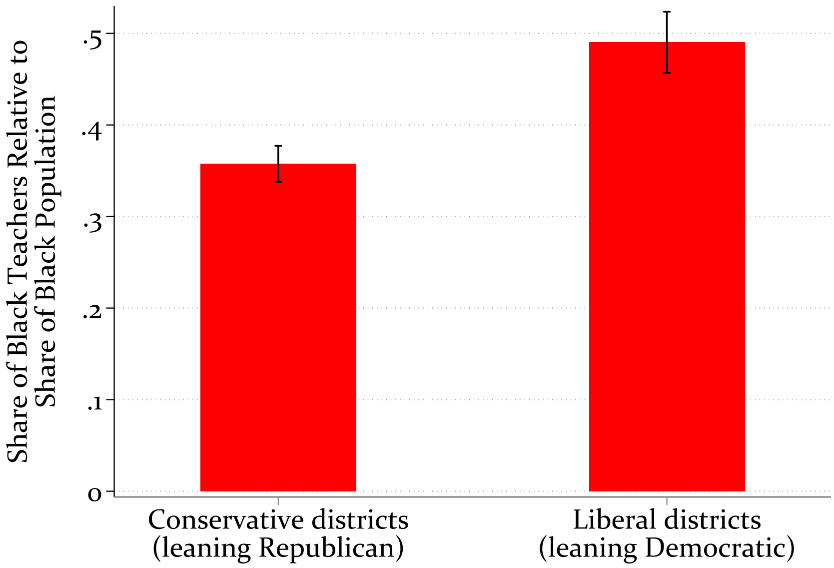
Vojtěch Bartoš and Ulrich Glogowsky and Johannes Rincke

May 21, 2025

*Bartoš: University of Milan, Department of Economics, Management, and Quantitative Methods, Via Conservatorio 7, 20122 Milano, Italy, vojtech.bartos@unimi.it. Glogowsky: Johannes Kepler University Linz, Department of Economics, Altenbergerstrasse 69, 4040 Linz, Austria, ulrich.glogowsky@jku.at. Rincke: University of Erlangen-Nuremberg, School of Business, Economics, and Society, Lange Gasse 20, 90403 Nuremberg, Germany, johannes.rincke@fau.de. The authors would like to thank Michal Bauer, Stefan Bauernschuster, Tobias Cagala, Alexander Cappelen, Julie Chytilová, Hajdu Gergely, Matthias Fahn, Susann Fiedler, Alex Imas, Yves Le Yaouanq, Jakob Möller, Elizabeth Levy Paluck, Alexander Todorov, Ricardo Perez-Truglia, and Bertil Tungodden for helpful comments. Seminar participants at various places provided valuable feedback. We gratefully acknowledge funding from the German Research Foundation through a research grant (project 444754857) and through CRC TRR 190 (project 280092119). Bartoš acknowledges financial support within the "Fund for Departments of Excellence academic funding" provided by the Ministero dell'Universita e della Ricerca (MUR), established by Stability Law, namely "Legge di Stabilita n.232/2016, 2017" - Project of the Department of Economics, Management, and Quantitative Methods, University of Milan. Glogowsky gratefully acknowledges financial support from the Linz Institute for Transformative Change (C-2022-1-118). This experiment was pre-registered at the AEA Social Science Registry as AEARCTR-0005812. Updates: AEARCTR-0007737, AEARCTR-0008563, AEARCTR-0009715. We obtained IRB approval from the Ethics Commission of the School of Business and Economics at the University of Erlangen-Nuremberg. Ülkü Bıçakçı, Irina Fux, and David Gstrein provided valuable research assistance.

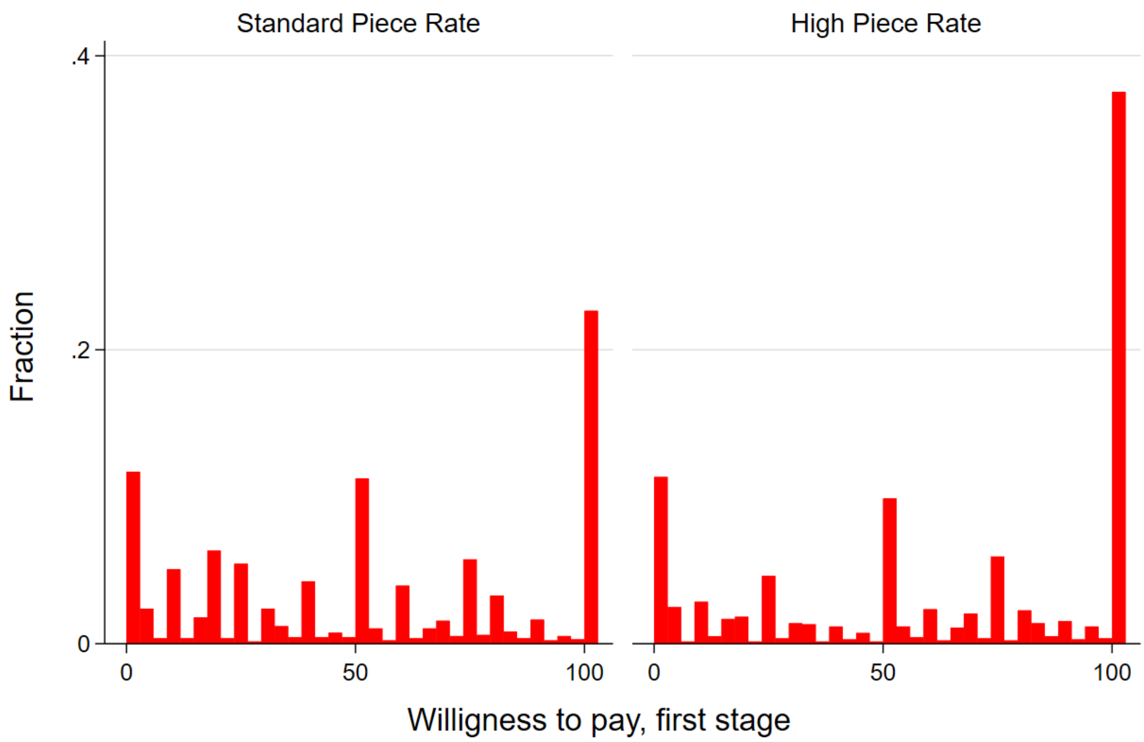
A Appendix Figures

Figure A1: Relative Under-Representation of Black Teachers



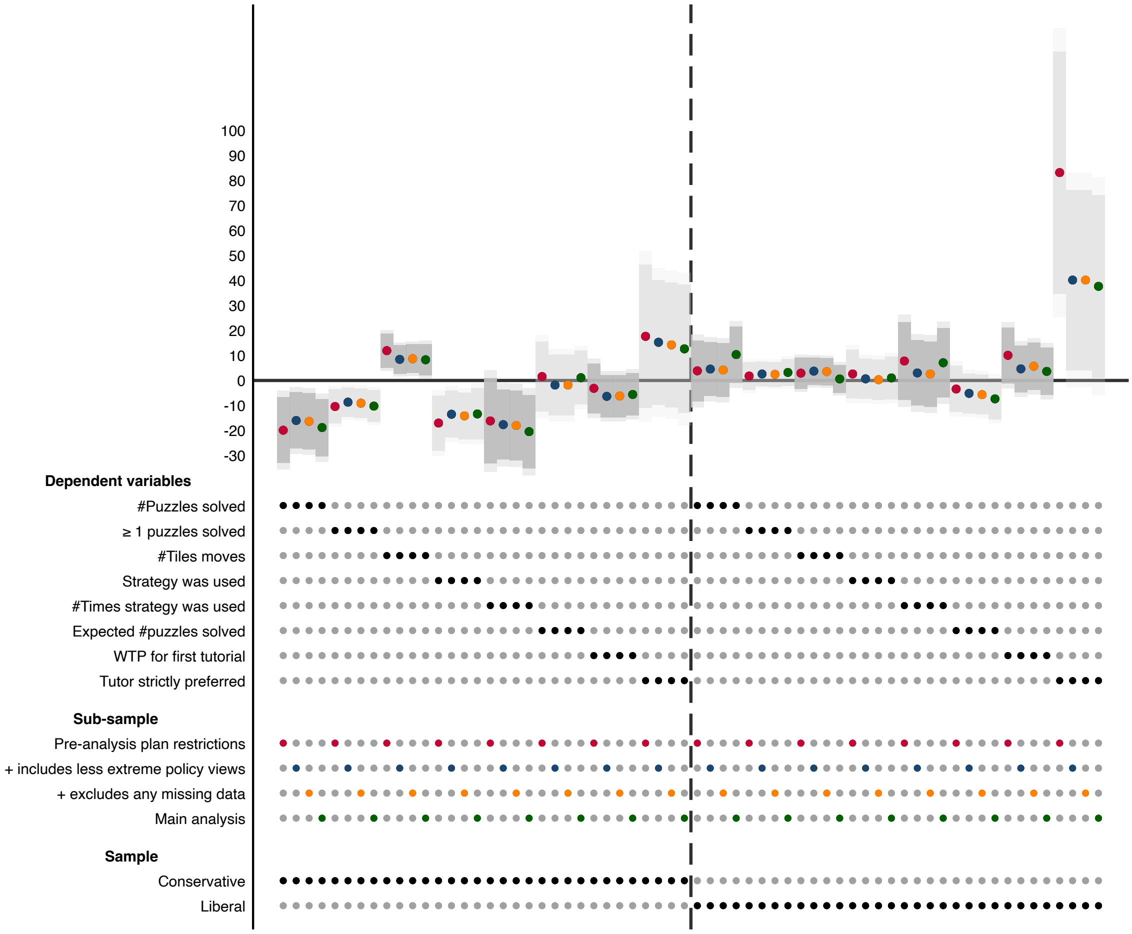
Notes: District-level share of Black teachers relative to the population share of Blacks in the respective district. The left (right) bar presents the mean for below-median (above median) school districts by democratic vote share in the 2016 presidential election. For most districts, data on teachers is from 2017-18. Sample: School districts with more than 50,000 inhabitants ($N = 1{,}121$). Sources: de Benedictis-Kessner *et al.* (2023) and data repository of Meckler and Rabinowitz (2019).

Figure A2: First-Stage Willingness to Pay: Distribution of Choices by Piece Rate



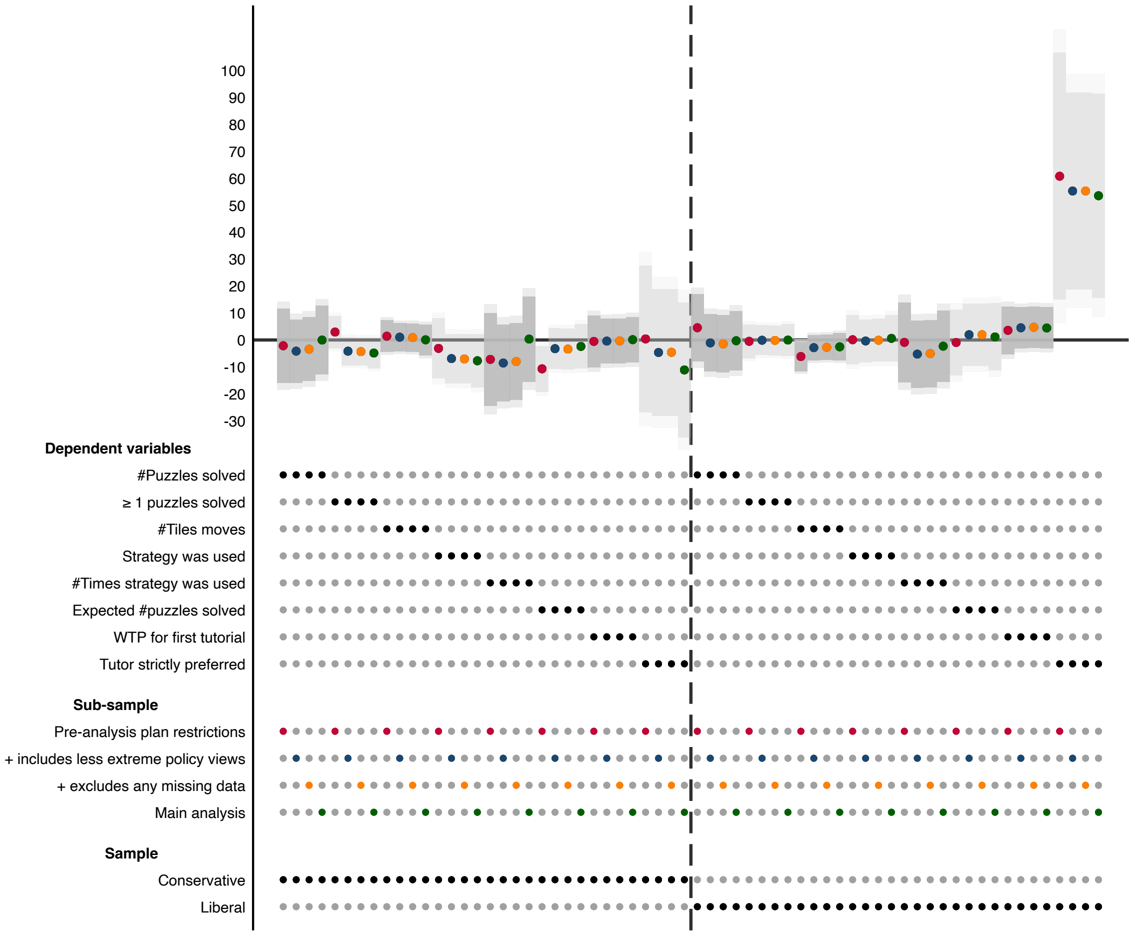
Notes: Histogram. Conservatives and liberals samples pooled. Split by piece-rate treatment.

Figure A3: Main Results under Standard Piece Rate: Stability by Sample Restrictions



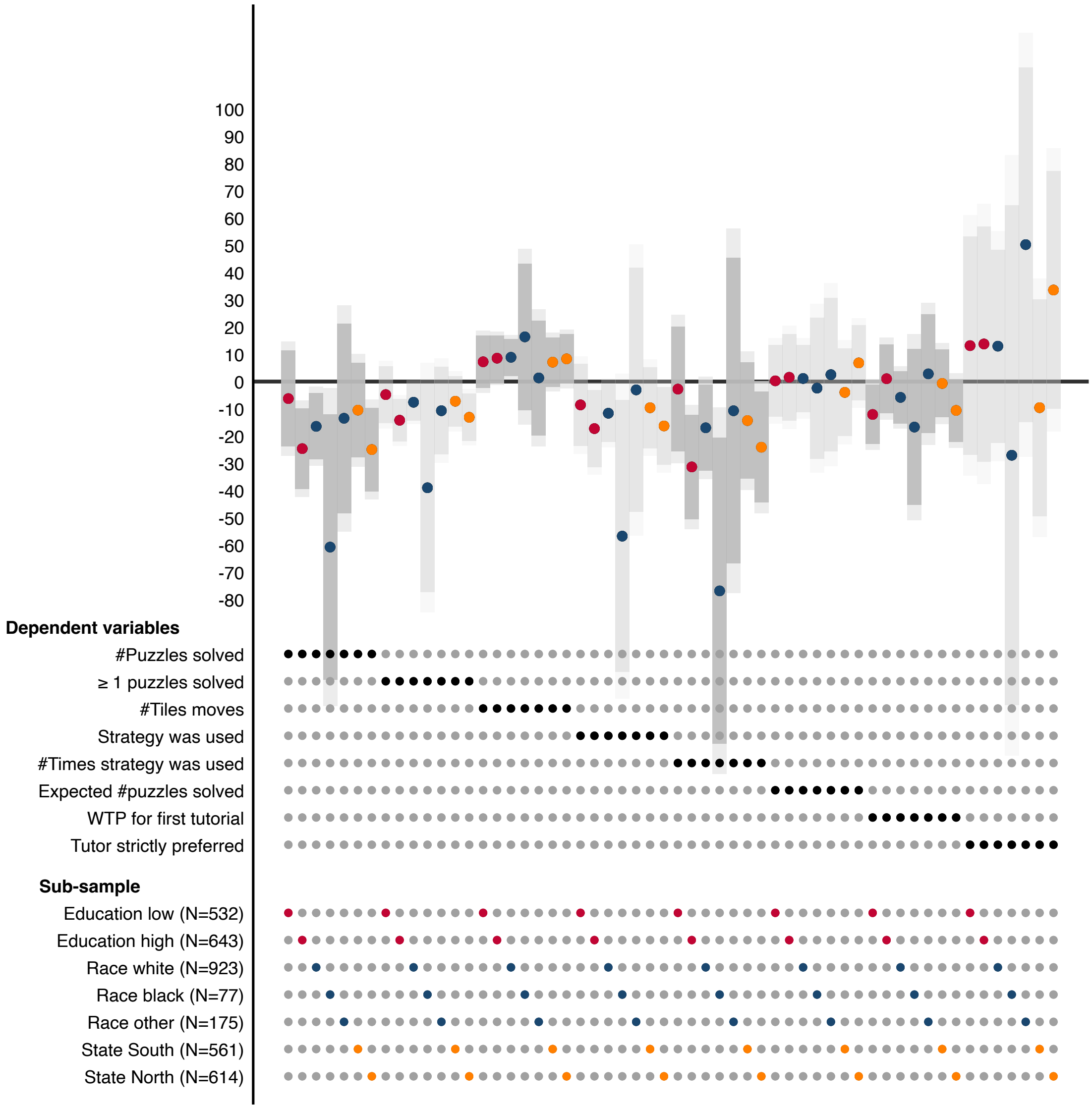
Notes: Effects of tutor race on learners' behavior. Specifications as in Figures 4 and 5, by sample restriction criteria. Dependent variables described just below the estimates: (1) number of puzzles solved in first stage, (2) dummy for learners who solved at least one puzzle in first stage, (3) average number of moves needed to solve a puzzle in first stage, (4) dummy for learners who solved at least one puzzle in first stage using first-stage strategy, (5) number of puzzles solved in first stage using first-stage strategy, (6) expected number of puzzles solved in first stage, (7) WTP for full tutorial in first stage, and (8) dummy indicating strict preference for a tutor in second stage. Markers show the estimated effects, the darker or lighter whiskers denote 90 and 95 percent confidence intervals based on robust standard errors (learner-level clusters in (3)), respectively. Overall sample: Conservatives and liberals under standard piece rate, separated by the dashed vertical line, respectively. Sample restriction criteria sub-samples described in the middle part (in different colors for readability). In red, Pre-Analysis Plan sampling restrictions. In blue, additionally including individuals with less extreme policy views (i.e., merging very conservative and conservative, and very liberal and liberal). In orange, further excluding individuals for whom the website did not record all the data. In green, main analysis sample. Table A4 reports numbers of observations by sample restriction.

Figure A4: Main Results under High Piece Rate: Stability by Sample Restrictions



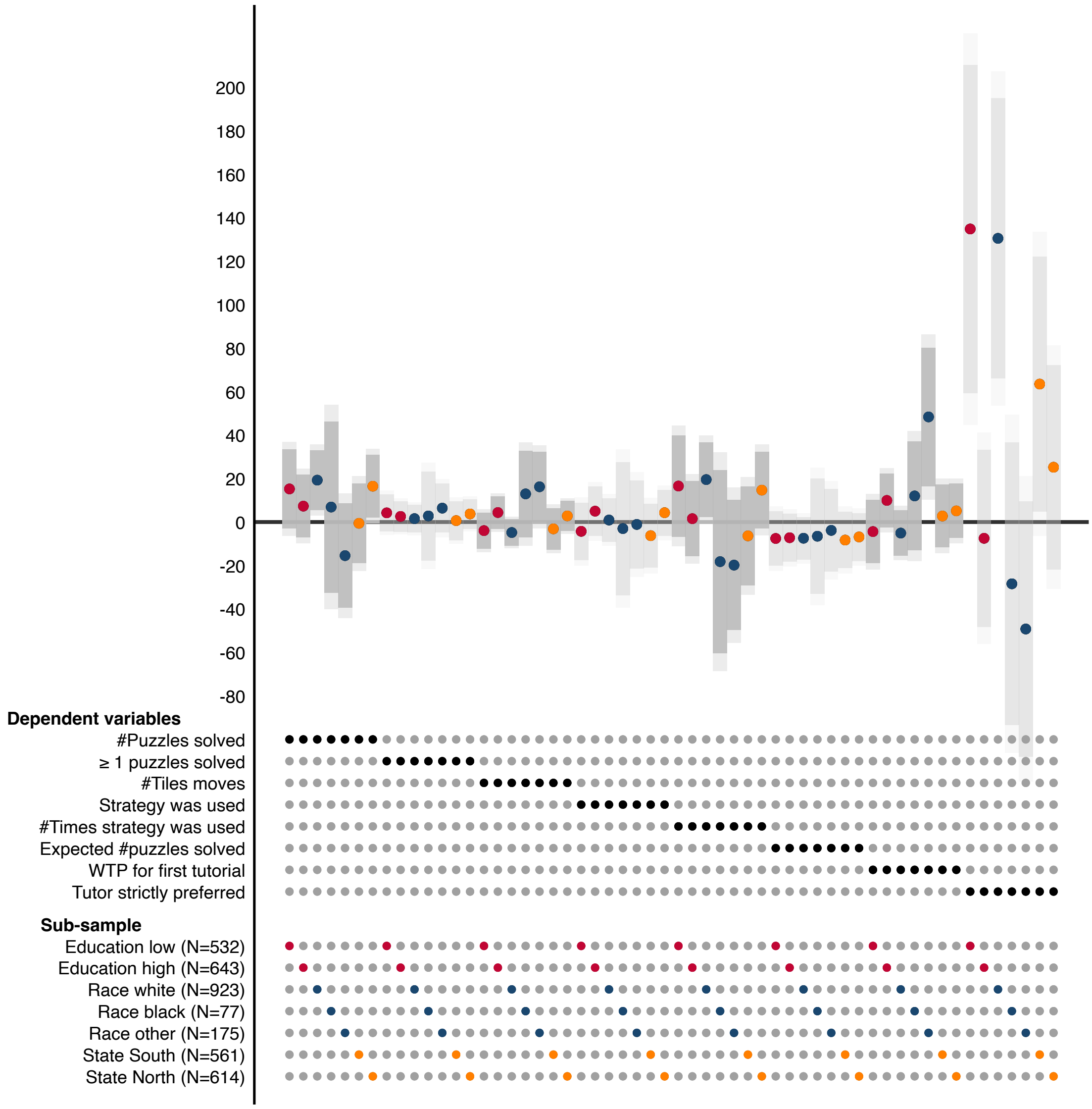
Notes: Effects of tutor race on learners' behavior. Specifications as in Figures 4 and 5, by sample restriction criteria. Dependent variables described just below the estimates: (1) number of puzzles solved in first stage, (2) dummy for learners who solved at least one puzzle in first stage, (3) average number of moves needed to solve a puzzle in first stage, (4) dummy for learners who solved at least one puzzle in first stage using first-stage strategy, (5) number of puzzles solved in first stage using first-stage strategy, (6) expected number of puzzles solved in first stage, (7) WTP for full tutorial in first stage, and (8) dummy indicating strict preference for a tutor in second stage. Markers show the estimated effects, the darker or lighter whiskers denote 90 and 95 percent confidence intervals based on robust standard errors (learner-level clusters in (3)), respectively. Overall sample: Conservatives and liberals under standard piece rate, separated by the dashed vertical line, respectively. Sample restriction criteria sub-samples described in the middle part (in different colors for readability). In red, Pre-Analysis Plan sampling restrictions. In blue, additionally including individuals with less extreme policy views (i.e., merging very conservative and conservative, and very liberal and liberal). In orange, further excluding individuals for whom the website did not record all the data. In green, main analysis sample. Table A4 reports numbers of observations by sample restriction.

Figure A5: Results for Conservatives Under Standard Piece Rate: Heterogeneity by Strata Variables



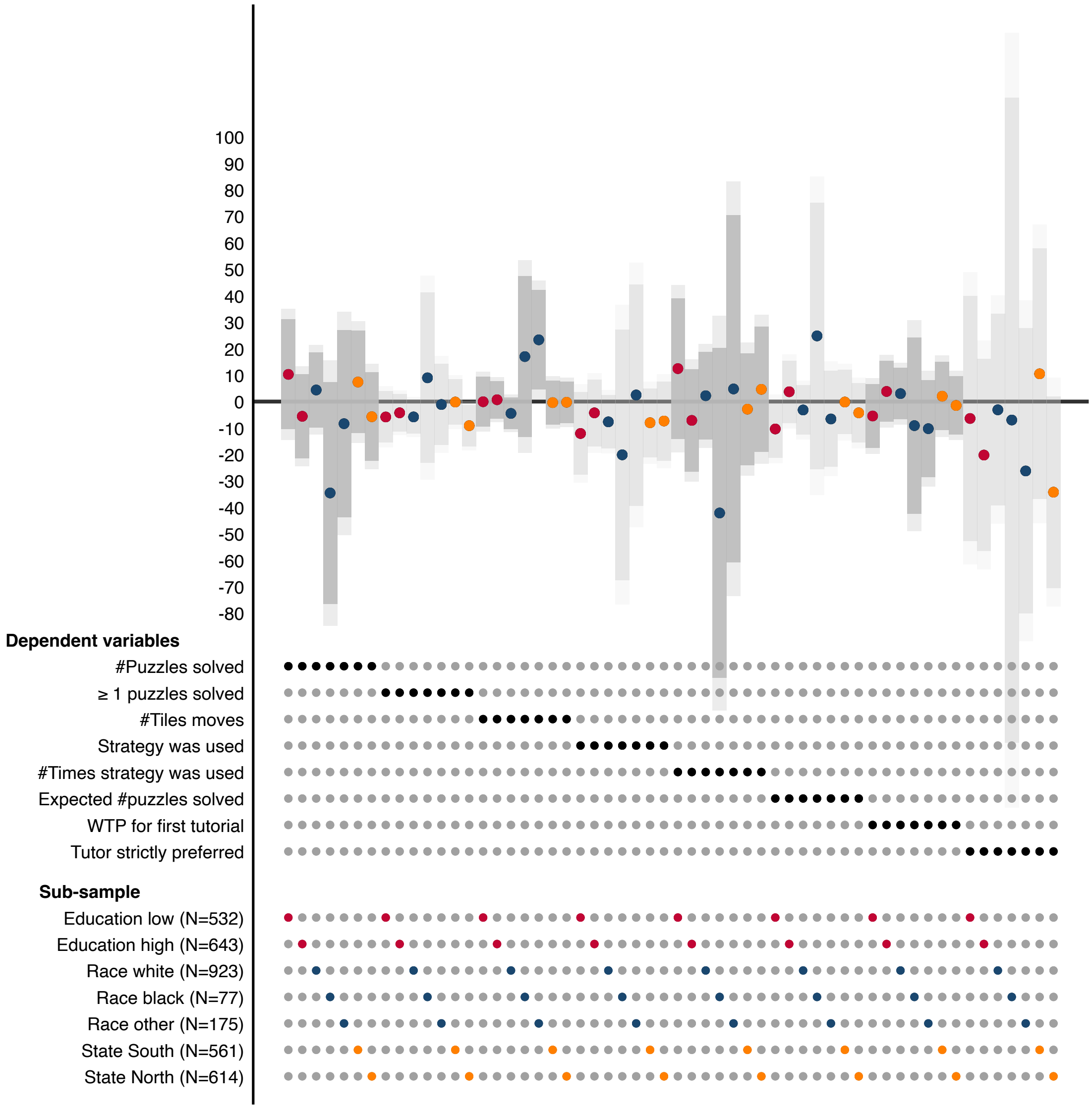
Notes: Effects of tutor being Black on learner's behavior. Specifications as in Figures 4 and 5, by strata-specific sub-samples. Dependent variables described just below the estimates: (1) number of puzzles solved in first stage, (2) dummy for learners who solved at least one puzzle in first stage, (3) average number of moves needed to solve a puzzle in first stage, (4) dummy for learners who solved at least one puzzle in first stage using first-stage strategy, (5) number of puzzles solved in first stage using first-stage strategy, (6) expected number of puzzles solved in first stage, (7) WTP for full tutorial in first stage, and (8) dummy indicating strict preference for a tutor in second stage. Markers show the estimated effects, the darker or lighter whiskers denote 90 and 95 percent confidence intervals based on robust standard errors (learner-level clusters in (3)), respectively. Overall sample: Conservatives under standard piece rate. Strata-specific sub-samples described in the bottom part (different colors mark groups of sub-samples for ease of readability).

Figure A6: Results for Liberals Under Standard Piece Rate: Heterogeneity by Strata Variables



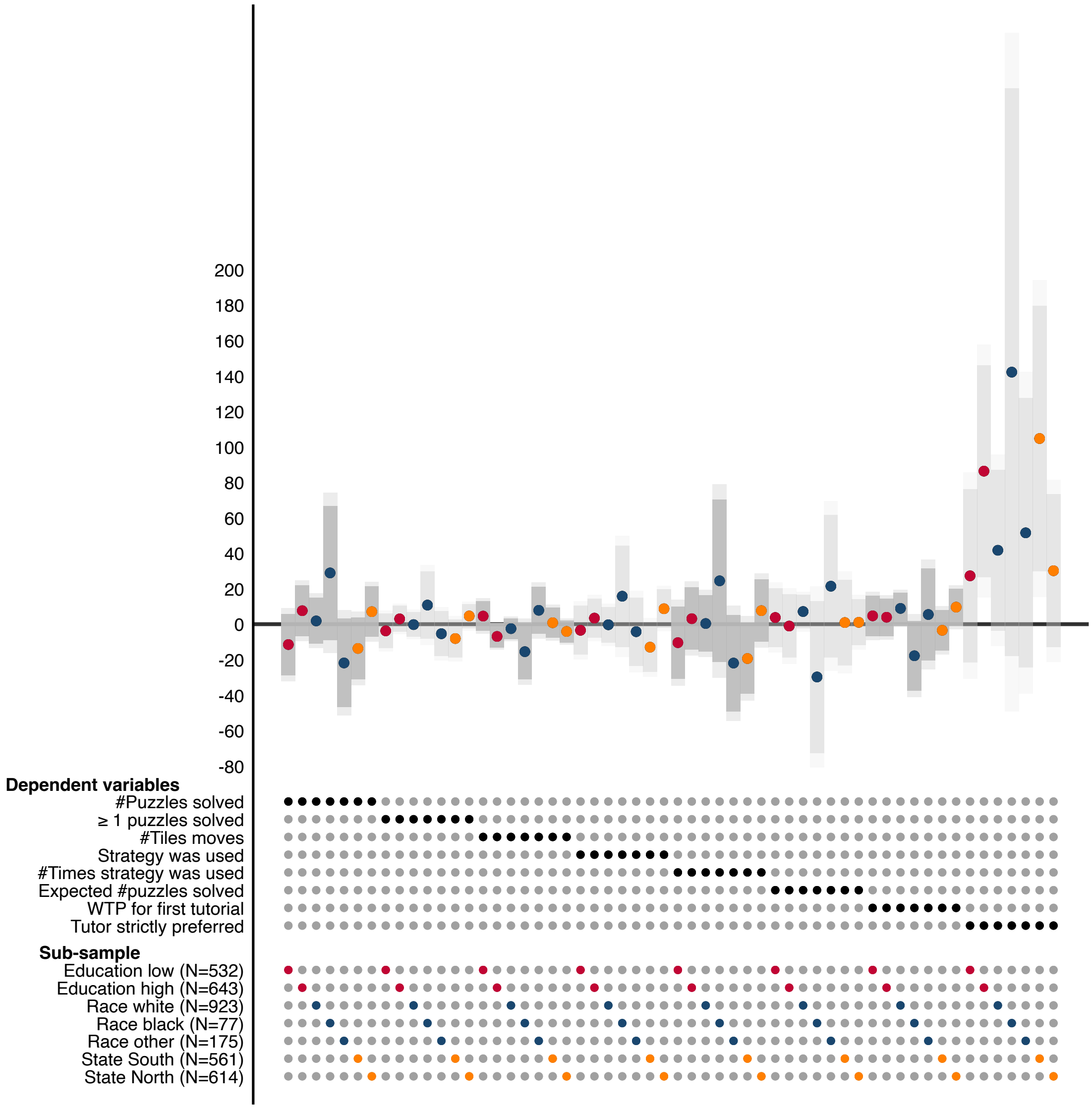
Notes: Effects of tutor being Black on learner's behavior. Specifications as in Figures 4 and 5, by strata-specific sub-samples. Dependent variables described just below the estimates: (1) number of puzzles solved in first stage, (2) dummy for learners who solved at least one puzzle in first stage, (3) average number of moves needed to solve a puzzle in first stage, (4) dummy for learners who solved at least one puzzle in first stage using first-stage strategy, (5) number of puzzles solved in first stage using first-stage strategy, (6) expected number of puzzles solved in first stage, (7) WTP for full tutorial in first stage, and (8) dummy indicating strict preference for a tutor in second stage. Markers show the estimated effects, the darker or lighter whiskers denote 90 and 95 percent confidence intervals based on robust standard errors (learner-level clusters in (3)), respectively. Overall sample: Liberals under standard piece rate. Strata-specific sub-samples described in the bottom part (different colors mark groups of sub-samples for ease of readability).

Figure A7: Results for Conservatives Under High Piece Rate: Heterogeneity by Strata Variables



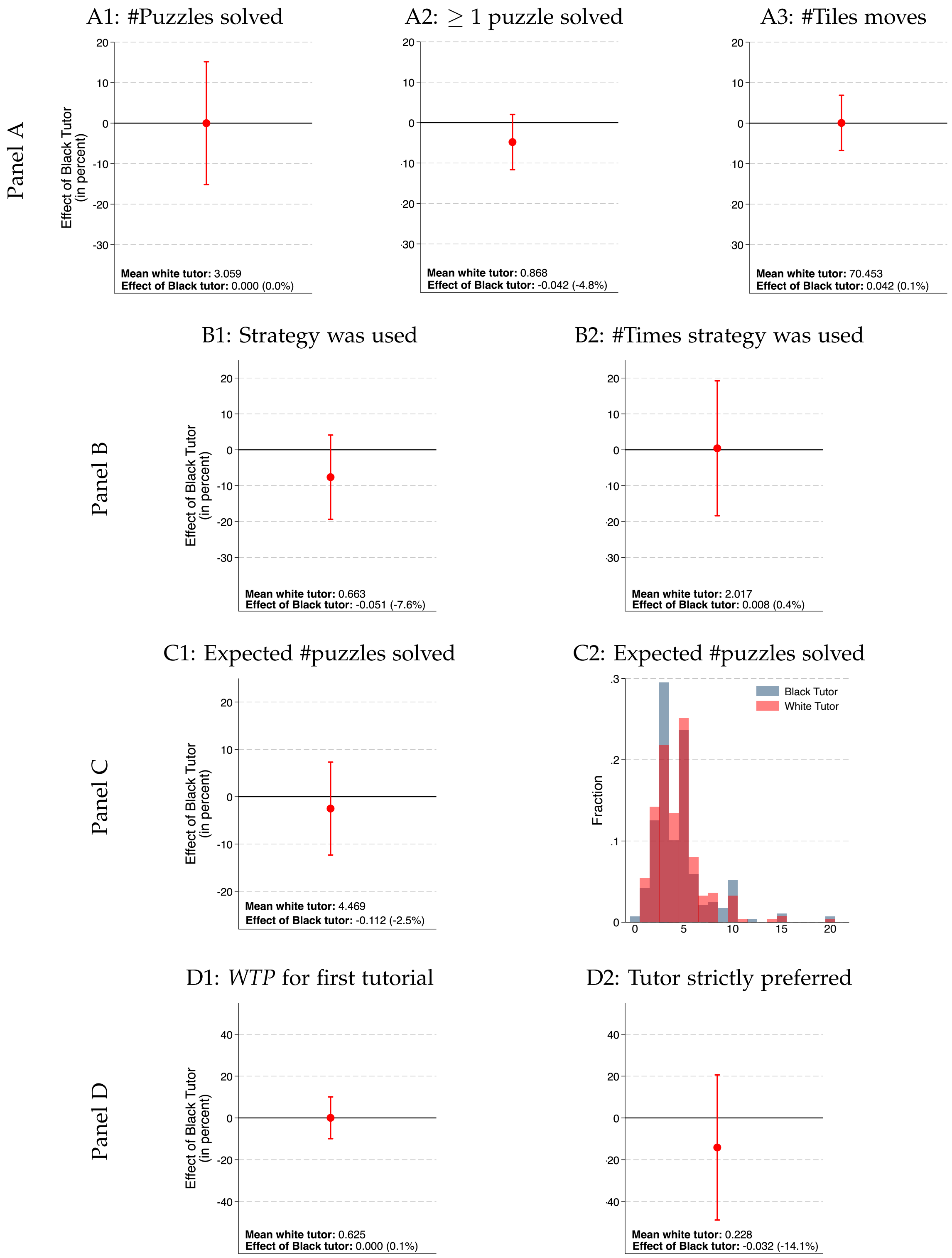
Notes: Effects of tutor being Black on learner's behavior. Specifications as in Figures 4 and 5, by strata-specific sub-samples. Dependent variables described just below the estimates: (1) number of puzzles solved in first stage, (2) dummy for learners who solved at least one puzzle in first stage, (3) average number of moves needed to solve a puzzle in first stage, (4) dummy for learners who solved at least one puzzle in first stage using first-stage strategy, (5) number of puzzles solved in first stage using first-stage strategy, (6) expected number of puzzles solved in first stage, (7) WTP for full tutorial in first stage, and (8) dummy indicating strict preference for a tutor in second stage. Markers show the estimated effects, the darker or lighter whiskers denote 90 and 95 percent confidence intervals based on robust standard errors (learner-level clusters in (3)), respectively. Overall sample: Conservatives under high piece rate. Strata-specific sub-samples described in the bottom part (different colors mark groups of sub-samples for ease of readability).

Figure A8: Results for Liberals Under High Piece Rate: Heterogeneity by Strata Variables



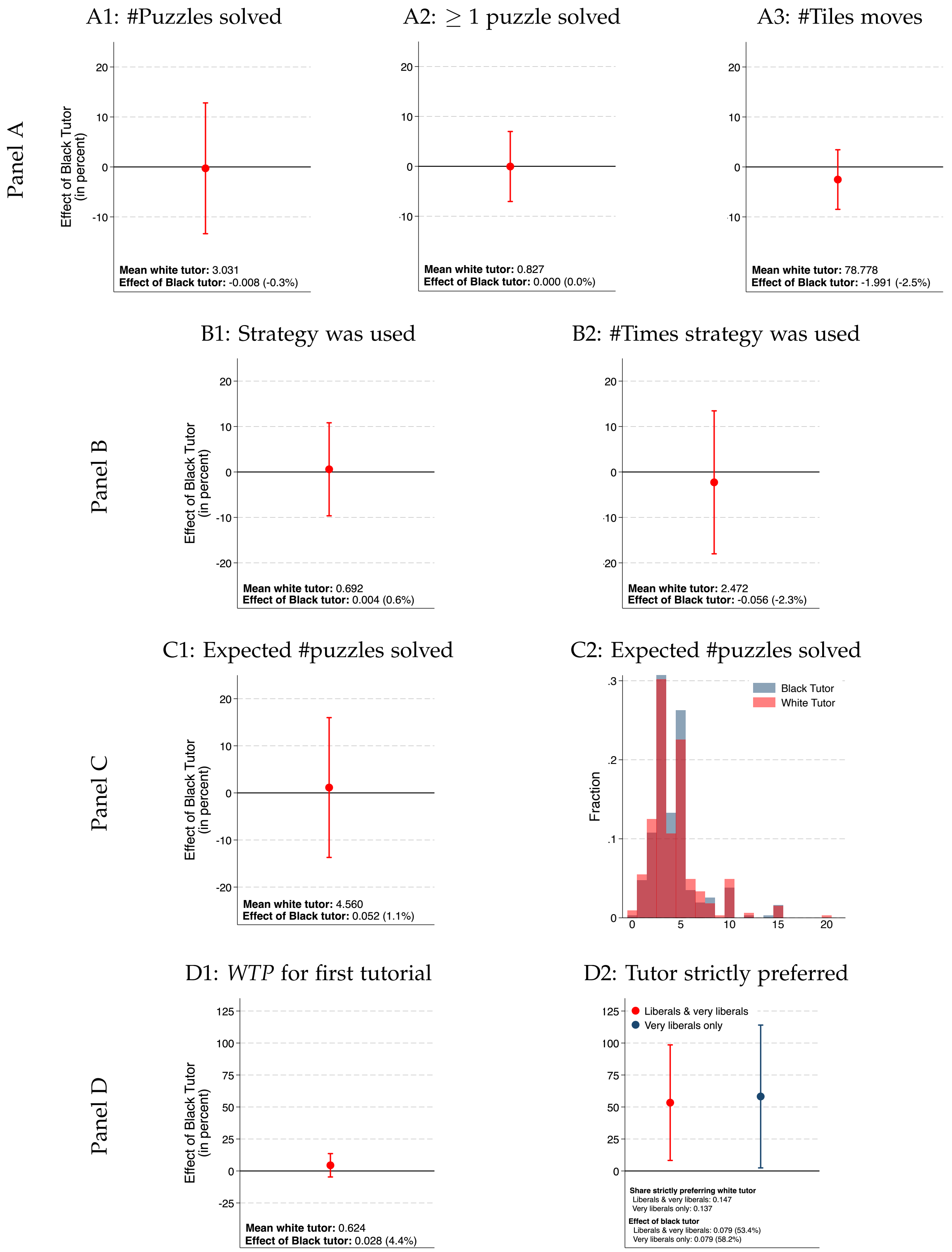
Notes: Effects of tutor being Black on learner's behavior. Specifications as in Figures 4 and 5, by strata-specific sub-samples. Dependent variables described just below the estimates: (1) number of puzzles solved in first stage, (2) dummy for learners who solved at least one puzzle in first stage, (3) average number of moves needed to solve a puzzle in first stage, (4) dummy for learners who solved at least one puzzle in first stage using first-stage strategy, (5) number of puzzles solved in first stage using first-stage strategy, (6) expected number of puzzles solved in first stage, (7) WTP for full tutorial in first stage, and (8) dummy indicating strict preference for a tutor in second stage. Markers show the estimated effects, the darker or lighter whiskers denote 90 and 95 percent confidence intervals based on robust standard errors (learner-level clusters in (3)), respectively. Overall sample: Liberals under high piece rate. Strata-specific sub-samples described in the bottom part (different colors mark groups of sub-samples for ease of readability).

Figure A9: Results for Conservatives Under High Piece Rate



Notes: Effects of tutor race on learners' behavior. Sample: Conservatives under high piece rate. Dependent variables: Number of puzzles solved in first stage (A9A1), dummy for learners who solved at least one puzzle in first stage (A9A2), average number of moves needed to solve a puzzle in first stage (A9A3), dummy for learners who solved at least one puzzle in first stage using first-stage strategy (A9B1), number of puzzles solved in first stage using first-stage strategy (A9B2), expected number of puzzles solved in first stage (A9C1), WTP for full tutorial in first stage (A9D1), dummy indicating a strict preference for a tutor in second stage (A9D2). Confidence intervals based on robust standard errors (learner-level clusters in A9A3 and A9D2). See also Appendix Table A6.

Figure A10: Results for Liberals Under High Piece Rate



Notes: Effects of tutor race on learners' behavior. Sample: Liberals under high piece rate. Dependent variables: Number of puzzles solved in first stage (A10A1), dummy for learners who solved at least one puzzle in first stage (A10A2), average number of moves needed to solve a puzzle in first stage (A10A3), dummy for learners who solved at least one puzzle in first stage using first-stage strategy (A10B1), number of puzzles solved in first stage using first-stage strategy (A10B2), expected number of puzzles solved in first stage (A10C1), WTP for full tutorial in first stage (A10D1), dummy indicating a strict preference for a tutor in second stage (A10D2). Confidence intervals based on robust standard errors (learner-level clusters in A10A3 and A10D2). See also Appendix Table A6.

B Appendix Tables

Table A1: Treatment Allocation

First Stage																
Black / White	B								W							
Actor	1				2				1				2			
Voice	1		2		1		2		1		2		1		2	
Second Stage																
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: Full set of 16 possible combinations of skin colors, actors, and voices used across stages 1 and 2. *B* stands for Black and *W* for white. The numbers represent different actors and voices.

Table A2: Summary Statistics

	Full	Conservative	Liberal	Difference (2)-(3)
	(1)	(2)	(3)	(4)
Participant white (d)	0.735	0.786	0.686	0.099***
	(0.442)	(0.411)	(0.464)	(0.000)
Participant black (d)	0.103	0.066	0.140	-0.074***
	(0.305)	(0.248)	(0.347)	(0.000)
Participant other ethnicity (d)	0.162	0.149	0.174	-0.025
	(0.368)	(0.356)	(0.379)	(0.097)
Participant from southern state (d)	0.432	0.477	0.389	0.088***
	(0.495)	(0.500)	(0.488)	(0.000)
Participant education low (d)	0.442	0.453	0.432	0.021
	(0.497)	(0.498)	(0.496)	(0.309)
Participant age	39.470	41.204	37.816	3.389***
	(12.671)	(12.843)	(12.284)	(0.000)
Participant female (d)	0.513	0.471	0.554	-0.083***
	(0.500)	(0.499)	(0.497)	(0.000)
Observations	2406	1175	1231	2406

Notes: Columns (1) to (3) show means and standard deviations. Column (4) depicts mean differences between conservatives and liberals (p-values from t-tests in parentheses). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Randomization Balance

	=1 Black		=1 other race		=1 southern state		=1 education low	
	Cons.	Lib.	Cons.	Lib.	Cons.	Lib.	Cons.	Lib.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black-White, No Info, High piece rate	0.015	0.032	0.069	0.034	-0.036	0.031	-0.075	0.139*
	(0.044)	(0.056)	(0.063)	(0.061)	(0.088)	(0.079)	(0.088)	(0.080)
Black-White, Info, Standard piece rate	-0.007	0.061	0.029	-0.040	-0.065	0.084	0.032	0.071
	(0.040)	(0.059)	(0.058)	(0.065)	(0.081)	(0.083)	(0.081)	(0.085)
Black-White, No Info, Standard piece rate	0.034	0.102*	0.073	0.045	-0.060	0.005	0.050	-0.014
	(0.041)	(0.057)	(0.058)	(0.062)	(0.082)	(0.080)	(0.082)	(0.081)
White-White, Info, High piece rate	-0.018	0.094	0.071	0.037	-0.092	0.053	0.108	-0.017
	(0.040)	(0.057)	(0.058)	(0.063)	(0.082)	(0.081)	(0.081)	(0.082)
White-White, No Info, High piece rate	0.046	0.023	0.109*	-0.010	-0.092	0.065	-0.033	0.056
	(0.040)	(0.056)	(0.058)	(0.062)	(0.082)	(0.080)	(0.081)	(0.081)
White-White, Info, Standard piece rate	0.013	-0.005	0.040	0.127**	-0.117	0.075	-0.033	0.005
	(0.041)	(0.058)	(0.059)	(0.063)	(0.083)	(0.082)	(0.083)	(0.083)
White-White, No Info, Standard piece rate	0.012	0.043	0.068	0.057	-0.051	0.000	0.093	0.043
	(0.040)	(0.059)	(0.057)	(0.064)	(0.080)	(0.083)	(0.080)	(0.084)
White-Black, Info, High piece rate	-0.028	0.057	0.125**	0.024	-0.056	0.024	0.042	0.062
	(0.041)	(0.056)	(0.059)	(0.061)	(0.083)	(0.079)	(0.083)	(0.080)
White-Black, No Info, High piece rate	-0.036	0.065	0.067	0.035	0.056	-0.015	0.053	-0.044
	(0.043)	(0.058)	(0.062)	(0.064)	(0.088)	(0.082)	(0.087)	(0.083)
White-Black, Info, Standard piece rate	0.008	0.039	0.124**	0.052	-0.047	0.018	-0.041	-0.039
	(0.041)	(0.058)	(0.058)	(0.064)	(0.082)	(0.082)	(0.082)	(0.083)
White-Black, No Info, Standard piece rate	-0.019	0.067	0.018	0.024	-0.021	0.060	0.050	-0.039
	(0.040)	(0.058)	(0.058)	(0.064)	(0.082)	(0.082)	(0.081)	(0.083)
Black-Black, Info, High piece rate	-0.055	0.042	0.081	0.008	-0.095	0.027	0.047	0.004
	(0.042)	(0.056)	(0.061)	(0.061)	(0.085)	(0.079)	(0.085)	(0.080)
Black-Black, No Info, High piece rate	-0.043	0.094*	0.047	0.014	-0.047	0.036	-0.002	0.001
	(0.041)	(0.056)	(0.058)	(0.061)	(0.082)	(0.078)	(0.082)	(0.079)
Black-Black, Info, Standard piece rate	0.001	0.064	0.043	0.036	0.036	0.001	0.062	0.033
	(0.041)	(0.059)	(0.060)	(0.065)	(0.084)	(0.084)	(0.083)	(0.085)
Black-Black, No Info, Standard piece rate	0.003	0.077	0.091	0.057	-0.093	0.043	-0.025	0.011
	(0.042)	(0.057)	(0.060)	(0.062)	(0.084)	(0.080)	(0.084)	(0.081)
Observations	1175	1231	1175	1231	1175	1231	1175	1231
p-value (F-test)	0.577	0.853	0.730	0.807	0.860	0.998	0.696	0.731

Notes: OLS regressions. Sample: Conservatives in Columns (1), (3), (5), and (7) and liberals in Columns (2), (4), (6), and (8). Dependent variables correspond to strata variables (indicators for race, state, and education level). Robust standard errors in parentheses. At the bottom of each column, we report the p-value of an F-test of joint significance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Treatment-Independent Attrition and Sample Restrictions

	Attention check fail (*N=11*) (1)	+ Not completing (*N=1203*) (2)	+ Missing data (*N=20*) (3)	+ Know puzzle (*N=297*) (4)	+ Endogenous tutor (*N=243*) (5)
Black-White, No Info, High piece rate	-0.004	0.010	0.007	-0.003	-0.021
	(0.004)	(0.039)	(0.040)	(0.043)	(0.044)
Black-White, Info, Standard piece rate	0.004	0.086**	0.082**	0.033	0.000
	(0.007)	(0.040)	(0.040)	(0.043)	(0.043)
Black-White, No Info, Standard piece rate	-0.004	0.042	0.034	0.011	-0.033
	(0.004)	(0.039)	(0.040)	(0.042)	(0.043)
White-White, Info, High piece rate	-0.000	0.021	0.017	-0.010	-0.025
	(0.005)	(0.039)	(0.039)	(0.042)	(0.043)
White-White, No Info, High piece rate	-0.000	0.038	0.026	-0.013	-0.064
	(0.005)	(0.039)	(0.040)	(0.042)	(0.043)
White-White, Info, Standard piece rate	-0.004	0.032	0.032	-0.006	-0.028
	(0.004)	(0.040)	(0.040)	(0.043)	(0.044)
White-White, No Info, Standard piece rate	-0.004	0.026	0.014	-0.022	-0.043
	(0.004)	(0.039)	(0.039)	(0.042)	(0.043)
White-Black, Info, High piece rate	-0.004	-0.011	-0.015	-0.031	-0.047
	(0.004)	(0.038)	(0.039)	(0.042)	(0.044)
White-Black, No Info, High piece rate	-0.000	0.059	0.051	0.053	0.041
	(0.005)	(0.040)	(0.040)	(0.043)	(0.044)
White-Black, Info, Standard piece rate	-0.000	0.002	-0.006	-0.018	-0.023
	(0.005)	(0.039)	(0.039)	(0.042)	(0.044)
White-Black, No Info, Standard piece rate	-0.004	0.034	0.026	-0.013	-0.026
	(0.004)	(0.039)	(0.040)	(0.042)	(0.044)
Black-Black, Info, High piece rate	-0.000	0.015	0.003	-0.036	-0.034
	(0.005)	(0.039)	(0.039)	(0.042)	(0.044)
Black-Black, No Info, High piece rate	-0.004	0.042	0.034	-0.016	-0.053
	(0.004)	(0.039)	(0.039)	(0.042)	(0.043)
Black-Black, Info, Standard piece rate	0.004	0.057	0.049	0.026	0.011
	(0.007)	(0.040)	(0.040)	(0.043)	(0.044)
Black-Black, No Info, Standard piece rate	-0.000	0.020	0.008	-0.008	-0.021
	(0.005)	(0.039)	(0.039)	(0.042)	(0.044)
Observations	4180	4180	4180	4180	4180
p-value (F-test)	0.275	0.711	0.741	0.854	0.761

Notes: OLS regressions. Sample: All individuals ever landing on the website ($N = 4{,}396$) minus individuals who did not even reach the initial instructions ($N = 216$) and hence were not assigned a treatment. Dependent variables: Indicator for whether individual failed attention check in Column (1), indicator adding condition for not completing the study in Column (2), indicator adding condition for having incomplete data due to interrupted connectivity to server in Column (3), indicator adding condition for participant reporting to know the slider puzzle task in Column (4), indicator adding condition capturing either (a) random draw of a price of $p > 0$ in the first stage or (b) advisee does not randomly encounter second-stage tutor in the second stage in Column (5). Robust standard errors in parentheses. At the bottom of each column, we report the p-value of an F-test of joint significance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Similarity Within Conservative and Liberal Groups by Intensity of Stated Political Views

Sample	**Conservatives**			**Liberals**		
	Very	Not very	Difference	Very	Not very	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Participant white (d)	0.789	0.775	0.014	0.693	0.667	-0.027
	(0.408)	(0.418)	(0.605)	(0.461)	(0.472)	(0.381)
Participant black (d)	0.063	0.073	-0.010	0.134	0.156	0.021
	(0.243)	(0.260)	(0.543)	(0.341)	(0.363)	(0.348)
Participant from southern state (d)	0.470	0.497	-0.027	0.386	0.397	0.010
	(0.499)	(0.501)	(0.420)	(0.487)	(0.490)	(0.745)
Participant education low (d)	0.460	0.434	0.026	0.431	0.435	0.004
	(0.499)	(0.496)	(0.422)	(0.496)	(0.497)	(0.909)
Participant female (d)	0.462	0.494	-0.032	0.561	0.533	-0.028
	(0.499)	(0.501)	(0.338)	(0.497)	(0.500)	(0.392)
Participant age	40.736	42.478	-1.742*	37.386	39.063	1.677*
	(12.846)	(12.770)	(0.039)	(12.263)	(12.280)	(0.037)
Follow-up: Republican voter (d)[a]	0.784	0.714	0.070*	0.010	0.014	0.004
	(0.412)	(0.453)	(0.019)	(0.099)	(0.119)	(0.540)
Follow-up: Voted Donald Trump (d)[a]	0.755	0.671	0.084**	0.012	0.036	0.024*
	(0.430)	(0.471)	(0.007)	(0.111)	(0.187)	(0.012)
Observations	859	316	1175	916	315	1231

Notes: Columns (1), (2), (4), and (5) show means and standard deviations. Columns (3) and (6) report mean differences between subjects classified by CloudResearch as "very conservative" ("very liberal") and those classified as "conservative" ("liberal"), with p-values in parentheses (t-tests). [a]We only have data for indicators "Republican voter" and "Voted Donald Trump" for participants in the follow-up study (see Appendix F; Very conservative: $N = 699$, Conservative: $N = 283$, Very liberal: $N = 806$, Liberal: $N = 278$). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Main Results Regressions

	First-stage performance			Utilization of first-stage advice		First-stage beliefs	Preferences for tutors	
	#Puzzles solved	≥ 1 puzzle solved	#Tiles moves	$= 1$ if strategy was used	#Times strategy was used	Expected #puzzles solved	WTP for first tutorial	Tutor strictly preferred
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Conservatives								
Black advisor (β_1)	-0.549***	-0.086***	6.128**	-0.084**	-0.394**	0.065	-0.032	0.029
	(0.202)	(0.032)	(2.812)	(0.039)	(0.171)	(0.387)	(0.029)	(0.039)
Black advisor $\times$ High piece rate (β_2)	0.549*	0.044	-6.086	0.033	0.402	-0.177	0.032	-0.061
	(0.312)	(0.044)	(3.731)	(0.056)	(0.258)	(0.446)	(0.043)	(0.056)
High piece rate	0.143	0.030	-2.543	0.044	0.127	-1.466***	0.052*	0.010
	(0.226)	(0.029)	(2.404)	(0.038)	(0.180)	(0.304)	(0.030)	(0.035)
Observations	1175	1175	3355	1175	1175	1175	1175	1152
Mean dep. var.: White advisor	2.924	0.842	71.964	0.705	2.370	5.160	0.565	0.222
$\gamma := \beta_1 + \beta_2$	0.000	-0.042	0.042	-0.051	0.008	-0.112	0.000	-0.032
$\gamma = 0$ (p-value)	1.000	0.167	0.986	0.202	0.965	0.616	0.992	0.424
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Liberals								
Black advisor (β_1)	0.292	0.027	0.465	0.007	0.162	-0.420	0.018	0.059*
	(0.191)	(0.028)	(2.805)	(0.037)	(0.189)	(0.290)	(0.029)	(0.035)
Black advisor $\times$ High piece rate (β_2)	-0.300	-0.028	-2.456	-0.003	-0.218	0.472	0.010	0.019
	(0.279)	(0.041)	(3.666)	(0.052)	(0.275)	(0.441)	(0.041)	(0.049)
High piece rate	0.253	-0.028	-2.253	-0.025	0.242	-1.266***	0.119***	-0.007
	(0.199)	(0.030)	(2.635)	(0.037)	(0.193)	(0.316)	(0.029)	(0.029)
Observations	1231	1231	3672	1231	1231	1231	1231	1208
Mean dep. var.: White advisor	2.924	0.842	80.696	0.705	2.370	5.160	0.565	0.152
$\gamma := \beta_1 + \beta_2$	-0.008	-0.000	-1.991	0.004	-0.056	0.052	0.028	0.079
$\gamma = 0$ (p-value)	0.967	0.991	0.406	0.910	0.776	0.880	0.340	0.020
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Sample: all participants. Dependent variables: Number of puzzles solved in first stage in Column (1), dummy for learners who solved at least one puzzle in first stage in Column (2), average number of moves needed to solve a puzzle in first stage in Column (3), dummy for learners who solved at least one puzzle in first stage using first-stage strategy in Column (4), number of puzzles solved in first stage using first-stage strategy in Column (5), expected number of puzzles solved in first stage in Column (6), WTP for full tutorial in first stage in Column (7), and dummy indicating strict preference for a tutor in second stage in Column (8). Column (3) uses puzzle level data and conditions on subjects solving at least one puzzle in the first stage. Column (8) restricted to participants assigned to tutors of different races across stages. All other columns use subject-level data. Puzzle fixed effects in Column (3). Column (8) controls for instructor's hand model, voice, and stage. All other regressions include strata controls. Robust standard errors in parentheses (learner-level clusters in Columns (3) and (8)). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Balance on Observables: Full Sample vs. Sub-Sample with Tutors of Different Skin Color Across Stages

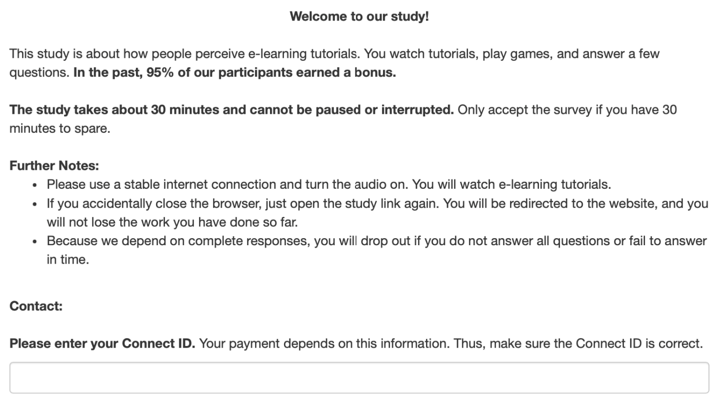
Sample	**Conservatives**			**Liberals**		
	Full	Sub-sample	Difference	Full	Sub-sample	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Participant white (d)	0.79	0.79	-0.01	0.69	0.70	-0.02
	(0.41)	(0.41)	(0.83)	(0.46)	(0.46)	(0.51)
Participant black (d)	0.07	0.07	-0.00	0.14	0.14	0.00
	(0.25)	(0.25)	(0.95)	(0.35)	(0.35)	(0.95)
Participant from southern state (d)	0.48	0.50	-0.04	0.39	0.38	0.01
	(0.50)	(0.50)	(0.20)	(0.49)	(0.49)	(0.64)
Participant education low (d)	0.45	0.45	0.01	0.43	0.43	-0.00
	(0.50)	(0.50)	(0.66)	(0.50)	(0.50)	(0.91)
Participant female (d)	0.47	0.47	-0.00	0.55	0.55	0.01
	(0.50)	(0.50)	(0.92)	(0.50)	(0.50)	(0.76)
Participant age	41.20	41.50	-0.58	37.82	38.48	-1.30
	(12.84)	(13.08)	(0.44)	(12.28)	(12.26)	(0.06)
Observations	1175	576	1175	1231	604	1231

Notes: Columns (1), (2), (4), and (5) show means and standard deviations. Columns (3) and (6) report mean differences between participants who in both stages faced tutors of the same skin color and participants who faced tutors of different skin color across stages, with p-values in parentheses (t-tests). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Experimental Instructions

This section documents the experimental instructions. See Figure 1 in the paper for a simplified timeline.

Screen 1: Login



Welcome to our study!

This study is about how people perceive e-learning tutorials. You watch tutorials, play games, and answer a few questions. **In the past, 95% of our participants earned a bonus.**

The study takes about 30 minutes and cannot be paused or interrupted. Only accept the survey if you have 30 minutes to spare.

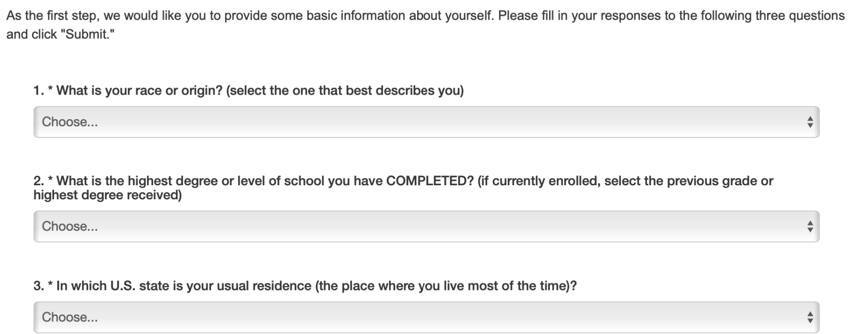
Further Notes:

- Please use a stable internet connection and turn the audio on. You will watch e-learning tutorials.
- If you accidentally close the browser, just open the study link again. You will be redirected to the website, and you will not lose the work you have done so far.
- Because we depend on complete responses, you will drop out if you do not answer all questions or fail to answer in time.

Contact:

Please enter your Connect ID. Your payment depends on this information. Thus, make sure the Connect ID is correct.

Screen 2: Demographics survey[1]



As the first step, we would like you to provide some basic information about yourself. Please fill in your responses to the following three questions and click "Submit."

1. * What is your race or origin? (select the one that best describes you)

Choose...

2. * What is the highest degree or level of school you have COMPLETED? (if currently enrolled, select the previous grade or highest degree received)

Choose...

3. * In which U.S. state is your usual residence (the place where you live most of the time)?

Choose...

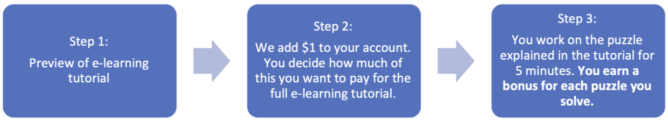
[1] Race: White / Black or African American / Hispanic, Latino, or Spanish / American Indian, or Alaska Native / Asian Indian / Chinese / Filipino / Japanese / Korean / Vietnamese / Other Asian / Native Hawaiian / Guamanian or Chamorro / Samoan / Other Pacific Islander / Some other race. Highest degree: 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) / Bachelor's degree (for example: BA, BS) / Master's degree (for example: MA, MS, MEng, MEd, MSW, MBA) / Professional degree beyond Bachelor's degree (for example: MD, DDS, DVM, LLB, JD) / Doctoral degree (for example: PhD, EdD).

Screen 3: General instructions[2]

Today, you will watch e-learning tutorials that explain how to solve a puzzle. You then work on some puzzles and apply what you have learned. **The more puzzles you solve, the higher your bonus.**

Timers: Note the two timers in the top right corner. You will be automatically redirected from one page to the next.

What's to come:



Particularly, the bonus for each puzzle solved is $1. Buying the full tutorial can help you to increase your payoff (as it explains how to solve the puzzle).

Your payoff will be as follows:

Payoff
= fixed reward (5$)
+ amount you can use to pay for the tutorial ($1)
− payment for the tutorial
+ bonus of $1 for each puzzle solved

Note: Once completed, steps 1 to 3 are repeated for a second round. At the end, a virtual coin flip will decide which of the two rounds counts towards the final payoff. Try your best in each round because it could be the round you will be rewarded for.

Enter the following number into the text box to show that you pay attention: 65

I agree with the above conditions and wish to proceed further.

Screen 4: Preview of video tutorial Stage 1

Preview started automatically. Instructor's hand [Black / white] visible at the beginning and towards the end of the preview. See Figure 3 for a visual impression of the preview.

Screen 5: Willingness to pay Stage 1[3]

Access to the tutorial: We will now determine whether you will access the full e-learning tutorial as follows.

Step	
1	We have **added the amount of $1** to your payoff account. You can use all or a part of this amount to access the tutorial.
2	**Use the slider** below to indicate the highest price you are willing to pay to watch the tutorial.
3	The computer will **randomly draw a price** for the tutorial. The price is a number between $0 and $1.

Rule: 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 **an 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.

Remember: 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 $1** for every puzzle you solve.
- The money you do not spend on the tutorial is added to your payoff.
- The money you spend on the tutorial is not distributed to the instructor.

Slider bar

$0.00 $1.00

[2] The piece rate ("bonus") was either $1 or $0.2.
[3] The piece rate ("bonus") was either $1 or $0.2.

Screen 6: Beliefs Stage 1

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.

1. * How many puzzles do you expect to solve in 5 minutes after having watched the FULL TUTORIAL? [Use numbers only]

2. * How many puzzles do you expect to solve in 5 minutes after having watched the ENTERTAINMENT VIDEO? [Use numbers only]

Screen 7: Video tutorial Stage 1[4]

Respective video started automatically. Instructor's hand [Black / white] visible again at the beginning of the video.

Given your willingness to pay, the random draw of the price determined that you will watch the full tutorial.

Screen 8: Sliding tile puzzle Stage 1

Real effort task: Participants had 5 minutes to solve as many puzzles as possible.

Screen 9: End Stage 1, beginning Stage 2

You just solved 0 puzzles.

Now you will learn a faster way to solve the sliding puzzle, and you will solve another set of puzzles.

Screen 10: Preview of video tutorial Stage 2

Preview started automatically. Instructor's hand [Black / white] visible at the beginning and towards the end of the preview video. See Figure 3 for a visual impression of the preview.

Screen 11: Instructor choice and information treatment Stage 2[5]

Your choice: Instead of choosing between the entertainment video and the tutorial, you will now select one of two potential instructors. You, hence, watch the second tutorial for sure. The tutorial will explain **a way to solve the puzzle that is faster than the one presented before**.

The instructors: 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**).

Important note: Regardless of which instructor is selected, the length of the tutorial will be the same. **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.**

Selection of instructors: 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 following table:

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**. • With a 30% chance, you will get the less preferred one. • 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.

1. * We now start the process of selecting your instructor. Please rank the two instructors by choosing one option:

- I prefer the first instructor to the second instructor.
- I prefer the second instructor to the first instructor.
- I am indifferent between the two instructors.

[4]If the stated WTP was lower than the price drawn, the screen read: "Given your willingness to pay, the random draw of the price determined that you will watch the entertainment video." Participants who watched the entertainment video are not part of our estimation sample.

[5]In the no-info treatment, the part "Furthermore, . . . explain the strategy." was left out.

Screen 12: Willingness to pay Stage 2[6]

Recall: You could choose between the following options:

1. I prefer the first instructor to the second instructor.
2. I prefer the second instructor to the first instructor.
3. I am indifferent between the two instructors,

and you indicated that you prefer option 1.

Now, suppose the following situation would occur: The computer-based random draw determines that the **second instructor** is selected (although you indicated that you prefer the **first instructor**). If this situation indeed occurs, **would you be willing to pay a small fee to get the first instructor for sure instead of the second instructor?**

Willingness to pay: Please state your willingness to pay as follows:

Step	
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 **first instructor** for sure.
2	**Use the slider** below to indicate the highest price you are willing to pay to watch the first instructor.
3	The computer will **randomly draw a price** for watching the first instructor. The price is a number between $0 and $1.

Rule: If your stated willingness to pay **is equal to or above** the price drawn, you will get the **first instructor**. If your stated willingness to pay **is lower** than the price drawn, you will not watch the **first instructor**. Instead, the **second 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.

Keep in mind: The following choice indeed allows you to influence which instructor will be chosen 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 **first instructor**.

Further notes:

- After having watched the tutorial, you will have 5 minutes to solve as many puzzles as possible.
- Again, you will earn **a bonus of $1** for every puzzle you solve.

Slider bar

$0.00 **$0.36** $1.00

Your stated willingness to pay is currently **$0.36**

Screen 13: Beliefs Stage 2

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).

1. * 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]

2. * 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]

Screen 14: Video Tutorial Stage 2

[First instructor:] Text TBA.

[Second instructor:] Text TBA.

Respective video started automatically. Instructor's hand [Black / white] visible again at the beginning of the video.

Screen 15: Sliding tile puzzle Stage 2

Real effort task: Participants had 5 minutes to solve as many puzzles as possible.

[6] The piece rate ("bonus") was either $1 or $0.2.

Screen 16: Final survey

Before concluding, we would like to ask you a few final questions.
[Recall that answering all questions is mandatory. YOU WILL ONLY GET PAID IF YOU ANSWER ALL QUESTIONS. There are 12 questions in total. Question 12 is optional but we would greatly appreciate your input.]

1. * What is your gender?

- Male
- Female

2. * How old are you (only numbers are allowed)?

3. * Did you already know how to solve the sliding puzzle before participating in this study?

- Yes
- No

4. You might have participated in an earlier version of our study on a different platform (e.g., MTurk or Prolific). If you did the study before, we would like to know how you experienced today's study compared to the previous one.

- I am NOT active on any other crowdsourcing platform.
- I am active on other crowdsourcing platforms, but I do not remember participating in an earlier version of this study.
- I remember doing the study on a different platform, and in today's study, I noticed the following differences (a text box will appear)

What do you remember about the first instructor?

5. The first instructor was experienced in giving advice to others.

- Yes
- No
- I did not watch the video / I don't remember

6. The first instructor's gender was:

- Male
- Female
- I did not watch the video / I don't remember

7. The first instructor's race was:

- White
- Black or African American
- I did not watch the video / I don't remember

What do you remember about the second instructor?

8. The second instructor was experienced in giving advice to others.

- Yes
- No
- I did not watch the video / I don't remember

9. The second instructor's gender was:

- Male
- Female
- I did not watch the video / I don't remember

10. The second instructor's race was:

- White
- Black or African American
- I did not watch the video / I don't remember

11. * When working on the puzzle for the second time, did you use the strategy that was presented in the second tutorial?

- Yes, always
- Yes, sometimes
- No, never
- (I did not watch both tutorials)

12. Please, write down any comments you might have regarding the study 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?)

Screen 17: Payoff

We thank you for participating in this study!

In total, you have earned $6. This amount includes (a) the $5 fixed reward, (b) the $1 you received as an additional amount to pay for the tutorial (net of the price for the video), and (c) the bonus. The bonus is based on your activity in period 2.

Last step: Please copy the following completion code (secret key) to the Connect platform:

dBXqHLfkzQKM

Afterwards, you can close the page. Have a good day!

D Evolution of Experimental Design and Pre-Analysis-Plans

This section documents the evolution of the experimental design and the pre-analysis-plan (PAP).

D.1 Original PAP

The original PAP was registered on May 11, 2020, as AEARCTR-0005812 and is available at https://www.socialscienceregistry.org/trials/5812.[7] All core elements of the design were identical to the design of the main experiment described in the current paper, with one key difference: The sampling did not condition on political views, and we did not plan to do the main data analysis separately by political views. The experiment was fielded on MTurk, as originally intended, and the data were collected as planned. The analysis of the data revealed no discrimination on average. This finding left us wondering if the MTurk population is somehow distinct—perhaps skewing more liberal or attracting individuals who are, on average, less likely to discriminate. To address this concern and to test more broadly whether people discriminate on MTurk, we designed a follow-up study to explore these questions further We thank Ricardo Truglia Perez for this suggestion.

D.2 First Update

The first update of the original PAP was registered on May 31, 2021, as AEARCTR-0007737 and is available at https://www.socialscienceregistry.org/trials/7737. As main motivation for the update, the registration stated that we wanted to test "if the sample in the original experiment exhibits discriminatory behavior on a different domain." The updated registration detailed the following changes to the original registration and PAP:

- Invite all participants who completed both rounds of the original experiment for another study, the "help-or-harm task" (Bartos *et al.*, 2021). For details on the task, see Appendix F.
- Construct an indicator for individuals acting as discriminators against African-Americans (strictly lower reward to the African-American recipient than to the white recipient).
- Use this binary classification to conduct a further heterogeneity analysis in the main experiment.

The analysis of the resulting data showed that (compared to a national representative sample) the original MTurk sample was, indeed, selected. Particularly, on MTurk, there were fewer individuals whom we classified as "white-favoring individuals" (individuals proposing higher rewards for a white than for a black recipient in the "help-or-harm task"). Given these findings, we decided to repeat our experiment in a national representative sample.

D.3 Second Update

The second update of the original PAP was registered on November 15, 2021, as AEARCTR-0008563 and is available at https://www.socialscienceregistry.org/trials/8563. As main motivation

[7] Here and in the following, we report initial registration dates.

for the update, the registration stated: "First, we aim at repeating our main experiment with a larger US national representative sample to test whether there is, on average, discrimination in seeking advice in a more representative population. Second, we aim at conducting sub-sample analysis along the individuals' types. Particularly, we separately study the behavior of three groups of individuals in our main experiment: (a) white-favoring individuals: individuals proposing higher rewards for a white than for a black recipient in a simple allocation task (b) black-favoring individuals: individuals proposing lower rewards for a white than for a black recipient (c) egalitarian individuals: individuals proposing the same reward for a white and a black recipient." The updated registration detailed the following changes:

- Increase statistical power such that we can study discrimination in seeking advice in the three subsamples of subjects (6,000 respondents in total). This large sample size was necessary because the design involved a two-stage process: we first needed to classify individuals based on their behavior in a separate allocation task and then recontact them for the main experiment. As a result, we expected to lose a considerable number of observations between classification and re-invitation.
- Use services by data provider Dynata to obtain a national representative sample.
- Invite participants for our classification survey ("help-or-harm task").
- Invite the same participants for our main experiment.
- Run the full analysis described in the PAP of AEARCTR-0005812. In addition, run the same analysis by type according to behavior in the classification task (i.e., separately for white-favoring individuals, egalitarians, and black-favoring individuals).

During the pre-tests for the pre-registered study, Dynata realized that they were unable to run our study. In particular, they were unable to recruit a sufficient number of representative subjects. As a consequence, we had to abort our cooperation with Dynata. We contacted several alternative access panels, but none could guarantee the sample size needed for studies that are as long as ours. Hence, we could not conduct the study pre-registered in AEARCTR-0008563. We then decided to collect the data in collaboration with CloudResearch and substitute the classification of subjects according to the "help-or-harm task" by pre-profiled information on the subjects' political views.

This decision was motivated by two factors. First, using CloudResearch's platform gave us access to both MTurk and Connect samples, which was essential for reaching our required sample size. In addition, data quality is higher on Connect, which implements attention checks and provides verified respondent profiles. Second, the shift to CloudResearch allowed us to replace the help-or-harm classification with a classification by political views (CloudResearch pre-screens participants in this dimension). This enables us to examine discrimination in advice-seeking behavior through a politically relevant and substantively meaningful lens, rather than simply replicating discrimination across different domains.

D.4 Third Update = Current PAP

The third update of the original PAP was registered on July 11, 2022, as AEARCTR-0009715 and is available at https://www.socialscienceregistry.org/trials/9715. The registration detailed the motivation to trace out the heterogeneity in discriminatory behavior in advice-seeking by studying discrimination in the main experiment along the individuals' political views. The updated registration detailed the following changes:

- Use CloudResearch to collect the data. Recruit 1,000 in each of the extreme categories of political views ("very conservative" and "very liberal") and 334 participants in each of the remaining categories ("conservative", "moderate", and "liberal").
- Invite participants for main experiment on discrimination described in AEARCTR-0005812.
- Invite the same participants for classification survey described in AEARCTR-0007737. This allows to analyze if the results on discrimination persist in a different domain.
- Run the full analysis described in the PAP of the main experiment for each type in terms of political views.

E Further Materials

E.1 Video Production

We highly standardized the video-production process. First, when recording the voice-overs and videos, the actors exactly followed written scripts. The instructions also described the actor's hand movements. We even provided example videos to illustrate the timing and hand gestures. Second, we formalized the terms of the video-production process through a written agreement. Consequently, also the video and post-production process was highly standardized (including the use of equipment such as the greenscreen). Third, in our quest for the optimal method and partner for skin-color manipulation, we engaged with various firms known for employing diverse techniques, and evaluated various samples.[8] Our selection process led us to conclude that video post-production yields the most satisfactory results. We then partnered with a video producer who previously has been involved in the production of many blockbuster movies (such as Star Wars, Black Widow, Terminator: Dark Fate, Aladdin, and The Expendables).

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:
- Out of each long video a short version will be cut
  - Short complicated: cut from final clips of long complicated

[8] Some firms, for example, offered to produce animated videos. Back then, however, these techniques did not produce material of sufficiently high quality.

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

  - 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

E.2 Performance Using First-Stage and Second-Stage Strategies

The instructions informed participants that in the second stage they would learn a faster strategy to solve the sliding tile puzzle. We tested if this statement is true. For that purpose, we extended the "A^* pathfinding algorithm".[9] The algorithm counts the minimum number of moves needed to solve a puzzle using a given strategy and allows us to evaluate the strategies' efficiency. The Python file is available upon request.

Appendix Table A8 presents the starting positions used in the study. In each stage, participants could work on 15 puzzles in a fixed and randomly chosen order. If a participant solved more than 15

[9]See `http://theory.stanford.edu/ ∼ amitp/GameProgramming/AStarComparison.html`.

puzzles, the puzzles were repeated in the same order. Table A8 shows starting positions translated into arrays. For example, the position

1	6	4
7	3	
5	8	2

would be translated to [1, 6, 4, 7, 3, 0, 5, 8, 2], where the number 0 indicates the blank position.

Table A8: Starting Positions of Puzzles

Puzzle n	Stage 1	Stage 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 A9 presents the results of the pathfinding algorithm for the puzzles presented in Table A8. First, the simple strategy is always faster than the complicated strategy. Considering the 30 puzzles in Table A8, on average, the algorithm executes 38.6 moves to solve the puzzle with the first-stage strategy and 28.1 moves with the second-stage strategy. Second, for all $n = 1, 2, \ldots, 15$, puzzle n in stage 2 can be solved faster when using the second-stage strategy compared to puzzle n in stage 1 when using the first-stage strategy.

E.3 Patterns in Data and Strategy Use

Complicated strategy We classify a game being solved using the complicated (stage 1) strategy when the following sequence of moves occurs in the data (see how to read the tile position array in Appendix Section E.1: [2, 0, 3, 1, x, x, x, x, x], [0, 2, 3, 1, x, x, x, x, x], and [1, 2, 3, 0, x, x, x, x, x], where x stands for any other number.

Simple strategy We classify a game being solved using the simple (stage 2) strategy when the following sequence of moves occurs in the data: [1, 3, 0, x, 2, x, x, x, x], [1, 0, 3, x, 2, x, x, x, x], and [1, 2, 3, x, 0, x, x, x, x]

Table A9: Performance of First-Stage and Second-Stage Strategies

Puzzle n	Starting Positions: First Stage		Starting Positions: Second Stage	
	First-Stage Strategy	Second-Stage Strategy	First-Stage Strategy	Second-Stage Strategy
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

E.4 Advisor Selection Mechanism

Our mechanism is similar to the one Toussaert (2018) proposed. It serves two primary objectives. First, it motivates participants to reveal their true preferences. Second, it prevents endogenous self-selection to advisors, making sure that advisees randomly encounter tutors of different skin colors. Our mechanism achieves incentive compatibility by guaranteeing that the expressed advisor preference, in principle, increases the probability that the preferred advisor is selected. Simultaneously, it incorporates a random component in the selection mechanism to fulfill the random-assignment objective. The mechanism consists of three elements:

First element: The first element elicits preferences over advisors with a simple survey. Specifically, after the second-stage preview, participants indicate with radio buttons if they

1. prefer the first-stage tutor to the second-stage one,
2. prefer the second-stage tutor to the first-stage one, or
3. are indifferent between the two tutors.

Second element: The second element consists of a first lottery. This lottery not only ensures that participants with a strict preference for one particular advisor reveal this preference but also guarantees the random assignment of advisors. As we mentioned in the main text, the lottery implements a 95% probability that the second-stage advisor will present the e-learning tutorial (case 1). In this case, the advisees' stated preference does not impact who will be selected as an advisor. Thus, by design, the lion's share of all participants randomly faces advisors. With the counter probability of 5%, participants can, in principle, influence the advisor selection with their stated preference (case 2). Thus, case 2 ensures the incentive compatibility. The instructions mention that this probability is positive without disclosing its exact values.

Third element: The third element combines a second lottery with a BDM mechanism. It allows us to double-check if (a) we only classify participants with a strict preference as such and (b) we do not classify indifferent participants as people with a strict preference. The potential existence of indifferent individuals vastly complicates the elicitation procedure. By construction, indifferent participants do not have a preferred tutor. Thus, they could randomly choose one option when asked about their preference (first element). The reason is that they perceive all options as equally good (such that they do not have a clear incentive to reveal their true preference). Our key challenge is, thus, to separate indifferent individuals (who select a "strict preference" radio button) from those with true strict preferences. The fundamental idea of our approach to tackle this problem is that only participants with a strict preference have a positive *WTP* to get their preferred advisor.

We implement a second lottery to check if this is the case. This lottery affects which advisor will be selected if participants end up in case 2. Specifically, it determines that in case 2, the preferred advisor presents the video with a probability of 70% (known to participants). With the counter probability, the non-preferred adviser delivers it. We then elicit the advisor's *WTP* to get the preferred advisor for sure in the hypothetical scenario in which the lottery picks the non-preferred advisor.[10] We use the same BDM mechanisms as in the first stage to elicit this *WTP*. Only participants who strictly prefer one of the two advisors should state a positive *WTP*.[11] Instead, participants who are indifferent between the two advisors should enter a *WTP* of zero. They do not care who will present the tutorial and, thus, should state a *WTP* of zero to get one advisor for sure. Note that, in this stage, we draw the price p from a uniform distribution.

E.5 Ensuring High Data Quality

We take several steps to make sure our sample does not suffer from a typical concern of low quality of data collected online. First, we use CloudResearch, a platform known for its high data quality compared to other providers (Hauser *et al.*, 2023; Stagnaro *et al.*, 2024). Second, the attention check, the fixed timing of each page, and sampling restrictions enable us to screen out inattentive participants. Third, the use of incentives motivates participants to exert effort and pay attention. Fourth, we conducted a follow-up study with the same set of participants (see Subsection 2.5). We collect basic demographic data in both surveys and compare responses to check consistency in responses. The data remains highly consistent across both studies, despite a substantial time gap between both studies and the use of different websites for data collection. Specifically, the correlations range from 0.93 (race) to 0.97 (gender) and 0.99 (age).

[10] If participants are indifferent between both advisors, one of the two advisors will be randomly picked as the hypothetical preferred advisor. The procedure then follows as if a preferred advisor was initially selected.

[11] They get their preferred advisor with a 5% probability by betting the entire dollar.

F Follow-Up Study: Help-Or-Harm Task

F.1 Design, Sample, and Results

Procedures: Several weeks after the main experiment, we invited all participants who completed it to participate in a follow-up experiment. We did not mention the main experiment and used a different data-collection platform (Qualtrics), making it very unlikely that participants recognized the connection between the two studies. Moreover, participants had previously given consent to be invited to future follow-up studies.

Experimental Design: We use the "Help-or-Harm task" originally introduced by Bartos *et al.* (2021) to measure discrimination. Appendix Section F.2 presents the full set of instructions.

In this experiment, using a slider, participants allocate monetary payoffs to recipients with different characteristics. The default payoff is $5. By actively moving the slider, participants can either increase, decrease, or maintain the default reward within the range of $0 and $10 (in increments of $0.01). The task does not feature pecuniary benefits or costs to the decision-maker. Therefore, the allocation decision demonstrates the decision makers' willingness to engage in hostile behavior towards the recipients (active reduction of the payoff below $5) or prosocial behavior (active increase of the payoff above $5). We rule out any strategic behaviour by not allowing unused amounts to be rewarded to other recipients.

The main treatment variation is that we experimentally vary the recipients' characteristics. Specifically, we introduce brief profiles of two recipients sequentially—one recipient is Black and the other is white. Participants view the profiles one at a time and allocate rewards immediately after seeing each profile. We do not inform participants beforehand about how many decisions they will make, and we randomize the order of the profiles.[12] Finally, we paid 30 randomly selected CloudResearch users, who matched the described characteristics, based on the choices made by 30 randomly selected participants.

Sample: In total, 2,057 (or 85%) of the main study participants completed the follow-up study. The treatments in the main experiment had no impact on the participation probability in the follow-up experiment. Moreover, the participation probability is uncorrelated with the observable characteristics of conservative and liberal participants (see Appendix Table F1).[13]

Results: Our main results are as follows. First, among conservatives, 15.8% of subjects allocate a higher amount to white than Black recipients. By contrast, only 3.0% of liberals exhibit such behavior. Second, 34.2% of liberals allocate more to Black recipients, while only 11.2% of conservatives give more to Black participants. Third, 17.4% of conservatives show hostility against Black recipients (by actively

[12] To reduce participants' awareness that we aim at studying discrimination, we describe recipients as male U.S. residents aged between 20 and 40 years. We explicitly state that recipients are not participants themselves, removing strategic incentives such as reciprocity. Additionally, we clarify that the two allocation choices are independent, eliminating any strategic interaction between them.

[13] CloudResearch experienced a technical problem when re-contacting a batch of conservative participants in our main study for the allocation task, resulting in a lower re-contact rate then usual. Appendix Table F1 therefore studies selection into follow-up participation both for the full sample and when excluding the batch of subjects where the sampling error occurred. Because the sampling error happened only when subjects were invited for the follow-up survey, we keep these subjects in the analysis of the main experiment, but exclude them when analyzing choices in the allocation task. Among the remaining subjects, re-contact rates among conservatives and liberals are very similar (88.1 and 87.9 percent, respectively). Despite the reduced statistical power, our main results hold when restricting the sample to follow-up study participants only, with the exception of the positive effect on the incentivized Black advisor ranking in the second stage by liberals (Appendix Table F3).

Table F1: Balance on Observables: Main Sample vs. Follow-up Study Sample

Panel A: Full sample

Sample	Conservatives				Liberals			
	Full	Follow-up	No follow-up	Difference	Full	Follow-up	No follow-up	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Participant white (d)	0.786	0.832	0.560	-0.272***	0.686	0.687	0.685	-0.002
	(0.411)	(0.374)	(0.498)	(0.000)	(0.464)	(0.464)	(0.466)	(0.958)
Participant black (d)	0.066	0.065	0.070	0.005	0.140	0.142	0.121	-0.022
	(0.248)	(0.246)	(0.256)	(0.779)	(0.347)	(0.350)	(0.327)	(0.478)
Participant from southern state (d)	0.477	0.507	0.335	-0.172***	0.389	0.388	0.396	0.008
	(0.500)	(0.500)	(0.473)	(0.000)	(0.488)	(0.488)	(0.491)	(0.855)
Participant education low (d)	0.453	0.474	0.350	-0.124**	0.432	0.431	0.443	0.012
	(0.498)	(0.500)	(0.478)	(0.001)	(0.496)	(0.495)	(0.498)	(0.777)
Participant female (d)	0.471	0.504	0.310	-0.194***	0.554	0.551	0.577	0.026
	(0.499)	(0.500)	(0.464)	(0.000)	(0.497)	(0.498)	(0.496)	(0.544)
Participant age	41.204	42.176	36.465	-5.711***	37.816	38.123	35.584	-2.539*
	(12.843)	(13.123)	(10.154)	(0.000)	(12.284)	(12.335)	(11.707)	(0.018)
Observations	1175	975	200	1175	1231	1082	149	1231

Panel B: Excluding problematic observations

Sample	Conservatives				Liberals			
	Full	Follow-up	No follow-up	Difference	Full	Follow-up	No follow-up	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Participant white (d)	0.825	0.831	0.785	-0.05	0.686	0.687	0.685	-0.00
	(0.380)	(0.375)	(0.413)	(0.192)	(0.464)	(0.464)	(0.466)	(0.958)
Participant black (d)	0.070	0.065	0.108	0.04	0.140	0.142	0.121	-0.02
	(0.256)	(0.247)	(0.311)	(0.077)	(0.347)	(0.350)	(0.327)	(0.478)
Participant from southern state (d)	0.491	0.502	0.408	-0.09*	0.389	0.388	0.396	0.01
	(0.500)	(0.500)	(0.493)	(0.043)	(0.488)	(0.488)	(0.491)	(0.855)
Participant education low (d)	0.485	0.479	0.531	0.05	0.432	0.431	0.443	0.01
	(0.500)	(0.500)	(0.501)	(0.270)	(0.496)	(0.495)	(0.498)	(0.777)
Participant female (d)	0.503	0.506	0.477	-0.03	0.554	0.551	0.577	0.03
	(0.500)	(0.500)	(0.501)	(0.531)	(0.497)	(0.498)	(0.496)	(0.544)
Participant age	41.847	42.316	38.369	-3.95**	37.816	38.123	35.584	-2.54*
	(13.061)	(13.123)	(12.082)	(0.001)	(12.284)	(12.335)	(11.707)	(0.018)
Observations	1094	964	130	1094	1231	1082	149	1231

Notes: Panel A shows the full sample. Panel B shows the sample excluding the batch of subjects affected by a technical problem when CloudResearch invited the subjects (see Footnote 13 for details). Columns (1), (2), (3), (5), (6) and (7) show means and standard deviations. Columns (4) and (8) report mean differences between subjects who participated in the follow-up study and subjects who did not, with p-values in parentheses (t-tests). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

reducing the reward below the default level of $5). Only 7.3% of liberals are hostile toward Blacks. The corresponding numbers are 13.0% and 18.6% for white recipients. In summary, conservatives are more likely than liberals to exhibit explicit discriminatory behavior against Blacks. Conversely, and in line with our results from the main study, we also observe that liberals are more likely than conservatives to show Black-favoring behavior. This finding is novel: Earlier studies compare behavior of conservatives and liberals in relative terms, without distinguishing between in-group favoritism and out-group bias. Table F2 presents the full set of results.

Table F2: Follow-Up Study Allocations and Behavior

Sample	**Conservatives**				**Liberals**			
HHT Recipient	Black (1)	White (2)	Difference (3)	p-value (4)	Black (5)	White (6)	Difference (7)	p-value (8)
Chosen reward	6.883	7.189	-0.306	<0.001	7.970	6.878	1.091	<0.001
Hostile behavior (d)	0.174	0.130	0.044	<0.001	0.073	0.186	-0.113	<0.001
Prosocial behavior (d)	0.647	0.682	-0.035	<0.001	0.773	0.593	0.179	<0.001
White > Black (d)			0.158				0.030	
Black > White (d)			0.112				0.342	
Observations	975				1082			

Notes: Columns (1), (2), (5), and (6) show means. Columns (3) and (7) report simple differences in means. Columns (4) and (8) show p-values of t-tests for differences between behavior towards Black and white recipients. Hostile (or Prosocial) behavior is an indicator for rewards below (or above) \$5. Sample: Follow-up study participants.

F.2 Instructions

Login screen: Welcome and thank you for joining our study!

We are a group of international academic researchers. Your participation in this survey contributes to the success of our research project.

The study takes around **3 minutes** to complete.

Notes:

- If you accidentally close the browser, just open the survey link again (using the same browser and same device). You will, again, be redirected to the website.
- Your participation in this survey is voluntary, and you may withdraw your participation or your data at any time without any penalty. The data will only be used for research purposes and never for identification purposes.

Survey: As the first step, we would like you to provide some basic information about yourself.

- **Please enter your Connect ID** *[Text field]*
- **What is your gender?** *[Male / Female / Other]*
- **How old are you?** *[Text field]*
- **What is your race or origin?** (select the one that best describes you) *[White / Black or African American / 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]*
- **What is the highest degree or level of school you have COMPLETED?** (if currently enrolled, select the previous grade or highest degree received) *[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) / Bachelor's degree (for example: BA, BS) / Master's degree (for example: MA, MS, MEng, MEd, MSW, MBA) / Professional degree beyond Bachelor's degree (for example: MD, DDS, DVM, LLB, JD) / Doctoral degree (for example: PhD, EdD)]

- **In which U.S. state is your usual residence** (the place where you live most of the time)? *[LIST OF US STATES]*
- **In politics today, do you consider yourself a Republican, Democrat or Independent?** *[Republican / Democrat / Independent / Don't know]*
- **Who did you vote for in the 2020 Presidential Election?** *[Donald Trump / Joe Biden / Other or Don't know / Didn't vote]*

Help-or-harm task: Now, you will make several decisions that can have a real impact on someone else's financial reward.

- We will ask you if you want to increase or decrease the reward of several persons.
- Each of them is a different person, but none of them is a participant in this survey.
- At the end of this survey, we will randomly select thirty participants and select one of their decisions that will determine someone else's reward. Therefore, please, make careful decisions, as each of your decisions can play a role.

Please make a decision for each person:

- If you do not change the reward, the person will receive 5 USD.
- You can choose to increase or decrease the reward to any amount between 0 USD and 10 USD.
- Please use the slider to specify a reward for each person.

To ensure that participants follow the instructions, we have included a question about your participation in earlier online surveys at the end of this page. Regardless of what the true answer is, just fill in the number "51" (without the quotation marks). Similar questions may be asked later.

How many online surveys have you ever participated in? *[Text field]*

[Next screen]

On the following pages, we ask you to determine the reward. We will provide a brief description of each person to be rewarded.

[Next screen]

A person who is male, between 20 and 40 years old, living in the US, and [TREATMENT: African-American / white]. *Use the slider to select the reward between 0-10 USD for this person.*

[Slider here]

Conclusions: We thank you for participating in this study. Please enter the secret key in Connect platform to indicate your participation in this study:

[Completion code]

F.3 Main Results for Follow-Up Study Participant Sub-Sample

Table F3: Main Results Regressions: Follow-Up Study Participant Sub-Sample

	First-stage performance			Utilization of first-stage advice		First-stage beliefs	Preferences for tutors	
	#Puzzles solved	$\geq$ 1 puzzle solved	#Tiles moves	= 1 if strategy was used	#Times strategy was used	Expected #puzzles solved	WTP for first tutorial	Tutor strictly preferred
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Conservatives								
Black advisor (β_1)	-0.414**	-0.068*	5.299*	-0.081*	-0.380**	0.087	-0.029	0.021
	(0.201)	(0.035)	(3.117)	(0.043)	(0.193)	(0.417)	(0.031)	(0.042)
Black advisor × High piece rate (β_2)	0.444	0.013	-5.888	0.007	0.269	-0.190	0.052	-0.027
	(0.310)	(0.049)	(4.236)	(0.061)	(0.285)	(0.485)	(0.046)	(0.060)
High piece rate	0.096	0.035	-2.851	0.051	0.106	-1.506***	0.057*	-0.007
	(0.214)	(0.031)	(2.745)	(0.042)	(0.197)	(0.327)	(0.032)	(0.037)
Observations	975	975	2656	975	975	975	975	954
Mean dep. var.: White advisor	2.976	0.848	75.930	0.705	2.394	5.156	0.565	0.203
$\gamma := \beta_1 + \beta_2$	0.030	-0.054	-0.589	-0.074	-0.111	-0.103	0.023	-0.006
$\gamma = 0$ (p-value)	0.899	0.102	0.837	0.086	0.599	0.668	0.494	0.888
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Liberals								
Black advisor (β_1)	0.204	0.020	1.009	0.002	0.116	-0.538*	0.016	0.055
	(0.203)	(0.030)	(3.072)	(0.040)	(0.204)	(0.303)	(0.031)	(0.037)
Black advisor × High piece rate (β_2)	-0.248	-0.020	-3.147	0.009	-0.169	0.526	0.011	0.026
	(0.298)	(0.043)	(3.956)	(0.055)	(0.295)	(0.451)	(0.043)	(0.052)
High piece rate	0.280	-0.028	-2.746	-0.024	0.237	-1.351***	0.116***	0.002
	(0.212)	(0.031)	(2.802)	(0.039)	(0.207)	(0.342)	(0.031)	(0.031)
Observations	1082	1082	3244	1082	1082	1082	1082	1074
Mean dep. var.: White advisor	2.976	0.848	80.680	0.705	2.394	5.156	0.565	0.156
$\gamma := \beta_1 + \beta_2$	-0.044	0.000	-2.139	0.011	-0.053	-0.013	0.027	0.082
$\gamma = 0$ (p-value)	0.837	0.991	0.397	0.775	0.802	0.971	0.386	0.027
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Sample: follow-up study participants. Dependent variables: Number of puzzles solved in first stage in Column (1), dummy for learners who solved at least one puzzle in first stage in Column (2), average number of moves needed to solve a puzzle in first stage in Column (3), dummy for learners who solved at least one puzzle in first stage using first-stage strategy in Column (4), number of puzzles solved in first stage using first-stage strategy in Column (5), expected number of puzzles solved in first stage in Column (6), WTP for full tutorial in first stage in Column (7), and dummy indicating strict preference for a tutor in second stage in Column (8). Column (3) uses puzzle level data and conditions on subjects solving at least one puzzle in the first stage. Column (8) restricted to participants assigned to tutors of different races across stages. All other columns use subject-level data. Puzzle fixed effects in Column (3). Column (8) controls for instructor's hand model, voice, and stage. All other regressions include strata controls. Robust standard errors in parentheses (learner-level clusters in Columns (3) and (8)). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

G Mediation Analysis

This subsection clarifies the details of our mediation analysis.

2SLS estimator: The details are as follows: We construct a panel dataset with two observations per participant—one for each stage's tutor—and restrict our analysis to participants who faced both a Black and a white tutor, one in each stage (see also footnote [38]). We then estimate the parameters of the following model with a two-stage least squares (2SLS):

$$\text{First stage:} \qquad Belief_{is} = \alpha_0 + \alpha_1 Black_{is} + \alpha_2 Black_{is} \times Inf_i + \alpha_3 Inf_i + X_{is}'\gamma + \varepsilon_i, \tag{4}$$

$$\text{Second stage:} \qquad Y_{is} = \beta_0 + \beta_1 Black_{is} + \beta_2 \widehat{Belief}_{is} + X_{is}'\delta + \epsilon_i, \tag{5}$$

where $Belief_{is}$ measures participant i's expected number of solved puzzles (in stage 2) when the stage-s tutor would present the stage-2 tutorial. Moreover, the dummy Inf_i measures if individual i receives the information treatment and $Black_{is}$ indicates if i's tutor in stage s is Black or not.

Causal effects: Models (4) and (5) identify two types of causal effects. First, equation (5) identifies the average direct effect ($\widehat{ADE} = \widehat{\beta}_1$), representing the direct causal effect of the skin-color treatment on the outcome, which is not transmitted through beliefs. A positive estimate of β_1 indicates a preference for a Black tutor. Second, we can also estimate the (local) average causal mediation effect as $\widehat{ACME} = \widehat{\alpha}_1 \times \widehat{\beta}_2$. Here, $\widehat{\alpha}_1$ shows how the tutor's race impacts beliefs in the no-information treatment. Finally, $\widehat{\beta}_2$ shows how an change in beliefs affects the probability of strictly preferring a tutor. Taken together, the $\widehat{ACME}$ measures the effect of the skin-color treatment on the outcome running through beliefs. The information treatment (designed to equalize beliefs across tutor) serves as an excluded instrument. It provides us with the necessary identifying variation in beliefs to estimate β_2 and allows us to estimate β_1 and β_2 simultaneously. The model assumes, among other things, that the exclusion restriction of the instrument holds (i.e., that the information treatment only affects the outcome through beliefs).

Results: Appendix Table G1 presents the results of the mediation analysis. Given that the tutor's race has no statistically significant effect on tutor selection for conservatives, we discuss only the results for liberals. The table shows that the tutor's race does not impact liberals' beliefs about their own performance (Column 1 in Panel B). This implies that the indirect effect running through beliefs ($ACME$) is small and statistically insignificant. The ADE, thus, almost entirely explains liberals' choice of Black over white tutors. In summary, conditional on having identical beliefs about their own productivity when being advised by Black and white tutors, liberals exhibit a strict preference for Black tutors.

Table G1: Mediation Analysis

	Belief about #puzzles solved (1)	Incentivized ranking of advisors (2)
Panel A: Conservatives		
Black advisor	0.033	-0.002
	(0.077)	(0.027)
Information	-0.300	
	(0.227)	
Black advisor × Information	-0.026	
	(0.106)	
Belief		0.125
		(0.115)
Observations	1152	1152
ACME		0.004
p-value		0.931
Controls		Yes
Panel B: Liberals		
Black advisor	0.044	0.054**
	(0.071)	(0.027)
Information	-0.302	
	(0.210)	
Black advisor × Information	0.075	
	(0.094)	
Belief		0.184
		(0.163)
Observations	1208	1208
ACME		0.008
p-value		0.850
Controls		Yes

Notes: Mediation analysis based on 2SLS regressions, estimated using panel data. Sample: All subjects who received advice from a Black advisor in one stage and from a White advisor in the other stage. First stage in Column (1). Dependent variables: Belief about the number of puzzled solved in the second stage under the first or second advisor in Column (1), and dummy indicating strict preference for a tutor in second stage in Column (2). All models include controls. Standard errors clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

H Derivations of Predictions Within Our Framework

In this Appendix, we derive stylized predictions sketching how non-discriminators (Subsection H.1), statistical discriminators (Subsection H.2), taste-based discriminators (Subsection H.3), and unconscious discriminators (Subsection H.4) behave in our experiment. We consider a simple model with two stages. The first stage is the *instructor-selection stage.* Here, individuals choose between one Black (B) and one white (W) instructor. The second stage is the *knowledge-acquisition-and-utilization stage*. For simplicity, individuals face a binary decision whether or not to acquire knowledge (i.e., to learn) from the instructor. They then apply the acquired knowledge in puzzle-solving. For ease of exposition, we refer to the learner using male pronouns. All results and conclusions apply equally regardless of gender.

H.1 Non-Discriminators

The following subsections have similar structures. They first define the considered type of individual within our utility framework (e.g., non-discriminators) and then solve the model for that type using backward induction. Hence, they always start with solving the knowledge-acquisition-and-utilization stage and then move on to the instructor-selection stage.

H.1.1 Expected Utility Function

Parameters: For non-discriminators, the parameters of the expected utility function apply consistently across instructors $j \in \{B, W\}$:

$$\begin{aligned} \text{Beliefs about effectiveness:} \quad & \phi^B = \phi^W = \phi \\ \text{Disutility of interacting with } j\text{:} \quad & \tau^B = \tau^W = 0 \\ \text{Conscious cost of learning:} \quad & c^B = c^W = 0 \\ \text{Unconscious cost of learning:} \quad & c^{B,u} = c^{W,u} = 0 \\ \text{General learning cost:} \quad & \bar{c} > 0. \end{aligned}$$

Expected utility: Because a non-discriminator i does not differentiate between Black and white instructors, his expected utility level is the same for both. The expected utility function for all $j \in \{W, B\}$ reads:

$$U_i^j(L_i) \;=\; r_i \cdot \left[L_i \cdot \phi \cdot E[P_i^{Max}] + (1 - L_i) \cdot E[P_i^{L_i=0}]\right] - L_i \cdot \bar{c}. \tag{6}$$

H.1.2 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Decision rule: To decide whether or not to acquire knowledge from instructor $j \in \{B, W\}$, i compares the expected utility of acquiring ($L_i = 1$) and not acquiring knowledge ($L_i = 0$). He will acquire knowledge (i.e., learn) from j if:

$$U_i^j(1) \geq U_i^j(0) \tag{7}$$

Optimal knowledge-acquisition decision: Inserting equation (6) into equation (7) leads to:

$$r_i \cdot \phi \cdot E[P_i^{Max}] - \bar{c} \geq r_i \cdot E[P_i^{L_i=0}].$$

Therefore, the individual decides to learn from j if:

$$\overbrace{r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right)}^{\text{Expected gains of acquiring knowledge}} \geq \overbrace{\overline{c}}^{\text{General learning cost}}$$

The outcome of this comparison results in the optimal action $\overline{L}_i^j \in \{0,1\}$. Importantly, the optimal decision is identical for Black and white instructors $\overline{L}_i^B = \overline{L}_i^W$.

Case of interest: In the following, we consider the scenario in which learning costs are sufficiently low, and the expected benefits are high enough to motivate individuals to learn from the instructor. Without imposing this condition, the experiment is irrelevant (as individuals always refuse to learn from j). For both instructors, we have:

$$\overline{L}_i^B = \overline{L}_i^W = 1.$$

Performance: In the current example, i decides to acquire knowledge from both instructors. The performance $E[P_i^j]$ under both instructors $j \in B, W$ is:[14]

$$E[P_i^B] = E[P_i^W] = \phi \cdot E[P_i^{Max}].$$

Hence, the belief parameter $\phi \in [0,1]$ determines the realized performance when learning. This reflects the idea that learners who expect to learn less, indeed, learn less as they engage less (e.g., by paying less attention or processing advice less deeply).

Belief about performance: Individual i is rational and expects to solve $\phi \cdot E[P_i^{Max}]$ puzzles under both instructors.

H.1.3 First Stage: Instructor-Selection Stage

Decision rule: Based on the optimal action $\overline{L}_i^j = 1 \ \forall \ j \in \{B, W\}$ from the second stage, the participant calculates the expected utility for each instructor j. He then chooses the one with the highest expected utility:

$$\overline{j} = \arg\max_{j \in \{B,W\}} U_i^{\overline{j}}, \tag{8}$$

Optimal instructor-selection decision: In the current example, the expected utility obtained from the Black instructor is:

$$U_i^{\overline{B}} = U_i^B(1) = r_i \cdot \phi \cdot E[P_i^{Max}] - \overline{c},$$

and that of the white instructor is:

$$U_i^{\overline{W}} = U_i^W(1) = U_i^{\overline{B}}.$$

Importantly, because $U_i^{\overline{W}} = U_i^{\overline{B}}$, non-discriminator i is indifferent between the two instructors.

H.1.4 Predictions

We can summarize the predictions for non-discriminators as follows:

- **Instructor selection:** They are indifferent between Black and white instructors.

[14]We implicitly assume that i follows the provided instructions perfectly.

- **Knowledge acquisition:** They equally likely learn from both instructors.
- **Performance:** Their performance is identical under both instructors.
- **Belief:** They expect the same performance.

H.2 Statistical Discriminators

H.2.1 Expected Utility Function

Parameters: *Statistical discriminators* engage solely in statistical discrimination without taste-based or unconscious biases. We can model them within the general expected utility function by setting the following parameters:

$$\begin{aligned}
\text{Beliefs about effectiveness:} \quad & \boldsymbol{\phi^B < \phi^W} = \phi \\
\text{Disutility of interacting with } j\text{:} \quad & \tau^B = \tau^W = 0 \\
\text{Conscious cost of learning:} \quad & c^B = c^W = 0 \\
\text{Unconscious cost of learning:} \quad & c^{B,u} = c^{W,u} = 0 \\
\text{General learning cost:} \quad & \bar{c} > 0.
\end{aligned}$$

Expected utility: A statistical discriminator i perceives white instructors as more effective than Black instructors. Therefore, his expected utility differs based on the instructor's race. The expected utility function for all $j \in \{B, W\}$ is:

$$U_i^j(L_i) = r_i \cdot \left[L_i \cdot \phi^j \cdot E[P_i^{Max}] + (1 - L_i) \cdot E[P_i^{L_i=0}]\right] - L_i \cdot \bar{c}. \tag{9}$$

H.2.2 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Decision rule: Statistical discriminator i decides whether to acquire knowledge from instructor j by comparing the expected utilities of acquiring it ($L_i = 1$) versus not acquiring it ($L_i = 0$). This decision is again guided by equation (7).

Optimal knowledge-acquisition decision: Substituting equation (9) into equation (7), for Black instructors, we obtain:

$$r_i \cdot \left(\phi^B \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \bar{c},$$

and for white instructors:

$$r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \bar{c}.$$

As $\phi^B < \phi$, statistical discriminator i is more likely to learn from white instructors.

Case of interest: Statistical discriminators perceive the instructions of white instructors to be more effective than those of Black ones ($\phi^W > \phi^B$). The most interesting case is the one in which the difference in ϕ^j is substantial enough that statistical discriminators choose to acquire knowledge from white but not Black instructors:

$$\overline{L}_i^W = 1 \quad \text{and} \quad \overline{L}_i^B = 0.$$

We concentrate on this case because, only here, statistical discrimination is strong enough to affect behavior.

Belief about performance: Statistical discriminator i knows that he will only acquire knowledge from white instructors, not Black ones (given his belief that $\phi^B < \phi^W = \phi$). Thus, his expected performance under the two instructors is:

$$\overbrace{E[P_i^W] = \phi \cdot E[P_i^{Max}]}^{\text{Expected performance when learning from } W} \qquad \text{and} \qquad \overbrace{E[P_i^B] = E[P_i^{L_i=0}]}^{\text{Expected performance when not learning from } B} \quad .$$

Because $\phi^B \leq \phi$, i's expected performance with W is higher than with B.

Performance: Individual i follows his beliefs and chooses to acquire knowledge from white but not Black instructors. As a result, his actual performance aligns with his expectations. In other words, although both instructors provide similarly effective instructions in our experiment, i's performance is lower under Black instructors. The reason is that i expects W to provide more effective instructions than B. Therefore, he is more likely to acquire knowledge from the white than Black instructor.[15] Moreover, even if statistical discriminators were to acquire knowledge from Black instructors, their belief that $\phi^B < \phi^W$ would still lead them to engage less. Consequently, their performance would still be lower than with white instructors.

H.2.3 First Stage: Instructor-Selection Stage

Decision rule: Statistical discriminators also base their instructor-selection decision on equation (8). They choose the instructor $\bar{j}$, delivering the highest expected utility (given their optimal knowledge acquisition decision from the second stage).

Optimal instructor-selection decision: From the previous analysis, individual i decides to learn from the white instructor ($\overline{L}_i^W = 1$) but not from the Black instructor ($\overline{L}_i^B = 0$). The expected utility for the white instructor W is:

$$U_i^W(1) = r_i \cdot \phi \cdot E[P_i^{Max}] - \bar{c},$$

and that for the Black instructor B corresponds to:

$$U_i^B(0) = r_i \cdot E[P_i^{L_i=0}].$$

As mentioned, individual i compares the expected utilities $U_i^W(1)$ and $U_i^B(0)$ to decide which instructor to select. Two factors ensure that $U_i^W(1) \geq U_i^B(0)$, leading i to select W over B. First, because i decided to learn from W, we know that $U_i^W(1) \geq U_i^W(0)$. Second, the utility without learning is identical for both instructors: $U_i^W(0) = U_i^B(0)$. Consequently, we find that $U_i^W(1) \geq U_i^W(0) = U_i^B(0)$. Therefore, individual i prefers the white instructor over the Black one.

[15] Strictly speaking, this effect occurs only if individuals hold biased beliefs about the instructions' effectiveness in our experiment. Our design can reveal such biases. Thus, we can determine whether participants bring pre-existing beliefs about Black instructors into the study.

H.2.4 Predictions

We summarize the predictions for statistical discriminators as follows:

- **Instructor selection:** Statistical discriminators prefer white over Black instructors.
- **Knowledge acquisition:** They only acquire knowledge from white instructors.
- **Performance:** They perform better under white than Black instructors.
- **Belief:** They expect a higher performance under white instructors.

H.3 Taste-Based Discriminators

H.3.1 Expected Utility Function

Parameters: *Taste-based discriminators* experience disutility when interacting with Black instructors due to prejudice or bias (Becker, 1957). We model them within the general expected utility function by setting the following parameters:

$$
\begin{aligned}
\text{Beliefs about effectiveness:} \quad & \boldsymbol{\phi^B \leq \phi^W} = \phi \\
\text{Disutility of interacting with } j\text{:} \quad & \boldsymbol{\tau^B > \tau^W} = 0 \\
\text{Conscious cost of learning:} \quad & \boldsymbol{c^B \geq c^W} = 0 \\
\text{Unconscious cost of learning:} \quad & c^{B,u} = c^{W,u} = 0 \\
\text{General learning cost:} \quad & \bar{c} > 0.
\end{aligned}
$$

Expected utility: A taste-based discriminator i experiences disutility when interacting with Black instructors ($\tau^B > 0$) and possibly higher conscious learning costs ($c^B \geq c^W = 0$). Moreover, he may believe Black instructors are less effective than white instructors ($\phi^B \leq \phi^W = \phi$). For $j = W$, his expected utility is:

$$U_i^W(L_i) = r_i \cdot \left[L_i \cdot \phi \cdot E[P_i^{Max}] + (1 - L_i) \cdot E[P_i^{L_i=0}]\right] - L_i \cdot \bar{c}, \tag{10}$$

and for $j = B$:

$$U_i^B(L_i) = r_i \cdot \left[L_i \cdot \phi^B \cdot E[P_i^{Max}] + (1 - L_i) \cdot E[P_i^{L_i=0}]\right] - L_i \cdot (\bar{c} + c^B) - \tau^B. \tag{11}$$

H.3.2 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Decision rule: Taste-based discriminator i decides to acquire knowledge from instructor j by comparing the expected utilities of acquiring ($L_i = 1$) versus not acquiring knowledge ($L_i = 0$). Equation (7) also guides his decision.

Optimal knowledge-acquisition decision: Substituting equation (11) into equation (7), for Black instructors, we obtain:

$$r_i \cdot \left(\phi^B \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq (\bar{c} + c^B).$$

Note that the disutility τ^B cancels out because it is present in both $U_i^B(1)$ and $U_i^B(0)$. Equivalently, we plug equation (10) into equation (7) and get for white instructors:

$$r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \bar{c}.$$

There are two potential reasons why taste-based discriminator i is more likely to learn from white than Black instructors. First, he potentially needs to pay the additional B-specific learning cost (c^B). Second, due to his distaste, he may expect the instructions of Black instructors to be less effective ($\phi^B \leq \phi$).

Case of interest: We focus on the scenario in which taste-based discrimination is strong enough that i chooses to learn from the white instructor but not from the Black one.

$$\overline{L}_i^W = 1 \quad \text{and} \quad \overline{L}_i^B = 0.$$

Specifically, i decides to learn from W because the expected gains from learning exceed the general learning cost ($\bar{c}$). Instead, i refuses to learn from B because the combination of lower perceived effectiveness ($\phi^B \leq \phi$) and higher conscious learning costs ($c^B \geq 0$) render learning sufficiently unattractive. We concentrate on this case because only here is taste-based discrimination strong enough to affect behavior.

Performance: Taste-based discriminator i decides to learn from the white instructor but not from the Black one. Therefore, his performance is:

$$\overbrace{E[P_i^W] = \phi \cdot E[P_i^{Max}]}^{\text{Performance when learning from } W} \qquad \text{and} \qquad \overbrace{E[P_i^B] = E[P_i^{L_i=0}]}^{\text{Performance when not learning from } B} \quad .$$

Belief about performance: Individual i is fully aware of his biases (i.e., he knows his parameters ϕ^B, ϕ^W, τ^B, τ^W, c^B, and c^W). He, thus, correctly predicts his performance with white and Black instructors.

H.3.3 First Stage: Instructor-Selection Stage

Decision rule: Taste-based discriminators also base their instructor-selection decision on equation (8). They choose the instructor $\bar{j}$, delivering the highest expected utility (given their optimal knowledge acquisition decision from the second stage).

Optimal instructor-selection decision: We study the case in which taste-based discriminator i decides to learn from the white instructor ($\overline{L}_i^W = 1$) but not from the Black one ($\overline{L}_i^B = 0$). The expected utility for the white instructor W is:

$$U_i^W(1) = r_i \cdot \phi \cdot E[P_i^{Max}] - \bar{c}.$$

and that for the one B corresponds to:

$$U_i^B(0) = r_i \cdot E[P_i^{L_i=0}] - \tau^B.$$

Taste-based discriminator i compares the expected utilities $U_i^W(1)$ and $U_i^B(0)$ to decide which instructor to select. Which instructor does i select? The argument is similar to that for statistical discriminators:

First, because i decided to learn from W, we know that $U_i^W(1) \geq U_i^W(0)$. Second, due to the distaste i perceives when interacting with Black people (τ^B), even the utility without learning is smaller for Black than white instructors: $U_i^B(0) < U_i^W(0)$. Consequently, we know that $U_i^W(1) \geq U_i^W(0) > U_i^B(0)$. Individual i then selects the white instructor over the Black one. In conclusion, due to their distaste parameter τ^B, taste-based discriminators are likelier to choose white over Black instructors than statistical discriminators.

H.3.4 Predictions

In sum, taste-based discriminators behave in the same way as statistical discriminators:

- **Instructor selection:** They prefer white over Black instructors.
- **Knowledge acquisition:** They only acquire knowledge from white instructors.
- **Performance:** They perform better under white than Black instructors.
- **Belief:** They expect a higher performance under white instructors.

H.4 Unconscious Discriminators

H.4.1 Expected Utility Function

Parameters: *Unconscious discriminators* experience an unconscious cost of learning from Black instructors. We model them within the general expected utility function by setting the following parameters:

$$\begin{aligned}
\text{Beliefs about effectiveness:} &\quad \phi^B = \phi^W = \phi \\
\text{Disutility of interacting with } j\text{:} &\quad \tau^B = \tau^W = 0 \\
\text{Conscious cost of learning:} &\quad c^B = c^W = 0 \\
\text{Unconscious cost of learning:} &\quad \boldsymbol{c^{B,u} > c^{W,u}} = 0 \\
\text{General learning cost:} &\quad \bar{c} > 0.
\end{aligned}$$

Expected utility: Individual i believes that both instructors are equally effective ($\phi^B = \phi^W = \phi$), experience no disutility when interacting with either instructor ($\tau^B = \tau^W = 0$), and perceives no conscious learning costs ($c^B = c^W = 0$). However, he unknowingly incurs an unconscious cost of learning from Black instructors ($c^{B,u} > 0$). Unconscious discriminator i expects to make his (binary) learning decision based on the expected utility function:

$$U_i^j(L_i) = r_i \cdot \left[L_i \cdot \phi \cdot E[P_i^{Max}] + (1 - L_i) \cdot E[P_i^{L_i=0}] \right] - L_i \cdot \bar{c}. \tag{12}$$

However, his actual (unconscious) expected decision utility is:

$$U_i^{j,\text{actual}}(L_i) = U_i^j(L_i) - L_i \cdot c^{j,u}, \tag{13}$$

where $c^{B,u} > c^{W,u} = 0$.

H.4.2 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Perceived decision rule: Unconscious discriminator i decides whether to acquire knowledge from instructor j by comparing the expected utilities of acquiring ($L_i = 1$) versus not acquiring knowledge ($L_i = 0$). Being unaware of the unconscious costs, he expects to base this decision on the utility function $U_i^j(L_i)$. Put differently, he expects to learn from instructor j if:

$$U_i^j(1) \geq U_i^j(0). \tag{14}$$

Actual decision rule: In reality, his decision is based on his actual utility function $U_i^{j,\text{actual}}(L_i)$ that includes the unconscious cost $c^{j,u}$. He will actually learn from instructor j if:

$$U_i^{j,actual}(1) \geq U_i^{j,actual}(0). \tag{15}$$

Note that for white instructors, we have $c^{W,u} = 0$ such that $U_i^W(L_i) = U_i^{W,actual}(L_i)$.

Perceived knowledge-acquisition decision: Substituting equation (12) into equation (14), we find for both instructors $j \in \{B, W\}$:

$$r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \bar{c}.$$

Unconscious discriminators expect to acquire knowledge from instructor j whenever this condition is satisfied. Because they hold no conscious biases and expect both instructors to be equally effective, they apply the same perceived decision rule for each instructor.

Actual optimal knowledge-acquisition decision: The actual knowledge-acquisition decision, however, depends on the unconscious learning cost. Individual i will actually learn from a Black instructor if:

$$r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq (\bar{c} - c^{B,u}).$$

There is no unconscious learning for white instructors ($c^{W,u} = 0$). Thus, for white instructors, they decide based on the standard conditions:

$$r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \bar{c}.$$

In conclusion, due to the unconscious learning cost $c^{B,u}$, individual i is more likely to acquire knowledge from white than Black instructors.

Case of interest: We focus on the scenario in which individual i expects to learn from both instructors (as $r_i \cdot (\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]) \geq \bar{c}$), but the unconscious cost $c^{B,u}$ is large enough that:

$$\begin{array}{ccc} \overbrace{r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right)}^{\text{Expected gains of acquiring knowledge from } W} & \geq & \overbrace{\bar{c}}^{\text{Cost of learning from } W} \quad , \\ \underbrace{r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right)}_{\text{Expected gains of acquiring knowledge from } B} & < & \underbrace{(\bar{c} - c^{B,u})}_{\text{Cost of learning from } B} \quad . \end{array}$$

In this case, he actually learns from the white instructor but not from the Black one due to the unconscious cost $c^{B,u}$:

$$\overline{L}_i^W = 1 \quad \text{and} \quad \overline{L}_i^B = 0.$$

Performance: Because i learns from the white instructor but not from the Black instructor, his performance is:

$$\overbrace{E[P_i^W] = \phi \cdot E[P_i^{Max}]}^{\text{Performance when learning from } W} \qquad \text{and} \qquad \overbrace{E[P_i^B] = E[P_i^{L_i=0}]}^{\text{Performance when not learning from } B} \qquad .$$

Belief about performance: Because individual i is unaware of his B-specific learning cost, he believes that he will acquire the knowledge provided by both instructors. He, thus, expects to perform equally well under both instructors:

$$\overbrace{E[P_i^W] = \phi \cdot E[P_i^{Max}]}^{\text{Expected performance when learning from } W} \qquad \text{and} \qquad \overbrace{E[P_i^B] = \phi \cdot E[P_i^{Max}]}^{\text{Expected performance when learning from } B} \qquad .$$

However, due to not learning from the Black instructor, his actual performance under B is lower than anticipated.

H.4.3 First Stage: Instructor-Selection Stage

Decision rule: Importantly, unconscious discriminators choose the instructor that maximizes his expected utility based on his *perceived* (i.e., not actual) utility function $U_i^j(L_i)$. As discussed, i expects to learn from both instructors ($\overline{L}_i^W = \overline{L}_i^B = 1$). Therefore, he selects the instructor based on:

$$\overline{j} = \arg\max_{j \in \{B,W\}} U_i^j(1).$$

Actual instructor-selection decision: Substituting $L_i = 1$ into equation (12), we have:

$$U_i^W(1) = r_i \cdot \phi \cdot E[P_i^{Max}] - \overline{c},$$
$$U_i^B(1) = r_i \cdot \phi \cdot E[P_i^{Max}] - \overline{c}.$$

Unconscious discriminator i perceives no difference in expected utilities between the two instructors and is, thus, *indifferent* between them.

H.4.4 Predictions

We summarize the predictions for unconscious discriminators as follows:

- **Instructor selection:** They are indifferent between Black and white instructors.
- **Knowledge acquisition:** They expect to acquire knowledge from both instructors but, due to unconscious costs, actually learn only from white ones.
- **Performance:** They perform better under white than Black instructors.
- **Belief:** They expect to perform equally well with both instructors.

I Impact of Piece Rate on Discrimination

This section analyzes how changes in the piece rate r_i affect discriminatory behavior in our model. To that end, it examines the impacts of r_i on knowledge acquisition, instructor selection, and performance for each type of individual. Again, we consider the case in which (a) participants are randomly assigned to a white or Black instructor (no selection bias), (b) both instructors provide equally effective instructions (beliefs can differ due to factors outside the experiment), and (c) individuals receive no additional information on the instructors' quality (no-info treatment). Moreover, we begin by considering the baseline case where discrimination exists and then assess how increasing r_i influences behavior.

I.1 Statistical Discriminators

I.1.1 Baseline Scenario

In the baseline scenario, statistical discriminators (a) refuse to learn from Black instructors but learn from white ones (knowledge-acquisition stage) and (b) choose white over Black instructors (instructor-selection stage).

I.1.2 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Decision rule: Statistical discriminator i decides to acquire knowledge from instructor j if the expected gains from learning exceed the learning cost:

$$r_i \cdot \left(\phi^j \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \bar{c}, \tag{16}$$

where $\phi^B < \phi^W = \phi$.

Impact of r_i on learning: We can analyze the impact on learning by computing the derivative of the expected gain from acquiring knowledge from instructor j (i.e., the left-hand side of equation 16) with respect to r_i:

$$\frac{\partial}{\partial r_i}\left[r_i \cdot \left(\phi^j \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right)\right] = \phi^j \cdot E[P_i^{Max}] - E[P_i^{L_i=0}].$$

This derivative is (a) most likely positive for both instructors and (b) larger for the white ones (given that $\phi^W > \phi^B$). Therefore, increasing r_i increases the expected gains from learning from both instructors. However, the increase is larger for W than B. If r_i becomes sufficiently large such that:

$$r_i \geq \frac{\bar{c}}{\phi^B \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]},$$

then statistical discriminators will also stop refusing knowledge from Black instructors. In this case, we no longer observe statistical discrimination in (binary) knowledge acquisition.

Impact of r_i on performance: While a higher r_i may lead statistical discriminators to start learning from Black instructors, their belief that $\phi^B < \phi^W$ still affects how they engage with the instruction. Because they expect to learn less from Black instructors, they engage less (e.g., by paying less attention or processing advice less deeply), which leads to lower realized performance ($\phi^B \cdot E[P_i^{Max}] < \phi^W \cdot E[P_i^{Max}]$). As a result, they continue to perform (and expect to perform) better with white instructors.

I.1.3 First Stage: Instructor-Selection Stage

Decision rule: Statistical discriminator i selects the instructor j that maximizes his expected utility:

$$\bar{j} = \arg \max_{j \in \{B,W\}} U_i^j(\overline{L}_i^j), \tag{17}$$

where $\overline{L}_i^j$ is the optimal knowledge-acquisition decision for instructor j and

$$U_i^j(\overline{L}_i^j) = \begin{cases} r_i \cdot \phi^j \cdot E[P_i^{Max}] - \bar{c}, & \text{if } \overline{L}_i^j = 1, \\ r_i \cdot E[P_i^{L_i=0}], & \text{if } \overline{L}_i^j = 0. \end{cases}$$

Impact of r_i on the instructor-selection decision: We already know that i prefers the white instructor if he decides to acquire knowledge from the white ($\overline{L}_i^B = 1$) but not the Black one ($\overline{L}_i^B = 0$). What is the impact of r_i if r_i is already large enough so that $\overline{L}_i^B = \overline{L}_i^W = 1$? To answer this question, we can calculate the expected utilities under W and B for $\overline{L}_i^B = \overline{L}_i^W = 1$:

$$U_i^B(1) = r_i \cdot \phi^B \cdot E[P_i^{Max}] - \bar{c},$$
$$U_i^W(1) = r_i \cdot \phi^W \cdot E[P_i^{Max}] - \bar{c}.$$

Because $\phi^W > \phi^B$, we have $U_i^W(1) > U_i^B(1)$. Specifically, the positive difference in expected utilities is:

$$U_i^W(1) - U_i^B(1) = r_i \cdot (\phi^W - \phi^B) \cdot E[P_i^{Max}].$$

This difference increases linearly in r_i. Thus, as r_i increases, statistical discriminators will continue to prefer the white instructor. Indeed, the preference for white instructors increases in r_i.

Conclusion: Even if statistical discriminators start learning from both instructors due to higher r_i, they will still select the white instructor over the Black one. The reason is the higher expected utility associated with $\phi^W > \phi^B$. Thus, increasing r_i may reduce discrimination in knowledge acquisition but does not eliminate discrimination in instructor selection.

I.2 Taste-Based Discriminators

I.2.1 Baseline Scenario

In the baseline scenario, taste-based discriminators (a) refuse to learn from Black instructors but learn from white ones (knowledge-acquisition stage) and (b) choose white over Black instructors (instructor-selection stage).

I.2.2 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Decision rule: Taste-based discriminator i also decides to acquire knowledge from instructor j if:

$$r_i \cdot \left(\phi^j \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \bar{c} + c^j, \tag{18}$$

where $\phi^B \leq \phi^W = \phi$ and $c^B > c^W = 0$

Impact of r_i on learning: Again, we can analyze the impact on learning by computing the derivative of the expected gain from acquiring knowledge from instructor j (i.e., the left-hand side of equation 16) with respect to r_i:

$$\frac{\partial}{\partial r_i}\left[r_i \cdot \left(\phi^j \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right)\right] = \phi^j \cdot E[P_i^{Max}] - E[P_i^{L_i=0}].$$

The conclusions from this analysis are similar to the ones for statistical discriminators: Increasing r_i increases the expected gains from learning from both instructors. However, if $\phi^B < \phi^W$, the increase is larger for W than B. If r_i becomes sufficiently large such that:

$$r_i \geq \frac{\bar{c} + c^B}{\phi^B \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]},$$

then taste-based discriminators will stop refusing to acquire knowledge from Black instructors (i.e., no taste-based discrimination in learning). However, due to the additional costs (c^B) and perhaps lower ϕ^B, the required r_i to satisfy equation (18) is higher.

Impact of r_i on performance: The expected gains from learning from the white instructor remain higher if $\phi^W > \phi^B$. Therefore, while higher r_i may lead taste-based discriminators to start learning from Black instructors, they still may perform and expect to perform better under white ones ($\phi^B \cdot E[P_i^{Max}] < \phi^W \cdot E[P_i^{Max}]$). As with statistical discriminators, this reflects both perceived differences in effectiveness and reduced engagement when learning from Black tutors.

I.2.3 First Stage: Instructor-Selection Stage

Decision rule: Taste-based discriminator i selects the instructor j that maximizes his expected utility:

$$\bar{j} = \arg\max_{j \in \{B,W\}} U_i^j(\overline{L}_i^j), \tag{19}$$

where $\overline{L}_i^j$ is the optimal knowledge-acquisition decision for instructor j and

$$U_i^j(\overline{L}_i^j) = \begin{cases} r_i \cdot \phi^j \cdot E[P_i^{Max}] - (\bar{c} + c^j) - \tau^j, & \text{if } \overline{L}_i^j = 1, \\ r_i \cdot E[P_i^{L_i=0}] - \tau^j, & \text{if } \overline{L}_i^j = 0. \end{cases}$$

Impact of r_i on the instructor-selection decision: Again, the (new) interesting case is the one in which r_i is large enough so that i decides to learn from both instructors ($\overline{L}_i^B = \overline{L}_i^W = 1$). Here, the expected utilities are:

$$U_i^B(1) = r_i \cdot \phi^B \cdot E[P_i^{Max}] - (\bar{c} + c^B) - \tau^B,$$
$$U_i^W(1) = r_i \cdot \phi^W \cdot E[P_i^{Max}] - \bar{c}.$$

Individual i chooses the white over the Black instructor if the following difference in expected utilities is positive:

$$U_i^W(1) - U_i^B(1) = r_i \cdot (\phi^W - \phi^B) \cdot E[P_i^{Max}] + c^B + \tau^B.$$

Even if $\phi^B = \phi^W$ and $c^B = 0$, the disutility $\tau^B > 0$ ensures that $U_i^W(1) > U_i^B(1)$. Moreover, if $\phi^W > \phi^B$, an increase in r_i further increases the gap in expected utilities.

Conclusion: While an increasing piece rate r_i may lead taste-based discriminators to start learning from Black instructors, they will still prefer white over Black instructors. That is because (a) a disutility parameter ($\tau^B > 0$), (b) higher learning costs ($c^B \geq 0$), and (c) lower perceived effectiveness ($\phi^B \leq \phi^W$). Thus, discrimination in instructor selection may persist even as discrimination in knowledge acquisition decreases.

I.3 Unconscious Discriminators

I.3.1 Baseline Scenario

In the baseline scenario, unconscious discriminators do not acquire knowledge from Black instructors but learn from white ones (knowledge-acquisition stage). They, however, do not anticipate this behavior. Consequently, they are indifferent between selecting white and Black instructors (instructor-selection stage).

I.3.2 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Decision rule: Unconscious discriminator i will actually learn from instructor j if:

$$r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \bar{c} + c^{j,u}, \tag{20}$$

where $c^{B,u} > c^{W,u} = 0$

Impact of r_i on learning: Again, we can compute the derivative of the expected gain of learning to r_i to highlight the impact on the learning decision:

$$\frac{\partial}{\partial r_i}\left[r_i \cdot \left(\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right)\right] = \phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}].$$

The equation delivers two main insights: First, an increase in r_i increases the expected gains from learning equally for both instructors. Thus, a higher piece rate makes it more likely that i will acquire knowledge from both instructors. The costs of learning from the instructor are, instead, independent of r_i. Second, if r_i becomes sufficiently large such that:

$$r_i \geq \frac{\bar{c} + c^{B,u}}{\phi \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]},$$

then unconscious discriminators will also start learning from the Black instructor. Put differently, increasing the piece rate r_i can eliminate discrimination in knowledge acquisition among unconscious discriminators.

Impact of r_i on performance: Taste-based and statistical discriminators perform better under white instructors, even if they choose to learn from Black ones. The reason is that, due to the difference in the instructors' perceived effectiveness, the expected gains from learning from the white instructor remain higher. The predictions for unconscious discriminators are very different: As soon as they start learning from the Black instructor due to an increase in r_i, their performance increases from $E[P_i^{L_i=0}]$ to

$\phi \cdot E[P_i^{Max}]$. However, they also solve $\phi \cdot E[P_i^{Max}]$ puzzles with the white instructor. Put differently, once learning occurs, performance under both instructors is identical—eliminating any disparities caused by unconscious discrimination. In terms of beliefs, unconscious discriminators do not anticipate any performance differences, no matter whether they choose to learn or not.

I.3.3 First Stage: Instructor-Selection Stage

Decision rule: Unconscious discriminators select instructors based on their *perceived* expected utilities, which are equal for both instructors. The reason is that they are unaware of their unconscious learning cost ($c^{B,u} > c^{W,u} = 0$). Formally, we get:

$$\bar{j} = \arg \max_{j \in \{B,W\}} U_i^j(\overline{L}_i^j), \tag{21}$$

where $\overline{L}_i^j$ is the optimal knowledge-acquisition decision for instructor j and

$$U_i^j(\overline{L}_i^j) = \begin{cases} r_i \cdot \phi \cdot E[P_i^{Max}] - \bar{c}, & \text{if } \overline{L}_i^j = 1, \\ r_i \cdot E[P_i^{L_i=0}], & \text{if } \overline{L}_i^j = 0. \end{cases}$$

Impact of r_i on the instructor-selection decision: Unconscious discriminators expect to acquire knowledge from both instructors. Thus, their expected utility for both instructors is:

$$U_i^B(1) = U_i^W(1) = r_i \cdot \phi \cdot E[P_i^{Max}] - \bar{c}.$$

Therefore, as $U_i^W(1) - U_i^B(1) = 0$, they are indifferent between both instructors. An increase in r_i does not alter this result, as it equivalently raises the expected for both instructors.

Conclusion: Increasing r_i reduces discrimination in knowledge acquisition among unconscious discriminators without affecting their instructor selection.

I.4 Summary

We summarize the impact of increasing the piece rate r_i on discrimination:

- **Statistical discriminators:** A higher piece rate r_i may reduce discrimination in knowledge acquisition. The reason is that statistical discriminators start learning from Black instructors when expected gains from learning outweigh its associated costs. However, statistical discriminators will continue to select white over Black instructors due to perceived differences in the instructors' effectiveness ($\phi^W > \phi^B$). Thus, discrimination in instructor selection persists.
- **Taste-based discriminators:** A higher r_i may also counteract taste-based discrimination in knowledge acquisition. This counteracting effect unfolds if the expected gains from learning additionally compensate for the additional costs and disutility (c^B and τ^B). However, taste-based discriminators select white over Black instructors due to persistent disutility and perceived differences. Therefore, discrimination in instructor selection remains for them.
- **Unconscious discriminators:** An increase in the piece rate r_i can eliminate discrimination in knowledge acquisition by overcoming unconscious costs, leading them to learn from both instructors equally. The instructor selection decision remains unbiased.

J Impact of Information Treatment on Discrimination

Next, we discuss how the information treatment affects statistical, taste-based, and unconscious discrimination in our model. The treatment provides individuals with information to equalize their perceived effectiveness of Black and white instructors (ϕ^j). Considering each type of discrimination separately, we begin by recapping the baseline case (without information) in which discrimination exists. We then assess how the information treatment changes knowledge acquisition, instructor selection, and performance.

J.1 Statistical Discriminators

J.1.1 Baseline Scenario

In the baseline scenario, statistical discriminators believe Black instructors are less effective than white instructors ($\phi^B < \phi^W = \phi$). Statistical discriminators (a) refuse to learn from Black instructors but learn from white ones (knowledge-acquisition stage) and (b) choose white over Black instructors (instructor-selection stage).

J.1.2 Information Treatment

A well-designed information treatment provides credible evidence that Black and white instructors are equally effective. Statistical discriminators are assumed to be unbiased. Thus, they should update their beliefs accordingly, leading to $\phi^B = \phi^W = \phi$.

J.1.3 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Decision rule: Statistical discriminator i decides to acquire knowledge from instructor j if the expected gains from learning exceed the learning cost:

$$r_i \cdot \left(\phi^j \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \bar{c}, \tag{22}$$

where the information treatment ensures that $\phi^B = \phi^W = \phi$.

Impact of the information treatment on learning: By equalizing ϕ^B and ϕ^W, the expected gain from learning (left-hand side of the inequality) becomes identical for both instructors. Consequently, statistical discriminators are now equally likely to acquire knowledge from Black and white instructors ($\overline{L}_i^B = \overline{L}_i^W = 1$). The information treatment eliminates discrimination in knowledge acquisition among statistical discriminators.

Impact of the information treatment on performance: The expected gains from learning from the white and Black instructors are now identical. Therefore, statistical discriminators (expect to) perform equally well under both types of instructors.

J.1.4 First Stage: Instructor-Selection Stage

Decision rule: Statistical discriminator i selects the instructor j that maximizes his expected utility:

$$\bar{j} = \arg\max_{j \in \{B,W\}} U_i^j(\overline{L}_i^j), \tag{23}$$

where $\overline{L}_i^j$ is the optimal knowledge-acquisition decision for instructor j and

$$U_i^j(\overline{L}_i^j) = \begin{cases} r_i \cdot \phi \cdot E[P_i^{Max}] - \overline{c}, & \text{if } \overline{L}_i^j = 1, \\ r_i \cdot E[P_i^{L_i=0}], & \text{if } \overline{L}_i^j = 0. \end{cases}$$

Impact of information treatment on the instructor-selection decision: Because $\phi^B = \phi^W = \phi$ and $\overline{L}_i^B = \overline{L}_i^W$, the expected utilities are now equal:

$$U_i^B(\overline{L}_i^B) = U_i^W(\overline{L}_i^W) = r_i \cdot \phi \cdot E[P_i^{Max}] - \overline{c}.$$

Statistical discriminators become indifferent between selecting Black and white instructors.

Conclusion: The information treatment eliminates discrimination in knowledge acquisition and instructor selection. It also equalizes i's performance across instructors.

J.2 Taste-Based Discriminators

J.2.1 Baseline Scenario

In the baseline scenario, taste-based discriminators believe Black instructors are less effective than white instructors ($\phi^B < \phi^W = \phi$). Taste-based discriminators (a) refuse to learn from Black instructors but learn from white ones (knowledge-acquisition stage) and (b) choose white over Black instructors (instructor-selection stage).

J.2.2 Information Treatment

The information treatment provides evidence that Black and white instructors are equally effective. However, due to their bias, it is not clear if taste-based discriminators consider this signal credible. We consider two cases. In the first case, taste-based discriminators update their perceived effectiveness, resulting in $\phi^B = \phi^W = \phi$. In the second case, taste-based discriminators reject the provided information (due to their distaste) and continue to believe that $\phi^B < \phi^W = \phi$.

J.2.3 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Decision rule: Taste-based discriminator i decides to acquire knowledge from instructor j if:

$$r_i \cdot \left(\phi^j \cdot E[P_i^{Max}] - E[P_i^{L_i=0}]\right) \geq \overline{c} + c^j, \tag{24}$$

where $c^B > c^W = 0$. There are situations in which there are clear (monetary) gains of learning, but these are not large enough to compensate for the additional learning cost c^B. Here, individuals are willing to earn less to avoid the additional learning cost.

Case 1: Impact of the information treatment on learning: When the information treatment equalizes perceived effectiveness ($\phi^B = \phi^W = \phi$), individual i perceives the gains from learning (left-hand side) to be the same for both instructors. This outcome removes taste-based discrimination driven by differences in perceived effectiveness. However, the additional learning cost c^B associated with Black instructors can still discourage i from acquiring knowledge from them. As a result, the information treatment does not eliminate discrimination in knowledge acquisition if the additional learning cost c^B is large enough ($\overline{L}_i^B = 0$ and $\overline{L}_i^W = 1$).

Case 1: Impact of the information treatment on performance: As always, differences in learning translate into (expected) performance differences.

Case 2: Impact of the information treatment on learning and performance: In the second case, taste-based discriminator i ignores the provided information ($\phi^B < \phi^W = \phi$). In that case, the information treatment does not affect his knowledge-acquisition behavior or expected performance.

J.2.4 First Stage: Instructor-Selection Stage

Decision rule: Taste-based discriminator i selects the instructor j that maximizes his expected utility:

$$\bar{j} = \arg \max_{j \in \{B,W\}} U_i^j(\overline{L}_i^j), \tag{25}$$

where $\overline{L}_i^j$ is the optimal knowledge-acquisition decision for instructor j and

$$U_i^j(\overline{L}_i^j) = \begin{cases} r_i \cdot \phi^j \cdot E[P_i^{Max}] - (\bar{c} + c^j) - \tau^j, & \text{if } \overline{L}_i^j = 1, \\ r_i \cdot E[P_i^{L_i=0}] - \tau^j, & \text{if } \overline{L}_i^j = 0. \end{cases}$$

Case 1: Impact of the information treatment on the instructor-selection decision: Even with equal perceived effectiveness ($\phi^B = \phi^W = \phi$), taste-based discriminators may continue to select white over Black instructors. The reason is the parameters $c^B \geq 0$ and $\tau^B > 0$ that ensure $U_i^W(1) > U_i^B(0)$ and $U_i^W(1) > U_i^B(1)$.

Case 2: Impact of the information treatment on the instructor-selection decision: In the second case, the information treatment does not affect i's knowledge-acquisition behavior and (expected) performance. Thus, it also does not change his instructor-selection decision.

Conclusion: The information treatment does not necessarily eliminate discrimination in instructor selection and knowledge acquisition among taste-based discriminators (if the distaste is strong enough). Consequently, performance differences remain.

J.3 Unconscious Discriminators

J.3.1 Baseline Scenario

In the baseline scenario, unconscious discriminators believe that Black and white instructors are equally effective ($\phi^B = \phi^W = \phi$). Unconscious discriminators do not acquire knowledge from Black instructors but learn from white ones (knowledge-acquisition stage). They, however, do not anticipate this behavior. Consequently, they are indifferent between selecting white and Black instructors (instructor-selection stage).

J.3.2 Information Treatment

Because unconscious discriminators already believe that $\phi^B = \phi^W = \phi$, the information treatment does not alter their perceived effectiveness.

J.3.3 Second Stage: Knowledge-Acquisition-And-Utilization Stage

Because (a) unconscious discriminators already believe that $\phi^B = \phi^W = \phi$ and (b) the information treatment should not affect i's unconscious learning cost $c^{B,u}$, it does not reduce discrimination in knowledge acquisition among unconscious discriminators. Also, the performance differences remain.

J.3.4 First Stage: Instructor-Selection Stage

Also, the information treatment does not affect instructor selection among unconscious discriminators. Being unaware of their bias, they are still indifferent between Black and white instructors.

Conclusion: The information treatment does not counteract the discriminatory behavior among unconscious discriminators.

J.4 Summary

We summarize the impact of the information treatment on discrimination:

- **Statistical Discriminators:** The information treatment eliminates discrimination in both knowledge acquisition and instructor selection by equalizing perceived effectiveness ($\phi^B = \phi^W$). Statistical discriminators become equally likely to learn from and select Black and white instructors, and leads to equal performance under both instructors.
- **Taste-based discriminators:** The information treatment does not necessarily eliminate discrimination in knowledge acquisition and instructor selection if the additional learning cost (c^B) and disutility (τ^B) are sufficiently large.
- **Unconscious Discriminators:** The information treatment does not affect unconscious discriminators. They already perceive both instructors as being equally effective. Thus, discrimination in knowledge acquisition persists due to the unconscious learning cost ($c^{B,u}$).

K Additional Results

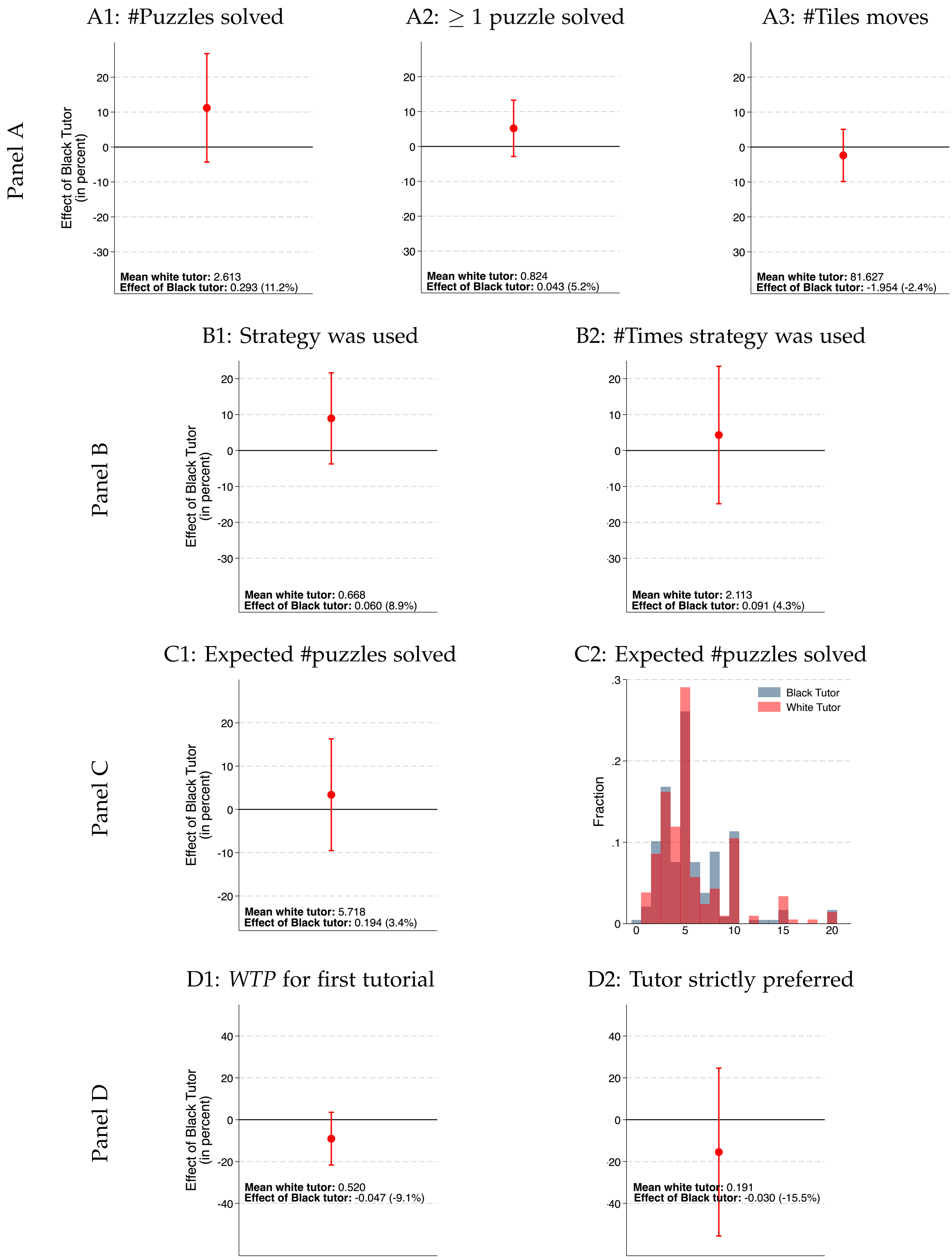
This section presents additional results defined in the pre-analysis plan (PAP). First, we present results of rank-ordered logit specifications defined in the PAP. Second, we present results for "Moderates" ($N = 933$). In the PAP we label individuals as moderates if they are classified by CloudResearch as "liberal" ($N = 316$), "moderate" ($N = 302$), or "conservative" ($N = 315$).

Table K1: Preferences for Tutors: Rank-Ordered Logit

	Conservative	Very Conservative	Liberal	Very Liberal
	(1)	(2)	(3)	(4)
Black advisor (β_1)	0.036	0.057	0.073	0.126**
	(0.051)	(0.062)	(0.044)	(0.050)
Black advisor $\times$ High piece rate (β_2)	-0.078	-0.071	0.028	-0.025
	(0.072)	(0.082)	(0.061)	(0.069)
Observations	1152	854	1208	902
Mean dep. var.: White advisor	1.222	1.211	1.152	1.135
$\gamma := \beta_1 + \beta_2$	-0.042	-0.014	0.101	0.100
$\gamma = 0$ (p-value)	0.420	0.797	0.019	0.038
Controls	Yes	Yes	Yes	Yes

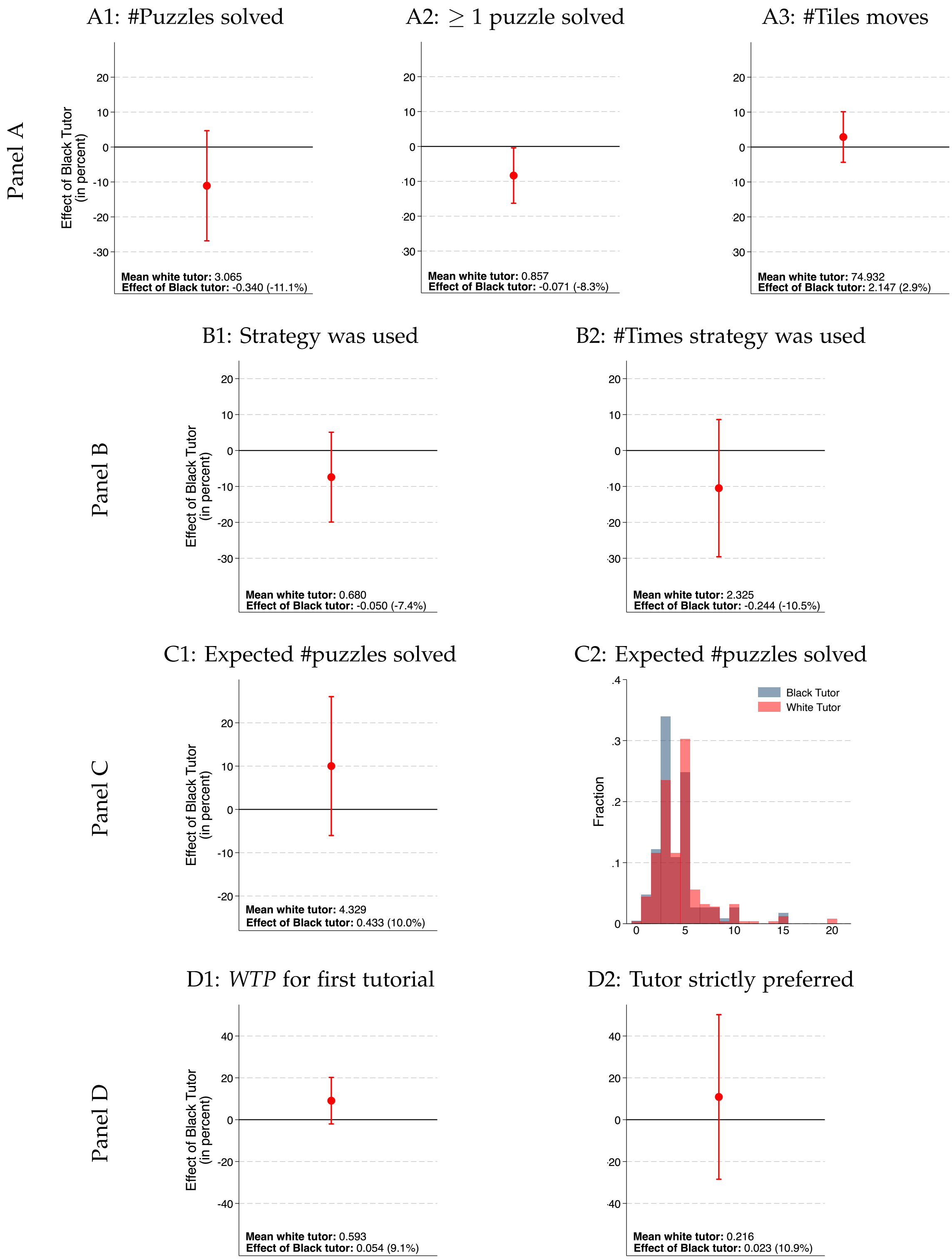
Notes: Rank-ordered logit estimations. Sample: Restricted to participants assigned to tutors of different races across stages, sub-samples defined in column titles. Dependent variable: Indicator for ranking of tutors elicited in the second stage, where value two (one) indicates a tutor who is (not) strictly preferred. The estimations control for instructor's hand model, voice, and stage. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure K1: Results for Moderates Under Standard Piece Rate



Notes: Effects of tutor race on learners' behavior. Sample: Moderates under standard piece rate. Dependent variables: Number of puzzles solved in first stage (K1A1), dummy for learners who solved at least one puzzle in first stage (K1A2), average number of moves needed to solve a puzzle in first stage (K1A3), dummy for learners who solved at least one puzzle in first stage using first-stage strategy (K1B1), number of puzzles solved in first stage using first-stage strategy (K1B2), expected number of puzzles solved in first stage (K1C1), WTP for full tutorial in first stage (K1D1), dummy indicating a strict preference for a tutor in second stage (K1D2). Confidence intervals based on robust standard errors (learner-level clusters in K1A3 and K1D2).

Figure K2: Results for Moderates Under High Piece Rate



Notes: Effects of tutor race on learners' behavior. Sample: Moderates under high piece rate. Dependent variables: Number of puzzles solved in first stage (K1A1), dummy for learners who solved at least one puzzle in first stage (K1A2), average number of moves needed to solve a puzzle in first stage (K1A3), dummy for learners who solved at least one puzzle in first stage using first-stage strategy (K1B1), number of puzzles solved in first stage using first-stage strategy (K1B2), expected number of puzzles solved in first stage (K1C1), WTP for full tutorial in first stage (K1D1), dummy indicating a strict preference for a tutor in second stage (K1D2). Confidence intervals based on robust standard errors (learner-level clusters in K1A3 and K1D2).