Why Do We Discriminate? Directional Beliefs and Biased Information Processing*

Markus Eyting
Johannes Gutenberg University Mainz
Leibniz Institute for Financial Research SAFE

This is the accepted manuscript.
Publication: https://doi.org/10.1287/mnsc.2025.01818

Identifying the cause of discrimination is crucial to design effective policies and to understand discrimination dynamics. This paper provides a potential link between existing models of taste-based and statistical discrimination by studying directional belief formation through biased information search and processing. By selectively acquiring and processing information, individuals form directional beliefs and consequentially discriminate based on these beliefs. Through a series of experiments in which individuals endogenously acquire and process various signals, I document discrimination stemming from directional beliefs and demonstrate important differences in how it manifests compared to taste-based and statistical discrimination. Finally, I show that limiting individuals' scope to interpret information reduces such discrimination.

Keywords: discrimination; directional belief formation; biased information processing; information

*I thank Michael Kosfeld, Florian Hett, Aislinn Bohren, Alex Imas, Sendhil Mullainathan, Emir Kamenica, George Loewenstein, Vojtěch Bartoš, Joël J. van der Weele, Theo Offerman, Jeroen van de Ven, Guido Friebel, Navid Sabet, Jan Hausfeld, Simon Heß, Johannes Kasinger, Patrick Schmidt, and seminar and conference audiences at the JGU Mainz, Goethe University Frankfurt, University Chicago, Discrimination & Diversity workshop at University of East Anglia, ESA Bologna, Esa Lyon, EEA Barcelona, Behavioral Measurements conference Frankfurt, Discrimination and Disparities online Seminar, IMEBESS Lisbon, EGPROC at UvA, and the Identity workshop at University of Konstanz for many valuable comments and discussions. I gratefully acknowledge the financial support from the SAFE Experiment Center Frankfurt, the Interdisciplinary Public Policy Mainz and the Mainzer Wissenschaftsstiftung.

1 Introduction

Discrimination, i.e. the systematically disparate treatment of individuals from different social groups, is widespread. It has been documented in various contexts, such as the labor market (Blau and Kahn, 2017; Neumark, 2018), healthcare (Alsan et al., 2019), the justice system (Arnold et al., 2018; Arnold et al., 2021), policing (Goncalves and Mello, 2021), and education (Farkas, 2003). While discrimination from any source can have serious negative consequences for the discriminated, precise identification of the source of discrimination has important implications for policy, welfare analyses, and discrimination dynamics.

Traditionally, economists have categorized discrimination as either taste-based discrimination (Becker, 1957) or statistical discrimination (Phelps, 1972; Arrow, 1973; Aigner and Cain, 1977). Taste-based discrimination arises if an individual experiences an animus towards members of a particular group and therefore discriminates against them to not experience a disutility from interacting with them. Resulting discrimination is hence a direct consequence of the discriminator's taste and attempts to alleviate this form of discrimination include normative approaches such as blaming or sanctioning (Rathelot and Safi, 2022) or increasing social interactions between groups (Pettigrew and Tropp, 2006). In contrast, statistical discrimination is based on beliefs. When there is uncertainty about focal attributes, evaluators need to rely on their beliefs about group-level statistics to make inference about those attributes. If these beliefs vary systematically across groups, discrimination may arise. Typically, statistical discrimination considers the case where beliefs are accurate and hence beliefs about differences reflect actual differences. However, beliefs that generate discrimination may be inaccurate and resulting discrimination is described as inaccurate statistical discrimination. Correcting these inaccurate group-level beliefs by providing individuals with the true statistics is one potential intervention to reduce this form of discrimination (Bohren et al., 2023).

In this paper, I provide a potential link between these two long-standing sources of discrimination. I show that individuals hold systematically inaccurate beliefs about the productivity of different social groups, that the formation of these incorrect beliefs is driven by directional tendencies, and that discrimination can be a consequence of these biased beliefs rather than a direct consequence of underlying tastes.[1] I study

[1] I use the term "directional tendency" to refer to any pre-existing inclination to favor one group

the behavioral operationalization of the formation of directional beliefs and demonstrate that individuals may exploit ambiguity in information (e.g., if the information is about group-level statistics rather than individual performance, or if the veracity of the information source is unclear) to maintain their prior beliefs. Finally, I show that discrimination based on directional belief formation can still effectively be mitigated by carefully designed information interventions that limit individuals' wiggle room to systematically interpret information and uphold their directional beliefs. This contrasts with taste-based discrimination, which is not impacted by information. Consequently, this paper provides a micro-foundation for the formation and persistence of incorrect beliefs that can drive a particular form of inaccurate statistical discrimination – *discrimination based on directional belief formation.*

In a series of four stylized online hiring experiments, 'employers' repeatedly engage in incentivized binary hiring decisions between two 'workers' from different races. Throughout all experiments, employers were randomized into one of two treatments: in treatment 'Race', the races of the workers were visible, whereas in treatment 'Neutral', this information was withheld. This treatment variation activates directional tendencies to hold particular productivity beliefs for employers in treatment 'Race', but not for those in the treatment 'Neutral'. Treatment 'Neutral' therefore served as a control condition, allowing to benchmark hiring decisions in the absence of racial information. Looking at decisions between Hispanic and Asian workers, I find in experiment 1 that employers significantly discriminate against Hispanics after having been provided with actual group-level performance statistics that indicate equal productivity distributions between the two groups.

To identify the underlying mechanism that is driving this disparity, I explore the information acquisition behavior of employers. In experiments 2 and 3, participants are additionally provided with an initial piece of individual-level information about the two workers and have the opportunity to iteratively request more ambiguous (experiment 2) and uncertain (experiment 3) individual-level information signals about the workers during the hiring decisions. Ambiguity and uncertainty are established by varying the veracity of the information source and by varying the information that participants receive about the information structure. In experiment 2, participants do not know the likelihood with which a given information signal is true. In experiment

over another, whether motivational or cognitive in origin. I formally sketch how such tendencies may affect individuals' decisions as well as their information search and processing in Section 4.

3, participants know that there is a 60% likelihood that a given signal is true (see Section 2.2.3 for details). Discrimination based on directional belief formation implies that employers 'fish for good news' Chen and Heese (2021), which means they try to acquire information that supports their directional tendency, e.g., by seeking additional information when a previous piece of information contradicts their directional tendency and stopping to seek additional information once their information set is consistent with their directional tendency. In line with this, I find that if the initial piece of provided information contradicts the employers' directional tendency, they are more likely to search for a second piece of information and acquire more pieces of information in total, as compared to when the initial piece of provided information confirms their directional tendency.

I also study to what extent employers act consistently with the information, and how this depends on whether or not the information is in line with their directional tendency. Further corroborating the mechanism of discrimination based on directional belief formation, I find that employers in both experiments are more likely to act consistently with the acquired information if the information confirms their directional tendency.

Finally, in experiment 4, I confirm that discrimination can be reduced by limiting the 'wiggle room' Dana et al. (2007) of employers to interpret information. After reducing the ambiguity and uncertainty of provided information signals, I find decreasing levels of discrimination and thereby provide evidence of how discrimination based on directional belief formation can be avoided.

Discrimination based on directional belief formation is closely related to, but distinct from, taste-based discrimination. In particular the behavior in experiment 4 contrasts with standard forms of taste-based discrimination as the provision of individual-level information should not affect discrimination that is solely based on taste, but does affect discrimination based on directional belief formation. Similarly, (inaccurate) statistical discrimination without directional belief formation cannot fully explain the findings across all experiments, in particular not those in experiments 2 and 3, as absent directional tendencies, individuals should not selectively engage with presented information depending on its content.

Taken together, this paper makes three main contributions. First, it proposes a new explanation of discrimination that links taste-based discrimination and inaccurate statistical discrimination: discrimination based on directional belief formation.

By selectively acquiring and processing information, individuals form directional beliefs and consequentially discriminate based on these beliefs. Through a series of experiments, it provides evidence of the existence and the underlying mechanisms of this explanation. By documenting biased information acquisition as well as belief distortions it provides evidence of two potential mechanisms through which directional belief formation operates. Second, it demonstrates that discrimination based on directional belief formation differs from traditional forms of discrimination by documenting how it can explain discrimination in settings where the basic model of statistical discrimination (without directional belief formation) as well as taste-based discrimination cannot fully explain observed behavior. Finally, it shows how these insights can be utilized to design an effective policy intervention to reduce discrimination. By varying the ambiguity of individual-level information, this paper demonstrates that limiting individuals' wiggle room to systematically engage with information may significantly reduce discrimination.

The remainder of this paper is structured as follows. Next, I review related literature on different models of discrimination with a particular focus on belief-based models, and discuss existing work on belief distortions and biased information choices in discrimination. Section 2 provides details on the experimental design and Section 3 demonstrates to what extent group-level beliefs of employers were aligned between treatments. Based on this, Section 4 explains the theoretical background of discrimination based on directional belief formation and derives testable predictions. Main results are provided in Section 5. Finally, Section 6 concludes.

1.1 Related Literature

This paper connects to the broad economic literature on theories of discrimination and empirical methods to measure it.[2] By linking taste-based motivations to statistical discrimination, it contributes to the recently growing literature, that aims to extend the traditional taxonomy of taste-based discrimination (Becker, 1957) versus statistical discrimination (Phelps, 1972; Arrow, 1973; Aigner and Cain, 1977). Other papers

[2]See, e.g., Charles and Guryan (2011) for a review on challenges to measuring racial discrimination, Bohren et al. (2018) for a discussion about observational vs. experimental data, Bertrand and Duflo (2017), Heckman (1998), and Neumark et al. (2016) for discussions of field experiments, Onuchic (2022) for a recent review on theories of discrimination, Cornell and Welch (1996) for a model of ingroup vs. outgroup screening discrimination, and Bohren et al. (2022) for tools to model and measure systemic discrimination.

that aim to reveal the limits of this long-standing taxonomy by looking more closely at inaccurate beliefs as drivers for discrimination include Coffman et al. (2021), who document that beliefs about average group differences drive gender discrimination in an artificial online hiring experiment, or Barron et al. (2022), who differentiate between explicit and implicit belief-based discrimination between genders.

Setting the stage for this paper, Bohren et al. (2023) argue for the importance of considering incorrect beliefs and document them. They demonstrate that when economic agents have inaccurate beliefs about group attributes, resulting discrimination based on these beliefs can be mistaken for taste-based discrimination. The authors argue that eliciting beliefs or providing multiple information treatments can identify inaccurate beliefs versus animus as potential drivers of discrimination and reason that those who discriminate based on taste are unlikely to change their behavior in response to group-level information. This paper builds upon their findings and takes a significant next step by exploring how directional belief formation – the upholding of directional beliefs through biased information search and processing – can prevent inaccurate beliefs from being corrected by simply providing such information. It considers individuals' learning behavior and shows that individuals selectively acquire and process information in line with their directional tendencies. It demonstrates that after receiving credible group-level and/or ambiguous individual-level information, individuals use the 'wiggle room' (Dana et al., 2007) that this information provides for the formation of persistent inaccurate beliefs about unobserved individual-level characteristics. By showing that directional belief formation leads to inaccurate statistical discrimination, this paper provides a micro-foundation for the emergence of inaccurate beliefs in a discrimination setting.[3]

Besides this study, to date only a few other studies have argued for the importance of considering incorrect beliefs as a source of discrimination in the economics literature.[4] However, these studies have either not identified the mechanism for why

[3]Other potential mechanisms for non-Bayesian updating or model misspecification resulting in persistent inaccurate beliefs include but are not limited to memory or attention constraints (e.g. Bordalo et al., 2023), correlation neglect (e.g. Enke and Zimmermann, 2019), or failure of contingent reasoning (e.g. Martínez-Marquina et al., 2019). See Bohren and Hauser (2021) for a general framework and more examples of how misinterpreting information impacts learning.

[4]In an in-depth literature review Bohren et al. (2023) show that only 10.5% of 105 papers on discrimination that were published in 10 top economics journals between 1990 and 2018 differentiate between accurate and inaccurate beliefs. These and more recent studies include e.g. Fershtman and Gneezy (2001), Albrecht et al. (2013), Reuben et al. (2014), Bordalo et al. (2016), Bordalo et al. (2019), Bohren et al. (2019), Bursztyn et al. (2020), and Esponda et al. (2022).

these belief errors arise or how they can persist in the face of informative signals. As such Bohren et al. (2019) and Sarsons (2017) provide evidence for biased beliefs but do not identify their source, and Coffman et al. (2023a), Bordalo et al. (2016), and Bordalo et al. (2019) argue that incorrect beliefs may be driven by stereotypes but do not directly tie this to discriminatory actions nor does their mechanism explain the persistence of incorrect beliefs over time and in the context of endogenous information acquisition and selective processing.

By introducing an endogenous information acquisition setting in a discrimination setting this paper also adds to the literature on belief distortions and biased information choices through systematic information acquisition and processing in discrimination contexts. Bartoš et al. (2016) pushed the research frontier in this context by modeling and documenting in three field experiments that rational (in)attention can amplify discrimination. They demonstrate that employers pay less attention to a priori less attractive applicants in cherry-picking markets, but more attention to a priori less attractive applicants in lemon-dropping markets. They reason that attention allocation is determined by the likelihood that costly information would change the status quo decision which is to not hire an applicant in the cherry picking market and to hire an applicant in the lemon dropping market.

Focusing more on the extent to which provided information is processed in order to update beliefs in a hiring context, Mengel and Campos-Mercade (2021) attribute disparities in artificial hiring to signal neglect in the belief formation process. They show that employers conservatively update beliefs when confronted with new information and ultimately discriminate against disadvantaged workers.

In closer and very recent work, Coffman et al. (2023b) experimentally study how endogenous information acquisition amplifies discriminatory outcomes between genders in a simulated hiring experiment. While they do not find systematic gender differences in hiring, they observe that relative comparisons – the extent to which a candidate's gender outperforms the other within the candidate pool – bias decision-making and observe that relative (dis)advantages amplify through endogenous information search. Consistent with the findings of my study, these studies provide evidence that systematic attention allocation and systematic processing of information can impact discrimination. Building on this, I look closer at the mechanisms by studying how endogenous information search and processing can lead to persistent and inaccurate beliefs that generate racial discrimination.

In a non-discrimination context, Chen and Heese (2021) also study biased information choices and show that individuals are more (less) likely to continue acquiring information after they have received information that suggests that acting selfishly is harmful (harmless). Together with Bohren and Hauser (2023) who provide a more general framework for mechanisms of belief distortions and show in an application how self-image concerns can lead to inaccurate beliefs and consequential discrimination, they provide related theoretical foundations for the observed information acquisition and processing behavior observed in this study. Similarly, this paper builds on theoretical foundations and psychological underpinnings of motivated reasoning (Kunda, 1990; Epley and Gilovich, 2016; Kozakiewicz, 2021), which have been documented in various non-discrimination contexts, such as altruistic situations (Di Tella et al., 2015), in response to feedback (Zimmermann, 2020), or in light of uncertainty (Drobner, 2022). The behavioral mechanism studied in this paper – directional belief formation through biased information search and processing – is consistent with motivated reasoning as studied in this literature, while remaining agnostic about whether the underlying source is motivational or cognitive in origin. More practically, it builds on Thaler (2020), who provides a novel experimental design to identify directionally biased belief formation and shows that even uninformative messages can lead individuals' beliefs to become more polarized and less accurate. I apply a modified version of the proposed experimental design to identify directional belief formation in the context of discrimination. Relatedly, this paper closely connects to research on confirmation bias (Nickerson, 1998; Klayman, 1995; Rabin and Schrag, 1999) as a related manifestation of directionally biased information processing. While confirmation bias refers to the tendency to seek and interpret information in line with one's prior beliefs, my setting is slightly different. I first provide participants with information that aligns their group-level beliefs about the productivity of Asians and Hispanics. Despite this alignment, when faced with binary hiring choices, individuals still selectively attend to and interpret individual-level signals in ways that favor the Asian candidate. This implies that the bias cannot be explained purely as confirmation of prior beliefs, but rather points to a directional tendency that persists even after group-level beliefs have been corrected.

Two recent studies by Rackstraw (2022) and Stötzer and Zimmermann (2024) explore motivated reasoning as a driver of stereotypical beliefs. Both studies share an important commonality with mine, namely the implication that motivated reasoning

may fuel discrimination. At the same time, they pursue a rather different focus in how they study this phenomenon. For example, the note by Stötzer and Zimmermann (2024) has implications for discriminatory beliefs, but its main objective is to test whether individuals devalue out-group members in order to justify selfish behavior, rather than to analyze discrimination between members of different social groups. By contrast, my paper examines how individuals behave in potentially discriminatory choice environments. In a setting closer to mine, Rackstraw (2022) also suggest that motivated reasoning may bias beliefs about workers from different social groups in a labor market context. However, like Stötzer and Zimmermann (2024), they do not focus on the endogenous information search and processing behavior that is central to my study and through which I show a specific mechanism by which directional belief formation may fuel discrimination. Taken together, these differences imply that none of the existing studies speaks directly to how discrimination based on directional belief formation differs from more traditional forms of discrimination, i.e. statistical discrimination (without directional belief formation) or taste-based discrimination, or how insights from an endogenous information search might be leveraged to design effective policy interventions.

Studying cognitive, rather than motivational reasons for self-deception, Carlana (2019) demonstrates that teachers' implicit stereotypes about gender and math ability predict gender gaps in student achievement, providing compelling evidence that biased beliefs of evaluators translate into discriminatory outcomes in education. My paper complements this finding by studying the behavioral mechanism through which such biased beliefs persist: while Carlana documents the downstream consequences of implicit stereotypes, I study how individuals actively maintain and reinforce biased beliefs through selective information acquisition and processing, even after being provided with corrective group-level information.

Similarly, Miserocchi (2023) also explains discrimination despite identical underlying information, but through a different mechanism. Her paper emphasizes biased recall and memory constraints, whereby decision-makers selectively remember stereotype-consistent information and form distorted beliefs. In contrast, my paper focuses on directional belief formation, where discrimination arises from how individual-level information is selectively acquired and interpreted without memory limitations.

Cunningham and De Quidt (2022) study how implicit attitudes predict discrimi-

natory behavior and examine the conditions under which implicit biases translate into action. While they focus on identifying implicit preferences from intransitive choice patterns driven by context-dependent "dilution", my paper instead emphasizes directional belief formation, where discrimination arises from the selective interpretation or weighting of information rather than from multiple layers of preferences. Thus, the two approaches are complementary: their framework provides a preference-based account of implicit discrimination, whereas mine highlights belief-based discrimination. Additionally, the endogenous information environment in my experiments is critical as it reveals how directional beliefs are actively maintained.

In another complementary work, Esponda et al. (2023) show that individuals assess probabilities based on representativeness rather than standard Bayesian updating. In a labor market context, this may generate directional tendencies in how employers evaluate workers from different groups. Rather than studying the source of directional tendencies, I study the behavioral mechanism through which they translate into persistent discrimination – selective information acquisition and biased processing that responds to the amount of wiggle room in the information environment.

2 Experiment Design

The data collection involves one survey and a series of pre-registered online experiments, programmed in Otree (Chen et al., 2016) and implemented on Prolific, that has shown to provide superior data quality compared to various other platforms (Peer et al., 2022). The following subsections explain the experimental design. For more details and screenshots, including instructions, see Appendix Section A.

2.1 The pool of workers

In the survey, I collect answers from 96 US participants on a logic quiz, a dictator game, and a real effort task. This is meant to mimic an 'assessment center' that provides proxies for workers' cognitive ability, social competence, and perseverance. Based on their answers in all three tasks, I calculate a composite score that defines their 'productivity' for the subsequent experiments. Specifically, this score consists of the number of questions solved in the logic task plus the amount of dollars shared in the dictator game plus the number of tasks solved in the real effort task and

was meant to provide a simple, behavior-based proxy for overall worker productivity across cognitive, social, and effort-related dimensions. Additionally, I ask for race and other demographics, past school performance information, and psychological scales for resilience, cooperativeness, ambition, and diligence. This information is used to set up profiles of available experimental 'workers' for hire in the main experiments. To rule out statistical discrimination based on accurate beliefs as the cause for potential discrimination in the hiring experiment by design, I set up the final pool of workers for the hiring stage by selecting workers so that productivity distributions between race groups are equal. The final pool of workers that was used in the hiring experiments consists of 58 individuals from the US, equally balanced across gender, aged between 18 and 30 with a mean age of 22.83 years.[5]

2.2 The hiring experiments

The pre-registration and experimental design of the hiring experiments replicate an earlier study that I conducted in December 2021. Using the same design, I collected new data from $1,605$ participants in September 2025.[6] After applying the pre-registered exclusion criteria (pass at least 60% of the comprehension questions, attention check, AI check), the final sample consists of $1,395$ participants, evenly split across the following four experiments: 346 participants in the 'No Information' experiment, 354 participants in the 'Ambiguous Information' experiment 342 participants in the 'Uncertain Information' experiment, and 353 participants in the 'Tangible Information' experiment.[7] Details of the experiments are provided in Section 2.2.3. On average, the experiments took between 15 minutes ('No Information') to 24 minutes ('Tangible Information').

[5]Of the 58 workers in the final pool, 17 identified as 'White', 16 as 'African American or Black', 13 as 'Asian', and as 12 'Hispanic or Latin' (hereinafter referred to as 'Hispanics').

[6]Results are robust across both waves of data collection; evidence from the earlier wave as well as pooled data is reported in Appendix Section B.

[7]The experiments are pre-registered under Eyting (2025). Representativeness of the sample is ensured in terms of age and gender and is established by Prolific (www.prolific.com). Out of the 210 excluded participants, 134 were dropped because they answered less then 60% of the test questions correctly, and 76 were dropped because they either failed an attention or AI check.

2.2.1 Inducing directional tendencies

All experiments involve the same two treatments in a between-subjects design. In both treatments, the employers make hiring decisions between two workers from different races. The two treatments differ in the way the workers are labeled. In treatment 'Race', the employers could observe the races of the workers, whereas in treatment 'Neutral' the race labels are replaced by neutral shape labels, e.g. 'Triangle' workers, or 'Diamond' workers. Importantly, the composition of workers in each hiring decision remains constant between treatments; only the labels are changed. This treatment variation activates directional tendencies to hold particular productivity beliefs for employers of treatment Race, but not for those of treatment Neutral. Hence, results from employers in treatment Neutral serve as baseline levels in the subsequent analysis.

2.2.2 Experiment procedure

The experiments consist of a belief stage followed by a hiring stage. In the belief stage, the employers are asked for their subjective belief distribution about the group productivities of the workers.[8] These beliefs indicate the direction of their tendencies. Subsequently, they are provided with the true productivity distributions for each group of workers to align beliefs between treatments. They are then again asked for their beliefs to check whether this belief alignment was successful. All employers were informed that the pool of workers consisted of selected workers from the survey.

As in Coffman et al. (2021), the data on beliefs reveals two things. First, it shows to what extent prior and posterior beliefs reflect the true productivity distributions. Second, and more importantly, it demonstrates the differences in beliefs between the two treatments. Updating beliefs between employers in treatment Race and employers in treatment Neutral is intended to generate identical and correct ex-post beliefs about the productivity distributions among employers in both treatments. This renders the potential directional tendencies based on the workers' race the only difference between

[8]Consistent with the argument by Armantier and Treich (2013) that stakes in the event as well as hedging opportunities through related decisions may bias belief reports, the belief elicitation was not incentivized. The presence of employers' directional tendencies (a non-monetary stake in their beliefs) as well as their subsequent incentivized hiring decisions (a potential hedging opportunity) could have interacted in ways that distort belief reporting, which discouraged the use of monetary incentives in this context.

the two treatments prior to the hiring decisions.[9]

In the hiring stage, I ask the employers to make a sequence of 20 incentivized hiring decisions between two workers from the constructed pool of workers.[10] Each employer is repeatedly presented with a pair of workers and asked to hire one of them. After the experiment, one hiring decision is randomly chosen. If they hired the worker with the higher productivity score in this decision the employer gets a bonus payment of \$2. The measurement of productivity has previously been explained to all employers. The workers did not receive an additional payment for being hired. After all hiring decisions the employers answer a few demographic questions before the experiment ends.[11]

2.2.3 Varying wiggle room

Discrimination based on directional belief formation implies that less wiggle room for the employers to form their beliefs about individual workers can decrease discrimination. Therefore, the experiments differ in the way *individual-level* information is provided in each hiring decision and thus in the extent of wiggle room that the employers have.

In Experiment 1 ('No Information'), employers have the most wiggle room as they do not receive any individual-level productivity information after having received the accurate group-level information during the belief stage. This means, that in each hiring decision in treatment Race employers are simply asked to hire e.g. the Asian worker or the Hispanic worker without receiving any further information about these particular two workers. In each hiring decision in treatment Neutral they are accordingly asked to hire e.g. the Triangle worker or the Diamond worker without receiving any further information about the two workers.[12] In total 346 subjects successfully completed this experiment, 186 subjects in treatment Race and 160 subjects in treatment Neutral.

In Experiment 2 ('Ambiguous Information'), employers also receive the accurate

[9] Until here, the basic structure of the experiments is partly inspired by the design of Coffman et al. (2021) who study gender discrimination in binary hiring decisions, after aligning beliefs between employers in a gender treatment and a birth-month treatment.

[10] Seven relevant decisions between Asian and Hispanic workers, and thirteen obfuscating decisions.

[11] The project received IRB approval at the joint ethics board of Goethe University Frankfurt and JGU Mainz.

[12] Importantly, the Triangle (Diamond) worker in treatment Neutral is the same worker as the Asian (Hispanic) worker in treatment Race. I vary race-shape assignments between sessions.

group-level information during the belief stage and complete the same hiring task as the employers in the 'No Information' experiment. However, in each hiring decision employers in this experiment receive one initial piece of individual-level information and are able to request up to 9 additional signals. The signals consist of a binary message from one of two randomly-chosen information sources: True News or Fake News. The message from the True News source is always correct, the message from the Fake News source is never correct. The message reads "The better worker is X" where X is one of the two presented workers.[13]

Note, that if the message comes from the Fake News source it implies that the worker who is not mentioned is indeed the better worker. Importantly, the employers do not know whether or not a message comes from the True news source or from the Fake News source. In fact, in this experiment, the employers are not informed about the likelihood with which a message comes from the True News source or the Fake News source.[14] Since the messages in this experiment are theoretically uninformative, they provide the employer with wiggle room to subjectively interpret each message in line with their directional tendency.[15] Apart from this information structure during the hiring phase, this experiment is identical to the 'No Information' experiment. In total 354 subjects successfully completed this experiment, 185 subjects in treatment Race and 169 subjects in treatment Neutral.

In Experiment 3 ('Uncertain Information') employers are provided with the exact same information structure as in the 'Ambiguous Information' experiment. However, in this experiment employers have previously been told that each message has a 60% likelihood to come from the True News source and a 40% likelihood to come from the Fake News source. This adds transparency about the information structure. Following (Dana et al., 2007), who argue that transparency about actions may reduce wiggle room, this additional layer of transparency about the information is meant to reduce employers' wiggle room to subjectively interpret each message compared to

[13]This feature is inspired by a novel experimental design to identify directionally biased belief formation by Thaler (2020). I apply a modified version to identify directional belief formation in the context of discrimination.

[14]Subsequent elicitation of the perceived fraction of messages from the True News source reveals that the vast majority of employers guess that approximately 50% of all messages are true.

[15]One could also argue that this setting provides more wiggle room than the 'No Information' experiment, but I expected some individuals to unconditionally overreact to any signal they get and therefore considered any individual level signal provision as a wiggle room limiting factor. Results confirm that a fraction of individuals unconditionally follow these uninformative first signals.

the 'Ambiguous Information' and 'No Information' experiments. Other than that, this experiment is identical to the 'Ambiguous Information' experiment. In total 342 subjects successfully completed this experiment, 178 subjects in treatment Race and 164 subjects in treatment Neutral.

In Experiment 4 ('Tangible Information') employers do not receive messages that directly displayed the (supposedly) better worker. Instead, employers are given individual-level information about past performances of the two workers.[16] Again, for each decision, employers receive one initial piece of information (e.g. their college GPA) and can request up to nine additional random pieces in each hiring decision. Thereby, for each request, the same attribute was shown for both workers in a pair. The order of attributes varied randomly between decisions. This environment still leaves wiggle room for participants but reduces it further as the signals are now always true, but still not conclusively predictive of the better of the two workers. In total 353 subjects successfully completed this experiment, 188 subjects in treatment Race and 165 subjects in treatment Neutral.

Note that, since I used the actual self-reported first names of real U.S. individuals who participated in the survey, and given the evidence from Gaddis (2023) that names can also signal social class, it is conceivable that employers in my setting likewise inferred social class from workers' names, and that such inferences systematically differed between Asian and Hispanic names. If so, potential findings would not reflect disparate treatments based purely on race or ethnicity, but rather on a combination of race and perceived social class.

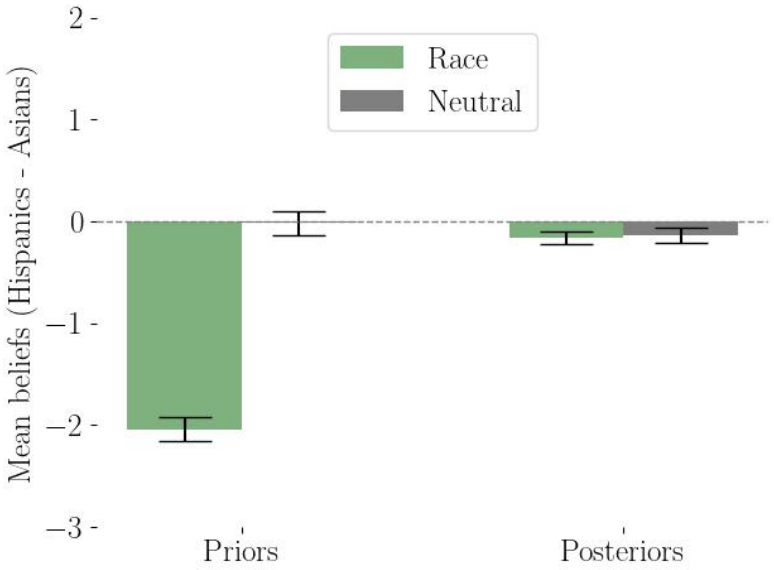
3 Beliefs about Group Productivities

Before describing the predictions and hypotheses in more detail I first present results of the belief stage that identify employers' directional tendencies. I then demonstrate that employers in treatment Race and treatment Neutral hold identical group-level beliefs about the productivity of workers from different races, once they enter the hiring stage.

The first bar of Figure 1 shows a large difference in mean prior beliefs about

[16]The past performance information includes their college and high school GPA, SAT and ACT score, final high school math and English grade, and psychological measurements of their level of ambition, resilience, diligence, and agreeableness (based on Duckworth et al., 2007; Rammstedt and John, 2005; Sinclair and Wallston, 2004).

Figure 1: Mean beliefs about group productivities



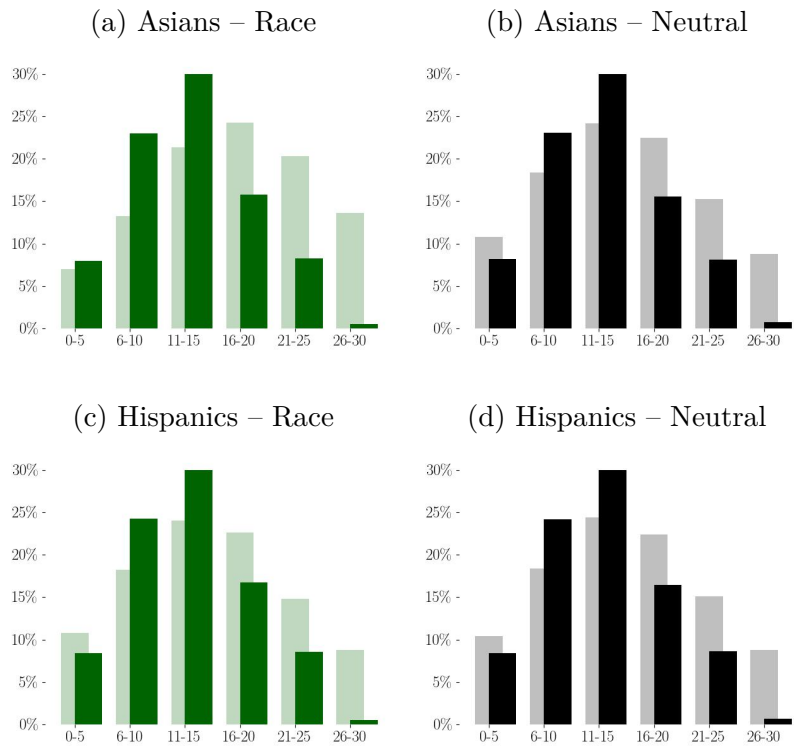
Notes: The first two bars show the mean productivity beliefs *before* the group-level information update of employers in treatment Race and treatment Neutral, respectively, the last two bars show the mean productivity beliefs *after* the group-level information update. Beliefs are plotted as the difference in mean beliefs about Hispanic workers and mean beliefs about Asian workers such that negative values indicate higher beliefs about the productivity of Asian workers than beliefs about the productivity of Hispanic workers, and vice versa. Error bars show the standard errors of the means.

respective productivity levels of Asians and Hispanics in treatment Race. Employers believed Asian workers to be significantly more productive (mean belief: 16.39) than Hispanic workers (mean belief: 14.35).[17] The second bar shows the differences in mean prior beliefs about Asian and Hispanic workers for employers in treatment Neutral and shows no significant difference (mean beliefs: 14.43 and 14.42, respectively). Considering that the two groups were presented to them as group Triangle and group Diamond, this result merely constitutes a verification. The difference in differences implies a potential for directional beliefs that Asian workers are more productive than Hispanic workers. Informed by a pilot study and to attenuate potential experimenter demand effects by diverting attention from the main decisions, I also elicited beliefs and hiring decisions about black and white workers. The analysis focuses on decisions between Asian and Hispanic workers for two reasons. First, as shown in Figure C1 in Appendix Section C, the largest gap in prior beliefs lies between Asian and

[17]This is consistent with previous literature on competence stereotypes (Fiske, 2018) and evaluation biases towards Asian students and against Hispanic students (Shi and Zhu, 2023).

Hispanic workers, providing the most promising context to test the hypotheses: it is the pair where the directional tendency is most clearly identified in the data. Second, comparing two minority groups reduces social desirability concerns that might lead employers to suppress discriminatory behavior, a risk particularly acute for White-vs-Black comparisons given heightened public attention to anti-Black discrimination. The theoretical framework predicts that discrimination based on directional belief formation may occur for any group comparison where employers hold pre-existing directional tendencies.

Figure 2: Prior and Posterior belief distributions



Notes: Figures (a) and (b) show employer belief distributions about the productivities of Asian workers before (low opacity) and after (high opacity) the group-level information update in treatment Race and treatment Neutral, respectively. Figures (c) and (d) show employer belief distributions about the productivities of Hispanic workers before (low opacity) and after (high opacity) the group-level information update in treatment Race and treatment Neutral, respectively.

Next, I check whether the group-level information intervention was successful in aligning beliefs. The two bars on the right of Figure 1 show that the treatment difference in differences in mean beliefs about group-level productivities between Asians and Hispanics disappeared after the information update. Importantly, this implies,

that group-level beliefs about the productivity of Asians and Hispanics were equal across treatments when employers entered the hiring stage. In fact, Figure 2 demonstrates that employers in both treatments learn that the average, variance, and skewness of the productivity distributions of Asian and Hispanic workers are equal. This result holds for all four experiments.

Figure C2 in Appendix Section C shows that beliefs were updated homogeneously across employers in each experiment. One could imagine that experimenter demand or confusion could still have caused some employers to report similar belief distributions across races, despite not actually holding these beliefs. It might hence be, that they have not truly absorbed this information and updated their prior beliefs. While asymmetric group-level beliefs and subsequent Bayesian updating could in principle explain part of the results below, it could still not explain employers' information search and processing behavior in the 'Ambiguous Information' treatment as shown in Section 5.2 and 5.3. Any systematic treatment differences in how this information was searched for or processed would still contradict Bayesian predictions, since the signals should be entirely uninformative to a rational Bayesian agent absent a specific motive. Additionally, employers knew that the pool of workers was artificially constructed from a larger pool of Asian and Hispanic workers. This implies that there was no reason not to believe the provided group-level information. In the next section, I present testable predictions for discrimination based on directional belief formation.

One might worry that employers simply failed to absorb the group-level information and reverted to their pre-existing stereotypes upon entering the hiring stage. If so, the observed patterns could reflect standard Bayesian updating from biased priors rather than directional belief formation. While I cannot fully rule out this possibility, several features of the data are difficult to reconcile with this explanation. First, posterior belief distributions were verified to be equal across treatments before the hiring stage (Figure 2), suggesting that employers did absorb the information at the time of elicitation. Second, employers knew that the worker pool had been artificially constructed, so they had no reason to doubt the accuracy of the provided distributions. Third, and most importantly, in the Ambiguous Information experiment, the signals are objectively uninformative: each message has an equal probability of coming from a True News or Fake News source, and employers were not told the likelihood. A rational Bayesian agent, even one with biased priors about group-level productivity, should not systematically engage with uninformative signals in a directional

manner. The fact that employers in treatment Race selectively sought additional, uninformative signals and differentially processed them depending on their content is inconsistent with standard Bayesian updating.

4 A brief framework

This section provides a brief conceptual framework of discrimination based on directional belief formation and describes the predictions and hypotheses that follow from it. The predictions depend on employers' wiggle room to systematically acquire and process information.

Consider an employer with a belief $b \in [0, 1]$ that the Asian worker is more productive than the Hispanic worker. The employer must decide whether to hire the Asian (A) or the Hispanc (H) worker, $a \in \{A, H\}$. Moreover, assume that the employer holds the following utility function:[18]

$$U(a,b) = \begin{cases} U(a = A, b \geq \frac{1}{2}) = r, \\ U(a = H, b \geq \frac{1}{2}) = -c, \\ U(a = A, b < \frac{1}{2}) = r - c, \\ U(a = H, b < \frac{1}{2}) = 0, \end{cases}$$

which can be summarized as:

$$U(a,b) = \begin{cases} r - \mathbf{1}_{\{b<\frac{1}{2}\}} \cdot c & \text{if } a = A \\ 0 - \mathbf{1}_{\{b\geq\frac{1}{2}\}} \cdot c & \text{if } a = H \end{cases}$$

where c is the (cognitive) cost of hiring the worker who is believed to be less productive and r reflects a positive payoff from hiring the worker favored by the directional tendency (in this case, the Asian worker). Note, that $c > r > 0$.

This yields the following hiring rule: hire the Asian worker if $b \geq \frac{1}{2}$ (as $r > -c$) and hire the Hispanic worker if $b < \frac{1}{2}$ (as $0 > r - c$). It follows, that the employer has an incentive to hold a belief $b \geq \frac{1}{2}$ (as $r > 0$).

[18] This utility function omits the monetary payoff that workers earn for hiring the more productive of the two workers. The reason is, that the monetary incentive encourages accurate hiring in both treatments equally. Since it does not vary across treatments, it drops out of the treatment comparison.

Even though group-level beliefs about the productivities of Asian and Hispanic workers are aligned before employers enter the hiring stage, beliefs about individual workers may still be distorted through directional belief formation if wiggle room to do so is provided. In the 'No Information' experiment, employers have considerable wiggle room to form any belief b when confronted with a hiring decision, e.g. by drawing selectively from two similar productivity distributions: they may infer that an Asian worker was drawn from the upper end of the distribution, while a Hispanic worker was drawn from the lower end. This allows employers to inflate their belief in a more productive Asian worker and a less productive Hispanic worker.

In the remaining experiments, two opposing forces operate. On the one hand, more informative individual-level signals should reduce discrimination if employers follow the information unconditionally. On the other hand, directional tendencies may distort how these signals are interpreted or searched for. Which force dominates depends on the wiggle room that the individual-level information provide. In the 'Ambiguous Information' experiment, employers retain substantial wiggle room to misperceive the informativeness of signals in ways consistent with their directional tendencies (see Sections 4.1 for the mechanism). Consequently, I expect significant discrimination to persist in this condition. While employers in both the 'No Information' experiment and the 'Ambiguous Information' experiment know that the two groups are equally productive on average, they are likely to reason that, between any particular Asian and Hispanic worker, the Asian is still more productive. As information becomes more precise, employers' wiggle room decreases. Therefore, I expect less discrimination in the 'Uncertain Information' and the 'Tangible Information' experiments. This reasoning yields the following two hypotheses:

Hypothesis 1a *There is significant discrimination against Hispanics in treatment Race, but not in treatment Neutral, among employers in experiments 'No Information' and 'Ambiguous Information'.*

Hypothesis 1b *There is less discrimination against Hispanics in decisions of employers in the 'Uncertain Information' and the 'Tangible Information' experiments, than in decisions of employers in the 'Ambiguous Information' or the 'No Information' experiments.*

Note, that since the signal distributions are identical across treatment groups

'Race' and 'Neutral', any observed disparities cannot be attributed to differences in information.

4.1 Mechanisms: how is the wiggle room used?

What are the underlying mechanisms of discrimination based on directional belief formation? In this section, I discuss two potential mechanisms, directional information processing and directional information search, and outline how they distinguish discrimination based on directional belief formation from both standard statistical discrimination and taste-based discrimination.

Mechanism 1: directional information processing We now turn to the formation of the belief b through asymmetric signal processing. Because hiring A with a belief $b \geq \frac{1}{2}$ yields a utility of r and hiring H with a belief $b < \frac{1}{2}$ yields a utility of 0, employers have an incentive to maintain beliefs that lead them to hire A. Formally, after receiving a signal s, the employer updates their belief b, according to

$$b = \hat{b} + \alpha(\tau)$$

where $\hat{b}$ denotes the Bayesian posterior and $\alpha(\tau)$ captures the directional adjustment depending on the signal precision τ, with $\alpha(\tau) > 0$ and $\alpha'(\tau) < 0$. This implies that employers upwardly bias their belief that the Asian worker is more productive than the Hispanic worker, with the magnitude of this bias depending on the signal precision τ. The clearer the signal, the smaller the bias.

Because signals favoring A are overweighted and signals favoring H are underweighted, employers are more likely to hire A after a majority of pro-A signals than to hire H after equally strong pro-H evidence.

Hypothesis 2a *If the majority of all considered signals in a decision suggests to hire the Hispanic worker in the 'Ambiguous Information' experiment, employers in treatment Race are less likely to act according to the suggestion than employers in treatment Neutral.*

Given that the individual-level information in the 'Uncertain Information' experiment and in the 'Tangible Information' experiment provides more meaningful

suggestions, i.e. τ increases and hence $\alpha(\tau)$ decreases, I hypothesize that directional belief-based discriminators reduce the systematically biased processing of information signals in those experiments.

Hypothesis 2b *If the majority of all considered signals in a decision suggests to hire the Hispanic worker, the difference in the likelihood to follow the suggestion between treatment Race and treatment Neutral is larger in the 'Ambiguous Information' experiment than in the 'Uncertain Information' and the 'Tangible Information' experiments.*

Mechanism 2: directional information search Finally, assume that the employer receives a first signal that suggest to hire the Asian worker $s = s_A$. Given that beliefs about the productivity of Asians and Hispanics have been aligned before the hiring decisions, i.e. $b = \frac{1}{2}$ the employer would hence increase their belief to $b > \frac{1}{2}$ and could hire the Asian worker to receive $U(a = A, b \geq \frac{1}{2}) = r$. However, after an initial signal that suggests to hire the Hispanic worker, the employer's belief would decrease to $b < \frac{1}{2}$ and the employer would only earn $U(a = H, b < \frac{1}{2}) = 0$ from hiring the Hispanic worker. In this case the employer has an incentive to request another signal to reinflate their belief to $b > \frac{1}{2}$. Hence, the likelihood to request a second signal as a means to inflate their belief is higher after an initial pro-H signal than after an initial pro-A signal.

Intuitively, this means, next to the biased information processing, discrimination based on directional belief formation predicts that individuals will 'fish for good news', meaning that they will request additional information signals if the previous signal contradicts their directional tendency, but will not request additional signals if the previous signal confirms their directional tendency. Since all individuals received one initial signal by default, I test for each hiring decision whether or not the content of the initial information signal affects (i) the likelihood to request a second information signal, and (ii) how many signals employers request per decision.

Hypothesis 3 *If the initial information signal suggests to hire the Hispanic worker in the 'Ambiguous Information' experiment, employers in treatment Race are more likely to request a 2nd signal and request more signals than employers in treatment Neutral.*

While I did not specify hypotheses about the behavior in the 'Uncertain Information' experiment and the 'Tangible Information' experiment, I do explore behavior in these experiments in the same way as I analyze the data of the 'Ambiguous Information' experiment.

Conceptually, the 'Uncertain Information' experiment and the 'Tangible Information' experiment are slightly different from each other. In the 'Uncertain Information' experiment only 60% of all signals are true and the information structure is rather abstract. In contrast, in the 'Tangible Information' experiment all signals are true and the information structure is very simple and includes specific, measurable indications of individuals' performance that can be directly understood and easily related to. This implies that in the former directional belief formation may operate through the evaluation of the truthfulness of a signal (as in the 'Ambiguous Information' experiment), whereas in the latter, it operates through the evaluation of the predictive power of a signal. In another exploratory analysis, I also study the extent to which directional belief formation driven by the evaluation of the truthfulness of constructed information in the 'Uncertain Information' experiment differs from directional belief formation driven by the evaluation of predictive power of the more relatable information in the 'Tangible Information' experiment.

By design, accurate statistical discrimination is ruled out as there are no group-level differences between Asians and Hispanics in my constructed pool of workers and hence potential beliefs about differences cannot reflect actual differences. Moreover, taste-based discrimination is unlikely, as it assumes that economic agents experience a disutility from interactions with certain individuals, but my experimental setting does not include any employer-worker interaction and employers are aware of this right from the start. Still, it is possible that individuals consider the mere selection of a worker from the less preferred group as a distasteful interaction with that worker. However, taste-based discriminators should not systematically process and search for information as their behavior is based on taste, not on beliefs. While it is possible that taste-based discriminators use the available information to justify their taste, the hypothesized different processing behavior between employers with wiggle room and employers with less wiggle room inconsistent with taste-based discrimination as the sole source of discrimination.

4.2 What is a directional tendency?

In general, for discrimination based on directional belief formation to occur, it is irrelevant what generates the directional tendency. A directional tendency is a pre-existing inclination to favor one group over another that may stem from preference-based sources (e.g., a distaste against workers from one particular group, or a taste for workers from another), or cognitive mechanisms (e.g., a stereotype that workers from one group are more productive, or the application of a representativeness heuristic whereby high productivity is more cognitively associated with one group; see Esponda et al. 2023). This paper does not take a stance on which of these underlies the observed behavior. Note that the model is silent on the source of r: a positive r for the Asian worker is consistent with a preference for Asian workers, a stereotype about group productivity, or a cognitive tendency to view Asian workers as more representative of high productivity. The behavioral predictions, directional information search and processing, are identical regardless of the source of r.

As discussed by Zimmermann (2020), several reasons for holding self-deceptive beliefs, which he calls the "demand side" of self-deception, have been proposed in the literature. Köszegi (2006) and Brunnermeier and Parker (2005), for example, emphasize belief-based utility, whereby individuals derive utility from holding favorable views (about oneself), while Festinger (1957) highlights the role of cognitive dissonance and the desire to avoid holding beliefs that conflict with prior views or actions. In the context of discrimination, self-deception may result from identity concerns (Bénabou and Tirole, 2011), for instance when beliefs about other groups help sustain a positive in-group identity, ideological or social alignment (Bursztyn and Yang, 2022), where individuals adopt beliefs consistent with prevailing narratives in their social environment, or outcome dependency (Kunda, 1990), when individuals are motivated to hold beliefs that justify preferred decisions or anticipated behavior.

Directionally biased belief formation need not, however, require a motivational component. As noted above, Esponda et al. (2023) show that a purely cognitive mechanism, the representativeness heuristic, can generate the same directional tendencies without any underlying desire to hold a particular belief. Similar to Zimmermann (2020), this paper does not take a stance on the demand side of self deception. What distinguishes this paper's contribution is the identification of the behavioral mechanism through which such tendencies translate into persistent discriminatory outcomes: directional information acquisition and processing that responds to the

amount of wiggle room in the information environment. Importantly, this source of discrimination differs from conventional inaccurate statistical discrimination, as it is characterized by endogenous information search and processing dependent on the direction of the tendency, even when group-level beliefs about the productivity of different groups are equal.

The implications of the analysis are robust to different underlying sources of the directional tendency. If individuals derive utility directly from holding particular directional beliefs, they will selectively search for and overweight information that confirms them. If driven by cognitive dissonance, they will process information in a way that avoids contradictory evidence. If directed by identity concerns, social alignment, or outcome dependency, individuals will similarly favor information consistent with their directional prior. If the tendency stems from a representativeness heuristic (Esponda et al., 2023), employers will systematically search for and process signals in a manner consistent with their cognitive prior about which group is more representative of high productivity. Across all these cases, the key behavioral implication is identical: individuals engage in directionally biased search and processing of information, leading to discriminatory hiring decisions. Importantly, this distortion arises only when individual-level information is sufficiently ambiguous to allow for selective interpretation; when information becomes precise and unambiguous, the scope for such distortions vanishes and the effect disappears. However, while improved information can mitigate these distortions by limiting the scope for selective interpretation, it primarily addresses the behavioral consequences rather than the underlying causes. Policies that directly target the roots of sustained directional beliefs, such as interventions reducing identity concerns or stereotype activation, have the potential to address the problem at its source. Future studies should therefore investigate the actual underlying sources that drive individuals' directional tendencies to inform effective targeted policies.

5 Results

I first consider to what extent employers discriminate across experiments. Subsequently, the analysis of the information processing and search behavior in the 'Ambiguous Information', 'Uncertain Information', and 'Tangible Information' experiments provides insights into the potential underlying mechanisms of directional belief

formation.

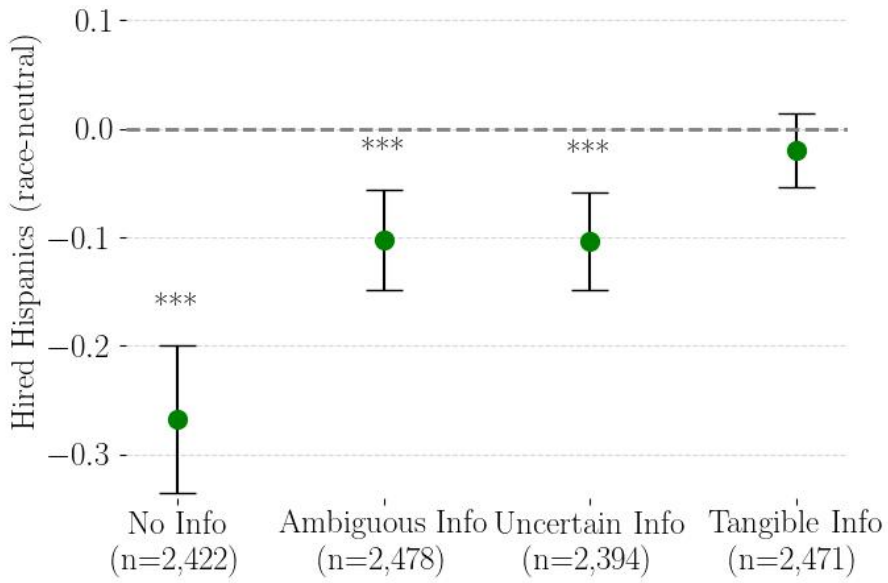
5.1 Disparities in hiring rates

In line with hypothesis 1a, Figure 3 shows that discrimination against Hispanics is substantial in the 'No Information' experiment and the 'Ambiguous Information' experiment. Providing employers with the race labels of workers in the 'No Information' experiment significantly decreases the hiring rate of Hispanics by 26.7pp. (p-value$<$ 0.001). In the 'Ambiguous Information' experiment, there is also a significant 10.2pp. (p-value$<$ 0.001) difference in hiring rates of Hispanics between the two treatments. Albeit this difference is lower than in the case with no additional information, both experiments show significant discrimination against Hispanics, even though employers are aware that group-level productivity scores between the two groups are equal.[19]

Usually, this source of discrimination could be described as taste-based discrimination despite the fact that employers and workers do not interact. In experiments 'Uncertain Information' and 'Tangible Information', I limit employers' wiggle room to systematically use the information to form directional beliefs by providing more meaningful individual-level information. If discrimination in the 'No Information' and the 'Ambiguous Information' experiments was based on taste and employers simply used the individual-level information to justify their taste, we should not observe any differences between behavior of employers in the 'Ambiguous Information' experiment and behavior of employers in the experiments with reduced wiggle room. This is because the hiring behavior of taste-based discriminators should not respond to variations in information. While one could argue that less wiggle room might complicate the ex-post rationalization of animus towards Hispanics, it does not affect the animus itself. The driver of discrimination would still be unaffected by changes in information and the hiring behavior should therefore not change with reduced wiggle room in experiments 'Uncertain Information' and 'Tangible Information' if discrimination was solely based on taste.

[19]The difference can be explained by the decision rule of some individuals to just follow the signals which lowered the discrimination rate in the 'Ambiguous Information' experiment compared to the one in the 'No Information'. However, discrimination rates were more extreme (45.7pp.) in the 'Ambiguous Information' experiment when subjects only requested one additional signal and ended up with two conflicting signals (see also Figure D1 in Appendix Section D.1) which supports the idea of directional belief formation.

Figure 3: Disparities in hiring rates across experiments



Notes: This figure shows the coefficients from an OLS regression of a dummy that is 1 if the Hispanic worker is hired and 0 otherwise, on the treatment dummy (Race = 1, Neutral = 0), controlling for employer fixed effects. *Negative* coefficients indicate discrimination against Hispanic workers. The horizontal axis groups results by experiment ('No Information', 'Ambiguous Information', 'Uncertain Information', and 'Tangible Information'). Standard errors are clustered at the individual level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

The coefficients for the 'Uncertain Information' experiment and the 'Tangible Information' experiment in Figure 3 show that the increased informativeness of the individual-level information translates into less discrimination. Absolute treatment differences in hiring rates of Hispanics in the experiments with less wiggle room for the employers remain at −10.3pp. (p-value$<$ 0.001) in the Uncertain Information experiment, but decrease substantially to an insignificant disparity of −0.02pp. (p-value= 0.254) in the Tangible Information experiment (see Table D1 in Appendix Section D.3). Overall, this supports hypothesis 1b and illustrates that tangible individual-level information that reduced individuals' wiggle room for interpretation can effectively decrease discrimination based on directional belief formation. As mentioned above, if discrimination was solely based on taste, we should not have observed differences in hiring rates of Hispanic vs. Asian workers in these two experiments compared to employers' behavior in the first two experiments.

Note that the reduction in discrimination between the 'Ambiguous Information'

and 'Uncertain Information' experiments is negligible (10.2pp. vs. 10.3pp.). This suggests that informing employers about the 60% likelihood of signals being true does not meaningfully reduce their wiggle room. Even with this information, signals remain probabilistic and employers can still selectively decide how much weight to place on any individual signal, exploiting the residual uncertainty to sustain their directional beliefs. The decisive reduction in discrimination occurs only when information becomes tangible and verifiably true, as in the 'Tangible Information' experiment. This pattern suggests that the critical factor for reducing discrimination based on directional belief formation is not merely transparency about the information structure, but the elimination of interpretive ambiguity in the signals themselves.

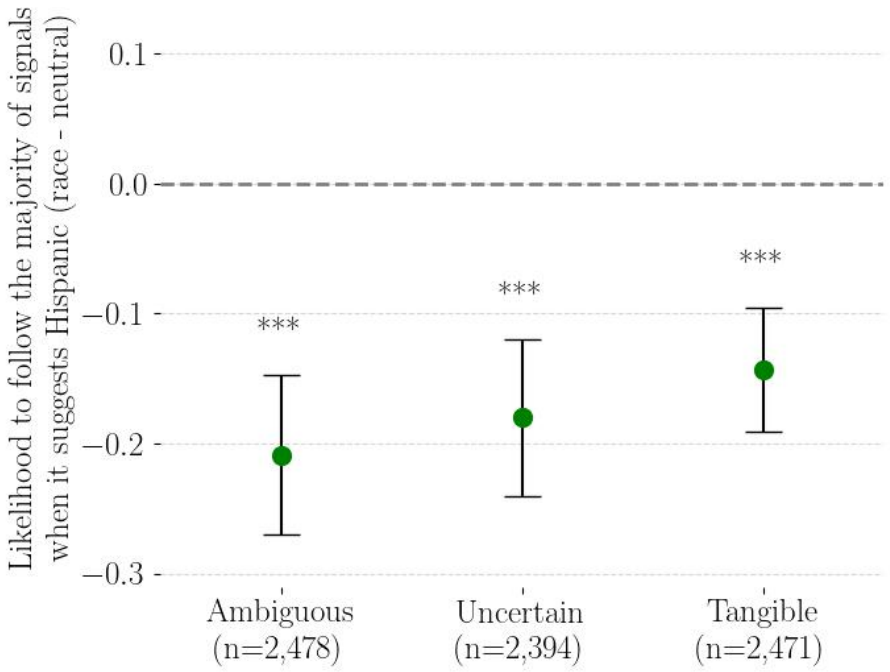
To further investigate the potential source of this discrimination, I now consider the information processing and search behavior of the employers.

5.2 Directional information processing

To gain some insights into the potential mechanisms, Figure 4 provides first evidence that the observed discrimination is based on directional belief formation. In the 'Ambiguous Information' experiment employers of treatment Race are significantly less likely to follow the signals if they suggest to hire the Hispanic worker than employers of treatment Neutral (coeff$=-20.9$, p-value<0.001). This lends support to hypothesis 2a.

Next, I consider the case where employers were given similar information as in the 'Ambiguous Information' experiment, except that employers are told that signals have a 60% likelihood to come from the True News source and a 40% likelihood to come from the Fake News source. The coefficient for the 'Uncertain Information' experiment of Figure 4 illustrate that reducing ambiguity and hence the potential wiggle room to interpret the information substantially decreases the treatment difference in the likelihood with which employers act in line with the acquired information. A majority of requested signals that suggests to hire the Hispanic worker, reduces the likelihood to follow the suggestion by 18.0pp more in treatment Race than in treatment Neutral. This corresponds to an insignificant reduction of 2.9pp (p-value$=0.513$) in the absolute effect size relative to the 'Ambiguous Information' experiment, consistent with the finding above that the 'Uncertain Information' experiment does not meaningfully

Figure 4: Systematically different information processing



Notes: This figure shows the coefficients from an OLS regression of a dummy that is 1 if the employer follows the majority of observed signals on the interaction between a treatment dummy (Race = 1, Neutral = 0) and a dummy equal to 1 if the majority of signals suggested hiring the Hispanic worker, controlling for employer fixed effects. *Negative* coefficients indicate that employers in the Race treatment were *less* likely to follow the majority of signals when they favored the Hispanic worker. The horizontal axis groups results by experiment ('Ambiguous Information', 'Uncertain Information', and 'Tangible Information'). Standard errors are clustered at the individual level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

reduce employers' wiggle room.[20]

Finally, I look at the behavior of employers that were provided with true information about past performances of the two workers in each decision. The estimated treatment difference in the likelihood to follow the signals in the 'Tangible Information' experiment decreases significantly by 6.5pp. (p-value= 0.099) relative to the 'Ambiguous Information' experiment to an absolute effect of −14.3 pp. Taken together, although there is still systematic processing of individual-level information under conditions with less wiggle room, this biased processing is significantly reduced. Taken together, the results indicate that the difference in the likelihood of following

[20]Figure B4 in Appendix Section B shows a significant decrease in the absolute effect size, from −26.0 to −9.6 (p-value of difference: < 0.001), when conducting the same analysis on wave 1 data. This is the only qualitative difference in the results between wave 1 and wave 2.

the suggestion between the Race and Neutral treatments is larger in the 'Ambiguous Information' experiment than in the 'Tangible Information' experiment, but not larger than in the 'Uncertain Information' experiment, providing only partial support for hypothesis 2b.

5.3 Directional information search

Finally, Figure 5 provides further pieces of evidence that discrimination is based on directional belief formation. Consistent with hypothesis 3 it shows that an initial signal that suggests to hire the Hispanic worker has a significantly larger positive effect in treatment Race than in the treatment Neutral on (1) the likelihood to acquire a second signal (coeff=3.6pp., p-value$<$ 0.050, Figure 5a) and insignificantly also on (2) the total number of signals requested per decision (coeff=0.16, p-value=0.135, Figure 5b). These results demonstrate that employers were 'fishing' for information signals that confirm their directional tendency if they have the necessary wiggle room to do so.

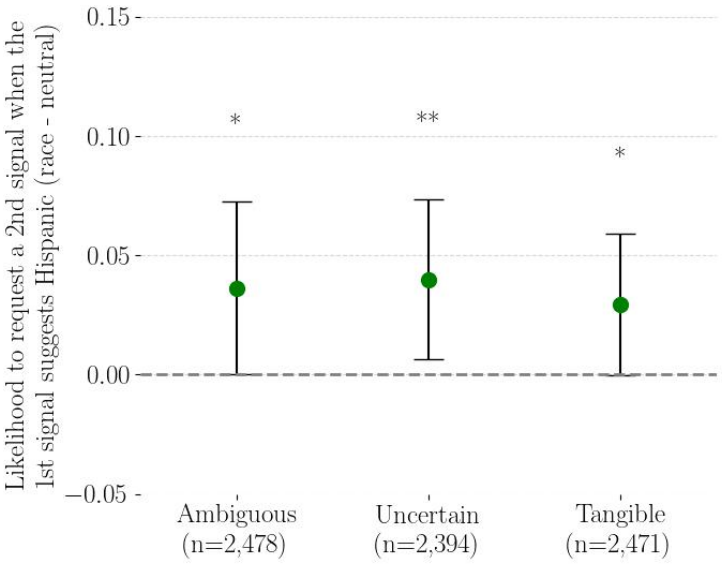
The information acquisition and processing behavior of employers in the 'Uncertain Information' and the 'Tangible Information' experiment is mixed. While an initial signal that suggests to hire the Hispanic worker has a significantly larger positive effect in treatment Race than in the treatment Neutral on the likelihood to acquire a second signal in both experiments, this only translates into a significantly higher number of requested signals in the setting with the least wiggle room, i.e. the 'Tangible Information' experiment.

In sum, the results confirm the existence and underlying mechanisms of discrimination based on directional belief formation. When employers had ample wiggle room to interpret ambiguous signals, they selectively sought and processed information in line with their directional beliefs, leading to substantial discriminatory behavior. As the informational environment became more structured and less open to interpretation, both the tendency to distort information and the resulting discrimination declined, with discrimination disappearing entirely when employers received fully tangible, individual-level information.[21]

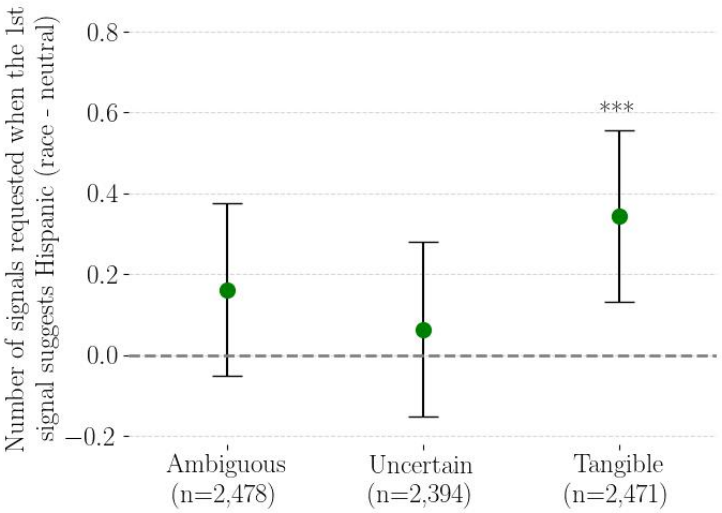
[21]Table D1 in Appendix Section D.3 reports the estimates underlying Figures 3–5. Figures D2 and D3 in Appendix Section D.2 shows results of the "counter-hypotheses" for individuals who stop requesting signals when the majority of signals suggests to hire the Asian worker and for individuals who receive a first signal that suggests to hire the Asian worker, respectively. Results are consistent

Figure 5: Systematically different information search

(a) Likelihood to request a 2nd signal



(b) Number of signals requested per decision



Notes: Panel (a) of this figure shows the coefficients from an OLS regression of a dummy that is 1 if the employer requests a second signal on the interaction between a treatment dummy (Race = 1, Neutral = 0) and a dummy equal to 1 if the first signal suggested hiring the Hispanic worker, controlling for employer fixed effects. *Positive* coefficients indicate that employers in the Race treatment were *more* likely to request a second signal when the first signal favored the Hispanic worker. Panel (b) shows the coefficients from an OLS regression of the number of requested signals per decision on the interaction between a treatment dummy (Race = 1, Neutral = 0) and a dummy equal to 1 if the first signal suggested hiring the Hispanic worker, controlling for employer fixed effects. *Positive* coefficients indicate that employers in the Race treatment requested *more* signals when the first signal favored the Hispanic worker. The horizontal axes group results by experiment ('Ambiguous Information', 'Uncertain Information', and 'Tangible Information'). Standard errors are clustered at the individual level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

As mentioned above, if discrimination was solely based on taste, we should not have observed differences in employers' behavior in the 'Uncertain Information' and the 'Tangible Information' experiments compared to employers' behavior in the 'No Information' and 'Ambiguous Information' experiments. Since we do observe significant differences in (i) the way employers engage with the information and (ii) hiring rates of Hispanic vs. Asian workers, the results cannot be explained by taste-based discrimination alone. Similarly, (inaccurate) statistical discrimination without directional belief formation cannot fully explain these findings. Absent directional tendencies, individuals should not systematically engage differently with presented information depending on its content, as prior beliefs had been aligned before the hiring stage and across treatments, leaving the race labels the only difference between treatments. While it is possible that in the Tangible Information experiment, (inaccurate) statistical discrimination without directional belief formation might still arise from different beliefs about the informativeness of the signal, this cannot explain observed individual behavior in the Ambiguous Information experiment and the Uncertain Information experiment. Hence, the only explanation that is consistent with behavior in all four experiments is discrimination based on directional belief formation.

5.4 Further evidence

I pre-registered and conducted two further experiments using the past performance information of the workers. These experiments were identical to the previous experiments, except for the provided individual-level information.

In the first additional experiment ('One Information'), employers are provided with individual-level information about past performances of the two workers (as in the Tangible Information experiment). However, for each decision, employers receive one piece of information and can not request any additional piece.

In the second additional experiment ('All Information'), employers are immediately provided with all ten available individual-level pieces of information about past performances of the two workers.

Both experiments reduce employers' wiggle room at least as much as the Tangible

and confirm directional information processing and search under wiggle room. Appendix Section D.4 presents robustness checks restricting the sample to participants who answered at least 80% (Table D2) and all (Table D3) comprehension questions correctly. The results remain unchanged.

Information experiment, as employers cannot engage in systematic information search and provided information is true and closely related to the respective workers. Hiring rates between treatments in these experiments are almost identical (see Figure E2 of Appendix Section E). This provides further evidence consistent with discrimination based on directional belief formation.

6 Conclusion

While the existence of discrimination in many contexts has been documented extensively, we still know much less about particular mechanisms of discrimination. This study contributes to filling this gap by taking a closer look at how individuals deal with individual-level information about others. I identify a particular source of inaccurate statistical discrimination that links documented forms of statistical discrimination to the well-known model of taste-based discrimination. When confronted with information that leaves a lot of wiggle room for interpretation, individuals make use of the inconclusiveness of the information and engage with the information systematically differently, depending on the informational content. They search for information that confirms their directional tendency and put less weight on information that contradicts it, and ultimately discriminate. When the wiggle room of individuals to interpret information decreases through more conclusive information structures, discrimination decreases. This implies that employers respond to information, but that this response is influenced by their directional tendency.

Distinguishing between various forms of discrimination has important implications for designing targeted policy interventions. This paper complements Bohren et al. (2023) who state that when group statistics are equal between two groups, remaining discrimination looks like taste-based discrimination but might actually be belief-based, as discriminators are not aware of the equal group statistics. I extend their argument by showing that even if discriminators are aware of equal group statistics, remaining discrimination still does not need to be taste-based. Discriminators might believe that particular individuals of one group outperform individuals of another group, and these beliefs may be upheld through directional information processing. In light of discrimination based on directional belief formation, it seems important that policy interventions take wiggle room of information into account to effectively fight discrimination. Since credible individual-level information is often difficult to

provide to decision-makers, the potential for discrimination is large in many contexts. As such, contexts in which individual-level information is sometimes not existent (e.g. about taxi or delivery drivers) or is mainly provided by the potentially discriminated (e.g. labor market) are prone to this source of discrimination. Decision-makers might reason that individual-level information from and about a particular individual is only true if it confirms their directional tendency but false if it contradicts it. It is therefore important that information interventions provide means to signal the validity of information, which reduces decision-makers' wiggle room to engage in directional belief formation.

Even though this study provides first evidence for the existence of discrimination based on directional belief formation the exact mechanisms of how information is processed such that decision-makers ultimately discriminate remain yet to be studied. Future research could employ process tracing techniques to look more closely at the information processing behavior (Chen and Fischbacher, 2016; Lahey and Oxley, 2016). Future studies could also study the interplay of individual-level and group-level information in discrimination contexts in more detail. It remains unclear to what extent different information acquisition and processing behavior can be linked to concepts like groupiness (Kranton and Sanders, 2017), or other individual characteristics and attitudes. Finally, studying discrimination based on directional belief formation in the field could yield first evidence of this source of discrimination in real-world settings.

References

Aigner, D. J. and Cain, G. G. (1977). "Statistical theories of discrimination in labor markets". *Ilr Review* 30 (2), pp. 175–187.

Albrecht, K., Von Essen, E., Parys, J., and Szech, N. (2013). "Updating, self-confidence, and discrimination". *European Economic Review* 60, pp. 144–169.

Alsan, M., Garrick, O., and Graziani, G. (2019). "Does diversity matter for health? Experimental evidence from Oakland". *American Economic Review* 109 (12), pp. 4071–4111.

Armantier, O. and Treich, N. (2013). "Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging". *European Economic Review* 62, pp. 17–40.

Arnold, D., Dobbie, W., and Hull, P. (2021). "Measuring racial discrimination in algorithms". *AEA Papers and Proceedings*. Vol. 111, pp. 49–54.

Arnold, D., Dobbie, W., and Yang, C. S. (2018). "Racial bias in bail decisions". *The Quarterly Journal of Economics* 133 (4), pp. 1885–1932.

Arrow, K. J. (1973). "The Theory of Discrimination". *Discrimination in Labor Markets*. Ed. by O. Ashenfelter and A. Rees. Princeton NJ: Princeton University Press.

Barron, K., Ditlmann, R., Gehrig, S., and Schweighofer-Kodritsch, S. (2022). "Explicit and implicit belief-based gender discrimination: A hiring experiment".

Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). "Attention discrimination: Theory and field experiments with monitoring information acquisition". *American Economic Review* 106 (6), pp. 1437–75.

Becker, G. S. (1957). *The economics of discrimination*. Oregon State monographs: Studies in economics, University Press.

Bénabou, R. and Tirole, J. (2011). "Identity, morals, and taboos: Beliefs as assets". *The quarterly journal of economics* 126 (2), pp. 805–855.

Bertrand, M. and Duflo, E. (2017). "Field experiments on discrimination". *Handbook of economic field experiments* 1, pp. 309–393.

Blau, F. D. and Kahn, L. M. (2017). "The gender wage gap: Extent, trends, and explanations". *Journal of economic literature* 55 (3), pp. 789–865.

Bohren, A., Imas, A., and Rosenberg, M. (2018). "The Language of Discrimination: Using Experimental versus Observational Data". *AEA Papers and Proceedings*. Vol. 108, pp. 169–74.

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.

Bohren, J. A. and Hauser, D. N. (2021). "Learning with heterogeneous misspecified models: Characterization and robustness". *Econometrica* 89 (6), pp. 3025–3077.

— (2023). *Behavioral Foundations of Model Misspecification*. Tech. rep. Penn Institute for Economic Research, Department of Economics, University of …

Bohren, J. A., Hull, P., and Imas, A. (2022). *Systemic Discrimination: Theory and Measurement*. Tech. rep. National Bureau of Economic Research.

Bohren, J. A., Imas, A., and Rosenberg, M. (2019). "The dynamics of discrimination: Theory and evidence". *American economic review* 109 (10), pp. 3395–3436.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). "Stereotypes". *The Quarterly Journal of Economics* 131 (4), pp. 1753–1794.

— (2019). "Beliefs about gender". *American Economic Review* 109 (3), pp. 739–73.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., and Shleifer, A. (2023). "Memory and probability". *The Quarterly Journal of Economics* 138 (1), pp. 265–311.

Brunnermeier, M. K. and Parker, J. A. (2005). "Optimal expectations". *American Economic Review* 95 (4), pp. 1092–1118.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). "Misperceived social norms: Women working outside the home in saudi arabia". *American economic review* 110 (10), pp. 2997–3029.

Bursztyn, L. and Yang, D. Y. (2022). *Misperceptions about others*. Tech. rep. forthcoming in Annual Review of Economics.

Carlana, M. (2019). "Implicit stereotypes: Evidence from teachers' gender bias". *The Quarterly Journal of Economics* 134 (3), pp. 1163–1224.

Charles, K. K. and Guryan, J. (2011). "Studying discrimination: Fundamental challenges and recent progress". *Annu. Rev. Econ.* 3 (1), pp. 479–511.

Chen, D. L., Schonger, M., and Wickens, C. (2016). "oTree – An open-source platform for laboratory, online, and field experiments". *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.

Chen, F. and Fischbacher, U. (2016). "Response time and click position: Cheap indicators of preferences". *Journal of the Economic Science Association* 2 (2), pp. 109–126.

Chen, S. and Heese, C. (2021). *Fishing for good news: Motivated information acquisition*. Tech. rep. University of Bonn and University of Mannheim, Germany.

Coffman, K., Collis, M. R., and Kulkarni, L. (2023a). "Stereotypes and belief updating". *Journal of the European Economic Association*, jvad063.

Coffman, K., Kostyshak, S., and Saygin, P. (2023b). *Choosing and Using Information in Evaluation Decisions*. Tech. rep.

Coffman, K. B., Exley, C. L., and Niederle, M. (2021). "The role of beliefs in driving gender discrimination". *Management Science* 67 (6), pp. 3551–3569.

Cornell, B. and Welch, I. (1996). "Culture, information, and screening discrimination". *Journal of political Economy* 104 (3), pp. 542–571.

Cunningham, T. and De Quidt, J. (2022). "Implicit preferences".

Dana, J., Weber, R. A., and Kuang, J. X. (2007). "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness". *Economic Theory* 33 (1), pp. 67–80.

Di Tella, R., Perez-Truglia, R., Babino, A., and Sigman, M. (2015). "Conveniently upset: avoiding altruism by distorting beliefs about others' altruism". *American Economic Review* 105 (11), pp. 3416–42.

Drobner, C. (2022). "Motivated beliefs and anticipation of uncertainty resolution". *American Economic Review: Insights* 4 (1), pp. 89–105.

Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). "Grit: perseverance and passion for long-term goals." *Journal of personality and social psychology* 92 (6), p. 1087.

Enke, B. and Zimmermann, F. (2019). "Correlation neglect in belief formation". *The Review of Economic Studies* 86 (1), pp. 313–332.

Epley, N. and Gilovich, T. (2016). "The mechanics of motivated reasoning". *Journal of Economic perspectives* 30 (3), pp. 133–40.

Esponda, I., Oprea, R., and Yuksel, S. (2022). "Discrimination Without Reason: Biases in Statistical Discrimination".

— (2023). "Seeing what is representative". *The Quarterly Journal of Economics* 138 (4), pp. 2607–2657.

Eyting, M. (2025). *Motivated Discrimination*. Tech. rep. Retrieved from https://osf.io/kewp2.

Farkas, G. (2003). "Racial disparities and discrimination in education: What do we know, how do we know it, and what do we need to know?" *Teachers College Record* 105 (6), pp. 1119–1146.

Fershtman, C. and Gneezy, U. (2001). "Discrimination in a segmented society: An experimental approach". *The Quarterly Journal of Economics* 116 (1), pp. 351–377.

Festinger, L. (1957). *A theory of cognitive dissonance*. Vol. 2. Stanford university press.

Fiske, S. T. (2018). "Stereotype content: Warmth and competence endure". *Current directions in psychological science* 27 (2), pp. 67–73.

Gaddis, S. M. (2023). "Signaling class: An examination of the treatment validity of names used to signal race in bias experiments with methodological recommendations for name selection". *Available at SSRN 3350739*.

Goncalves, F. and Mello, S. (2021). "A few bad apples? Racial bias in policing". *American Economic Review* 111 (5), pp. 1406–41.

Heckman, J. J. (1998). "Detecting discrimination". *Journal of economic perspectives* 12 (2), pp. 101–116.

Klayman, J. (1995). "Varieties of confirmation bias". *Psychology of learning and motivation* 32, pp. 385–418.

Köszegi, B. (2006). "Ego utility, overconfidence, and task choice". *Journal of the European Economic Association* 4 (4), pp. 673–707.

Kozakiewicz, M. (2021). *Belief-based utility and signal interpretation*. Tech. rep. Working paper.

Kranton, R. E. and Sanders, S. G. (2017). "Groupy versus non-groupy social preferences: Personality, region, and political party". *American Economic Review* 107 (5), pp. 65–69.

Kunda, Z. (1990). "The case for motivated reasoning." *Psychological bulletin* 108 (3), p. 480.

Lahey, J. N. and Oxley, D. (2016). "The power of eye tracking in economics experiments". *American Economic Review* 106 (5), pp. 309–13.

Martínez-Marquina, A., Niederle, M., and Vespa, E. (2019). "Failures in contingent reasoning: The role of uncertainty". *American Economic Review* 109 (10), pp. 3437–3474.

Mengel, F. and Campos-Mercade, P. (2021). "Irrational Statistical Discrimination". *Available at SSRN 3843579*.

Miserocchi, F. (2023). *Discrimination through biased memory*. Tech. rep. Working paper.

Neumark, D. (2018). "Experimental research on labor market discrimination". *Journal of Economic Literature* 56 (3), pp. 799–866.

Neumark, D., Burn, I., and Button, P. (2016). "Experimental age discrimination evidence and the Heckman critique". *American Economic Review* 106 (5), pp. 303–08.

Nickerson, R. S. (1998). "Confirmation bias: A ubiquitous phenomenon in many guises". *Review of general psychology* 2 (2), pp. 175–220.

Onuchic, P. (2022). "Recent Contributions to Theories of Discrimination". *arXiv preprint arXiv:2205.05994*.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., and Damer, E. (2022). "Data quality of platforms and panels for online behavioral research". *Behavior Research Methods* 54 (4), pp. 1643–1662.

Pettigrew, T. F. and Tropp, L. R. (2006). "A meta-analytic test of intergroup contact theory." *Journal of personality and social psychology* 90 (5), p. 751.

Phelps, E. S. (1972). "The statistical theory of racism and sexism". *American economic review* 62 (4), pp. 659–661.

Rabin, M. and Schrag, J. L. (1999). "First impressions matter: A model of confirmatory bias". *The quarterly journal of economics* 114 (1), pp. 37–82.

Rackstraw, E. (2022). "Bias-Motivated Updating in the Labor Market". *Available at SSRN 4278076*.

Rammstedt, B. and John, O. P. (2005). "Kurzversion des big five inventory (BFI-K)". *Diagnostica* 51 (4), pp. 195–206.

Rathelot, R. and Safi, M. (2022). "Taste-Based Discrimination". *Handbook on Economics of Discrimination and Affirmative Action*. Springer, pp. 1–32.

Reuben, E., Sapienza, P., and Zingales, L. (2014). "How stereotypes impair women's careers in science". *Proceedings of the National Academy of Sciences* 111 (12), pp. 4403–4408.

Sarsons, H. (2017). "Interpreting signals in the labor market: evidence from medical referrals". *Job Market Paper*, pp. 141–145.

Shi, Y. and Zhu, M. (2023). ""Model minorities" in the classroom? Positive evaluation bias towards Asian students and its consequences". *Journal of Public Economics* 220, p. 104838.

Sinclair, V. G. and Wallston, K. A. (2004). "The development and psychometric evaluation of the Brief Resilient Coping Scale". *Assessment* 11 (1), pp. 94–101.

Stötzer, L. S. and Zimmermann, F. (2024). "A note on motivated cognition and discriminatory beliefs". *Games and Economic Behavior* 147, pp. 554–562.

Thaler, M. (2020). "The Fake News Effect: Experimentally Identifying Motivated Reasoning Using Trust in News". *Available at SSRN 3717381*.

Zimmermann, F. (2020). "The dynamics of motivated beliefs". *American Economic Review* 110 (2), pp. 337–61.

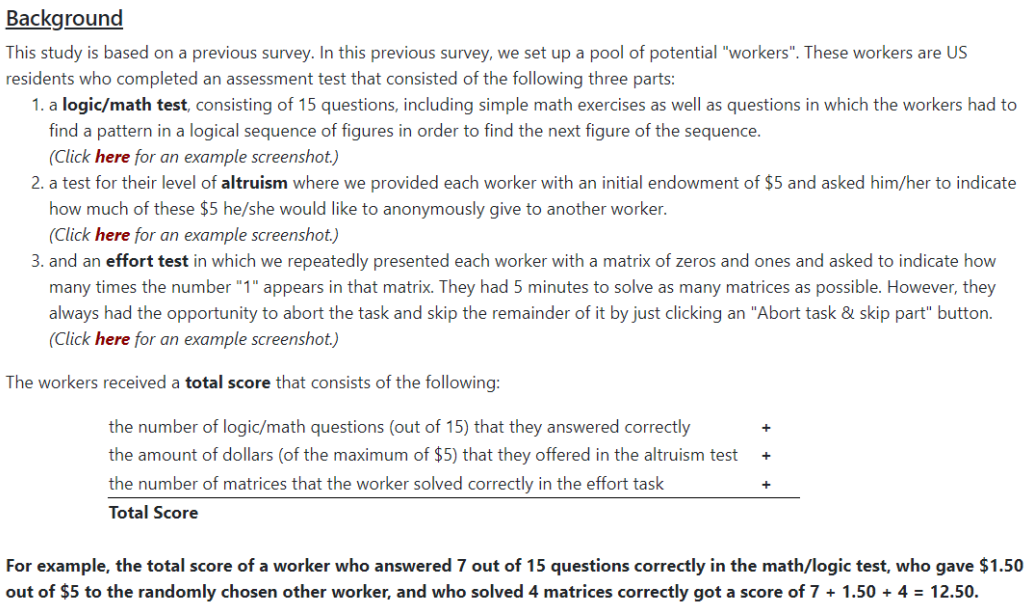
Online Appendix

A Experimental Design

This section provides example screenshots of all relevant screens for both treatments and across experiments.

A.1 Instructions

Figure A1: Background



Background

This study is based on a previous survey. In this previous survey, we set up a pool of potential "workers". These workers are US residents who completed an assessment test that consisted of the following three parts:

1. a **logic/math test**, consisting of 15 questions, including simple math exercises as well as questions in which the workers had to find a pattern in a logical sequence of figures in order to find the next figure of the sequence.
   *(Click **here** for an example screenshot.)*
2. a test for their level of **altruism** where we provided each worker with an initial endowment of $5 and asked him/her to indicate how much of these $5 he/she would like to anonymously give to another worker.
   *(Click **here** for an example screenshot.)*
3. and an **effort test** in which we repeatedly presented each worker with a matrix of zeros and ones and asked to indicate how many times the number "1" appears in that matrix. They had 5 minutes to solve as many matrices as possible. However, they always had the opportunity to abort the task and skip the remainder of it by just clicking an "Abort task & skip part" button.
   *(Click **here** for an example screenshot.)*

The workers received a **total score** that consists of the following:

the number of logic/math questions (out of 15) that they answered correctly	+
the amount of dollars (of the maximum of $5) that they offered in the altruism test	+
the number of matrices that the worker solved correctly in the effort task	+
Total Score	

For example, the total score of a worker who answered 7 out of 15 questions correctly in the math/logic test, who gave $1.50 out of $5 to the randomly chosen other worker, and who solved 4 matrices correctly got a score of 7 + 1.50 + 4 = 12.50.

Figure A2: Decision (group Neutral)

Your decision

In the main part of this study, you will play the role of an employer who is in charge of hiring some of the potential workers. You will repeatedly be presented with two workers and it is your task to hire one of them.
Note that we have set up four subgroups of workers, ◆, ■, ●, and ▲ and we have allocated each worker into one of these four groups.

Out of all hiring decisions that you make in this study, we will randomly select one hiring decision to be the **"decision-that-counts"**.

Your bonus payment

If you hired the better of the two workers in the decision-that-counts (i.e. the one with the higher total score in the assessment test), you will get a bonus payment of $2.00. However, if you hired the worse of the two workers in the decision-that-counts (i.e. the one with the lower score in the assessment test), you will get no bonus payment.

Data

All data that you enter in this survey will be treated anonymously and will solely be used for academic research. Neither the experimenter nor anyone else will be able to identify you from the answers that you give in this survey.

If you have read and understood these explanations, please click "Next" to answer **a few test questions** about the information on this page.

Figure A3: Decision (group Race)

Your decision

In the main part of this study, you will play the role of an employer who is in charge of hiring some of the potential workers. You will repeatedly be presented with two workers and it is your task to hire one of them.
Out of all hiring decisions that you make in this study, we will randomly select one hiring decision to be the **"decision-that-counts"**.

Your bonus payment

If you hired the better of the two workers in the decision-that-counts (i.e. the one with the higher total score in the assessment test), you will get a bonus payment of $2.00. However, if you hired the worse of the two workers in the decision-that-counts (i.e. the one with the lower score in the assessment test), you will get no bonus payment.

Data

All data that you enter in this survey will be treated anonymously and will solely be used for academic research. Neither the experimenter nor anyone else will be able to identify you from the answers that you give in this survey.

If you have read and understood these explanations, please click "Next" to answer **a few test questions** about the information on this page.

Figure A4: Test Questions

True or False: The workers completed a logic/math test, an altruism test and an effort test.

- ○ True
- ○ False

True or False: A very good worker to hire is one that has very good logic/math skills, is very altruistic and puts a lot of effort into tasks.

- ○ True
- ○ False

What is the total score of a worker who solved 5 math/logic questions correctly, who gave $2.44 to the randomly chosen other worker and who solved 3 matrices correctly in the effort task?

How high would your bonus payoff be, if you hired a worker with a total score of 15.3 instead of a worker with a total score of 10.5 in the decision-that-counts?

$

How high would your bonus payoff be, if you instead hired a worker with a total score of 2.60 instead of a worker with a total score of 5 in the decision-that-counts?

$

A.2 Belief Stage

This section shows screenshots of the belief stage. The screenshots show the elicitation of beliefs about Asian workers. The elicitation of beliefs about workers from the other groups was conducted identically.

Figure A5: Elicitation of Priors (group Neutral)

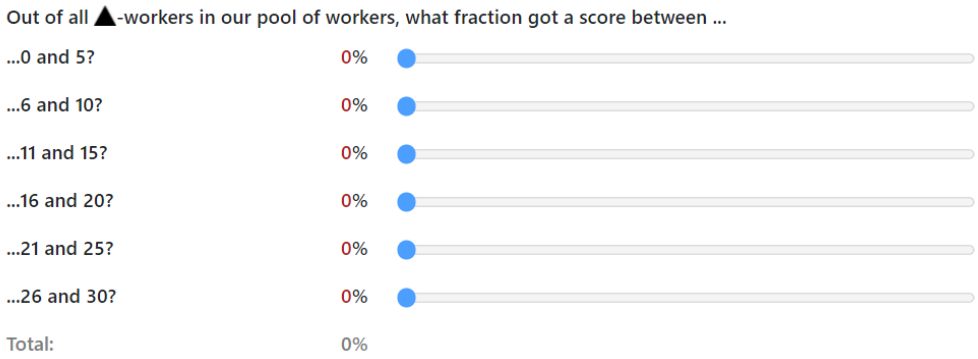


Figure A6: Elicitation of Priors (group Race)

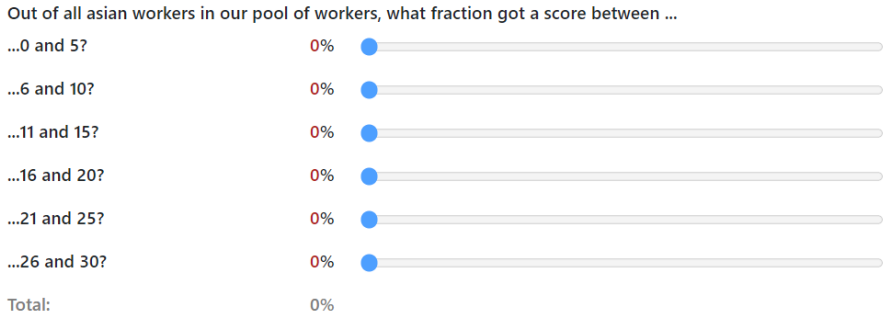


Figure A7: Elicitation of Posteriors (group Neutral)

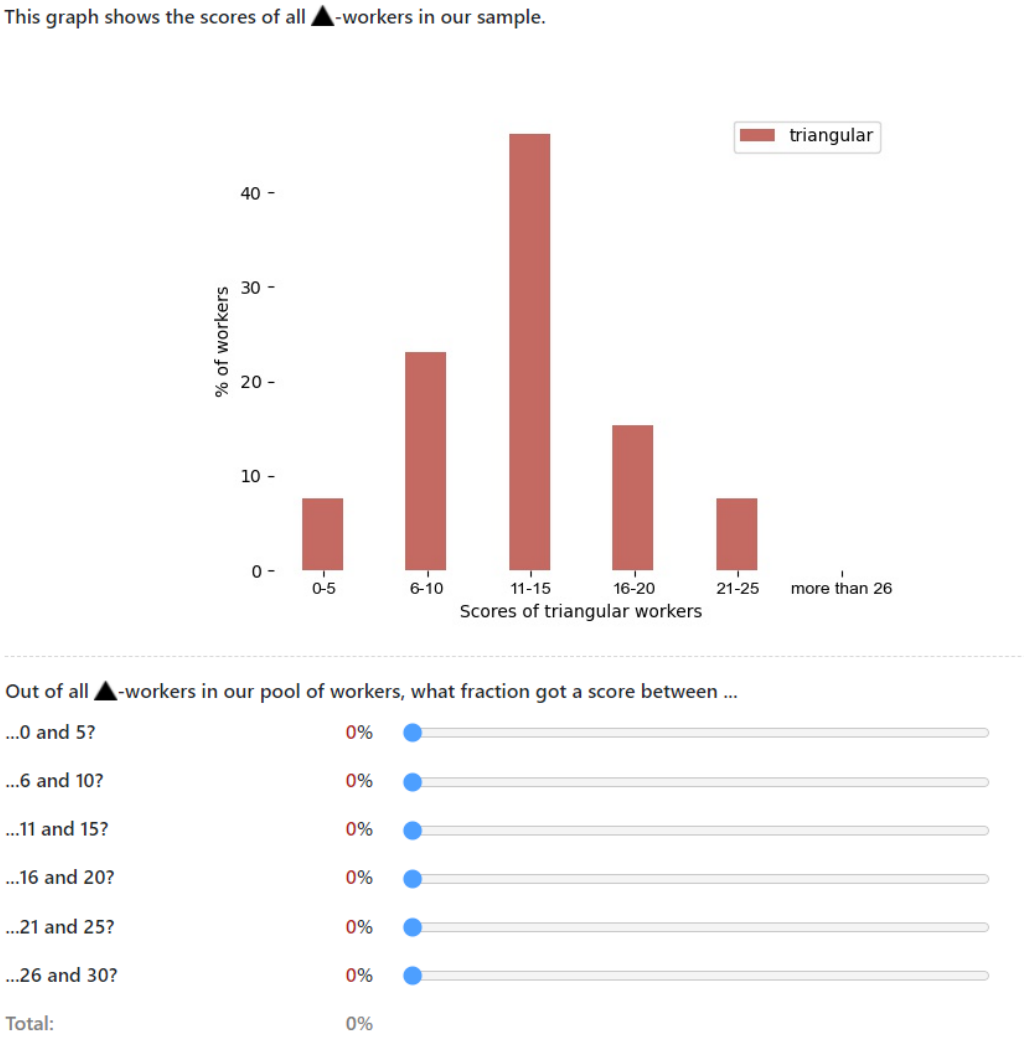
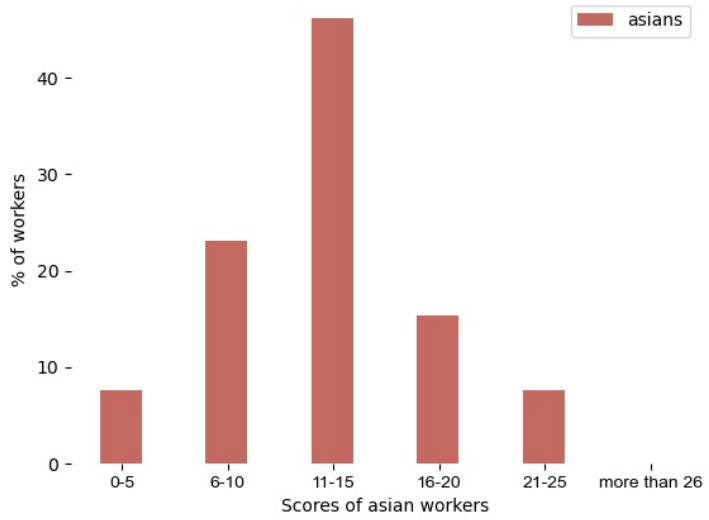


Figure A8: Elicitation of Posteriors (group Race)

This graph shows the scores of all asian workers in our sample.



Out of all asian workers in our pool of workers, what fraction got a score between ...

...0 and 5?	0%
...6 and 10?	0%
...11 and 15?	0%
...16 and 20?	0%
...21 and 25?	0%
...26 and 30?	0%
Total:	0%

A.3 Hiring Stage

Figure A9: Hiring Instructions (Experiment 'No Information')

You will now play the role of an employer who is in charge of hiring potential workers.

The task: You will repeatedly be presented with a randomly drawn pair of workers from our pool of workers. It is your task to hire one of these two workers.

Your bonus payment: In particular, in the end one of the decisions will be randomly drawn as the **decision-that-counts**. If you hired the better of the two workers in the decision-that-counts (i.e. the one with the **higher total score in the assessment test**), you will get a bonus payment of **$2.00**. However, if you hired the worse of the two workers in the decision-that-counts (i.e. the one with the **lower score in the assessment test**), you will get **no bonus** payment.

Figure A10: Hiring Instructions (Experiment 'Ambiguous Information')

You will now play the role of an employer who is in charge of hiring potential workers.

The task: You will repeatedly be presented with a randomly drawn pair of workers from our pool of workers. It is your task to hire one of the two workers.

Your bonus payment: In the end one of the decisions will be randomly drawn as the **decision-that-counts**. If you hired the **better** of the two workers in the decision-that-counts (i.e. the one with the **higher total score in the assessment test**), you will get a bonus payment of **$2.00**. However, if you hired the worse of the two workers in the decision-that-counts (i.e. the one with the **lower score in the assessment test**), you will get **no bonus** payment.

Important: On each hiring page, you will be given a message that is meant to tell you who the **better** worker is. However, the message comes from either a **True News** source or a **Fake News** source.

The True News source will always tell you the truth, whereas the Fake News source will never tell you the truth.

If you want, you can request more messages for each decision by clicking on the "**Request Another Message**" button below the two workers.

Whether a message comes from the True News source or the Fake News source is **randomly** determined for each message.

It is up to you how many messages (max. 10) you request and whether or not you follow them.

Figure A11: Hiring Instructions (Experiment 'Uncertain Information')

You will now play the role of an employer who is in charge of hiring potential workers.

The task: You will repeatedly be presented with a randomly drawn pair of workers from our pool of workers. It is your task to hire one of the two workers.

Your bonus payment: In the end one of the decisions will be randomly drawn as the **decision-that-counts**. If you hired the **better** of the two workers in the decision-that-counts (i.e. the one with the **higher total score in the assessment test**), you will get a bonus payment of **$2.00**. However, if you hired the worse of the two workers in the decision-that-counts (i.e. the one with the **lower score in the assessment test**), you will get **no bonus** payment.

Important: On each hiring page, you will be given a message that is meant to tell you who the **better** worker is. However, the message comes from either a **True News** source or a **Fake News** source.

The True News source will always tell you the truth, whereas the Fake News source will never tell you the truth.

If you want, you can request more messages for each decision by clicking on the "**Request Another Message**" button below the two workers.

The likelihood that a message comes from the **True News source is 60%**, the likelihood that a message comes from the **Fake News source is 40%**.

It is up to you how many messages (max. 10) you request and whether or not you follow them.

Figure A12: Hiring Instructions (Experiment 'Tangible Information')

You will now play the role of an employer who is in charge of hiring potential workers.

The task: You will repeatedly be presented with a randomly drawn pair of workers from our pool of workers. It is your task to hire one of these two workers.

Your bonus payment: In particular, in the end one of the decisions will be randomly drawn as the **decision-that-counts**. If you hired the better of the two workers in the decision-that-counts (i.e. the one with the **higher total score in the assessment test**), you will get a bonus payment of **$2.00**. However, if you hired the worse of the two workers in the decision-that-counts (i.e. the one with the **lower score in the assessment test**), you will get **no bonus** payment.

Important: On each hiring page, you will be given a random piece of information about the two workers.

If you want you can request more information for each decision by clicking on the **"Get more information"** button below the two workers.

It is up to you how many pieces of information (max. 10) you request and whether or not you follow them.

Figure A13: Hiring Decision (Experiment 'No Information', group Neutral)

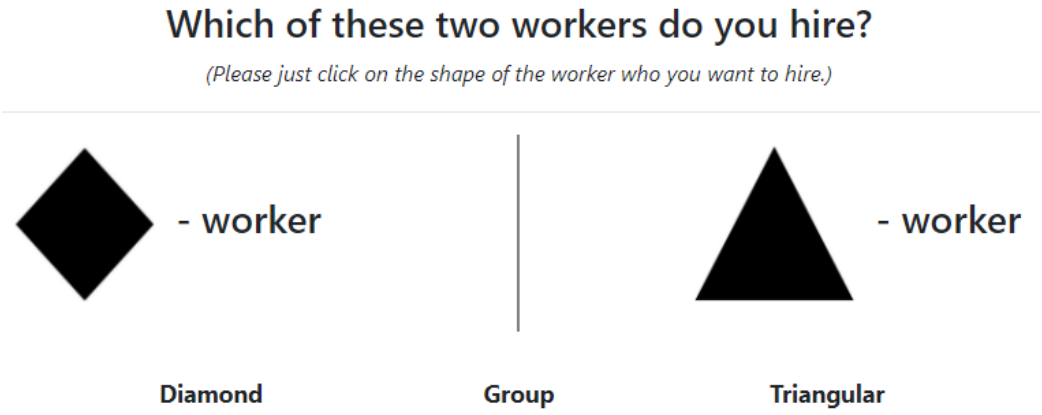


Figure A14: Hiring Decision (Experiment 'No Information', group Race)

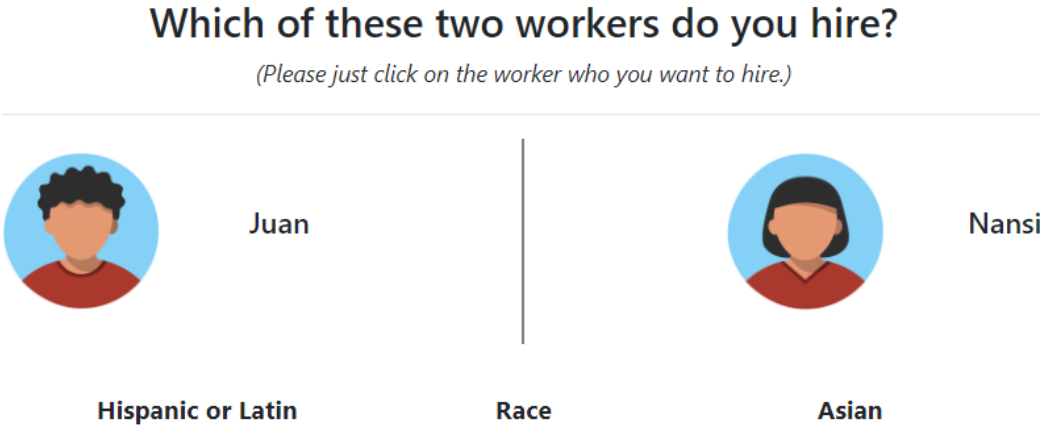


Figure A15: Hiring Decision (Experiment 'Ambiguous Information' & 'Uncertain Information', group Neutral)

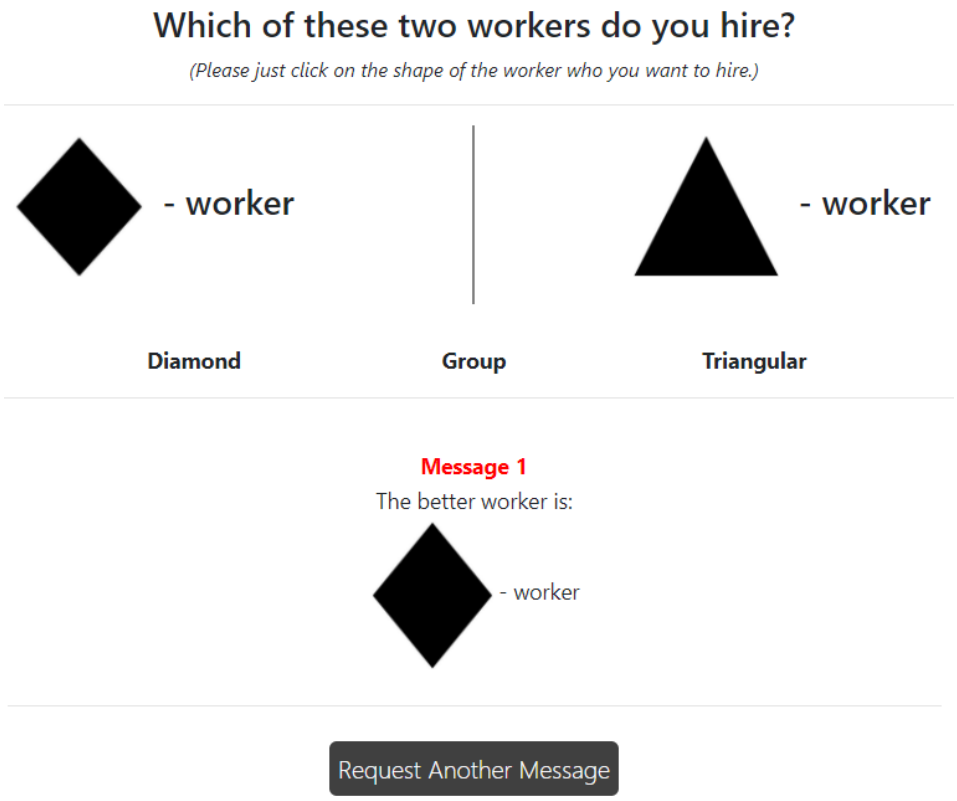


Figure A16: Hiring Decision (Experiment 'Ambiguous Information' & 'Uncertain Information', group Race)

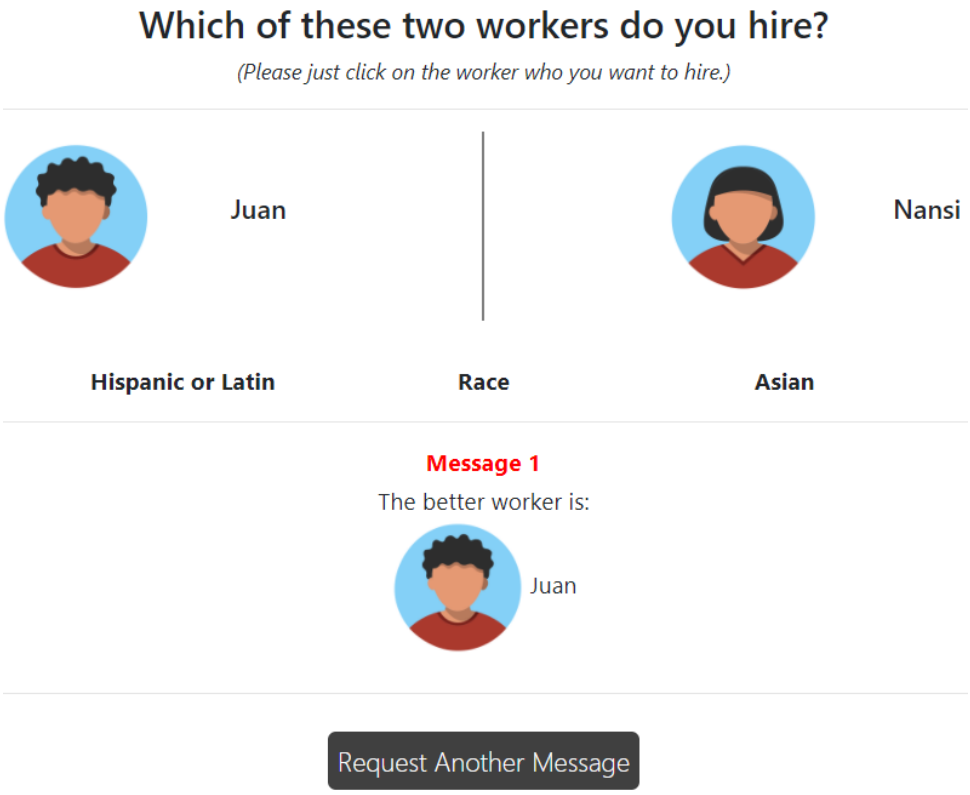


Figure A17: Hiring Decision (Experiment 'Tangible Information', group Neutral)

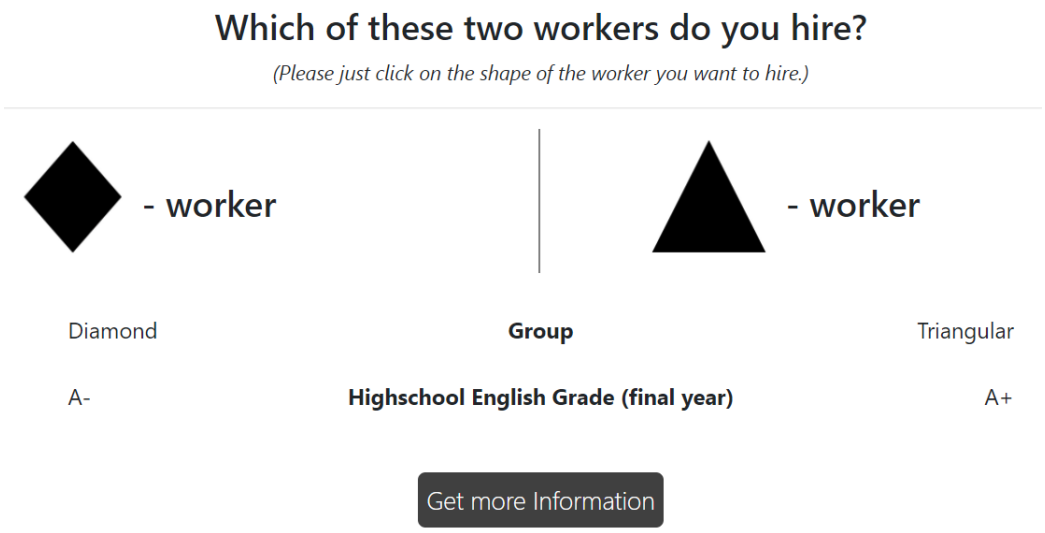
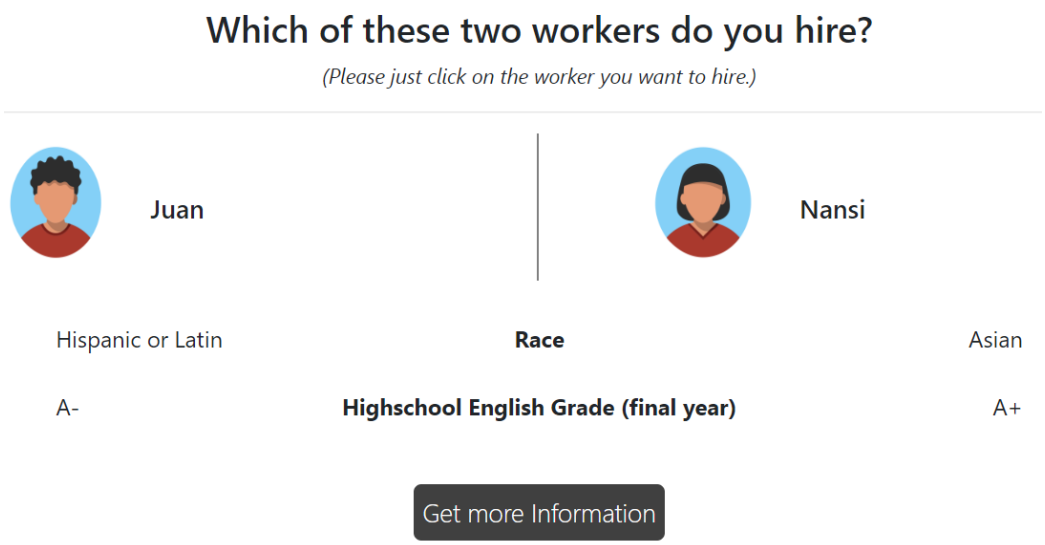


Figure A18: Hiring Decision (Experiment 'Tangible Information', group Race)



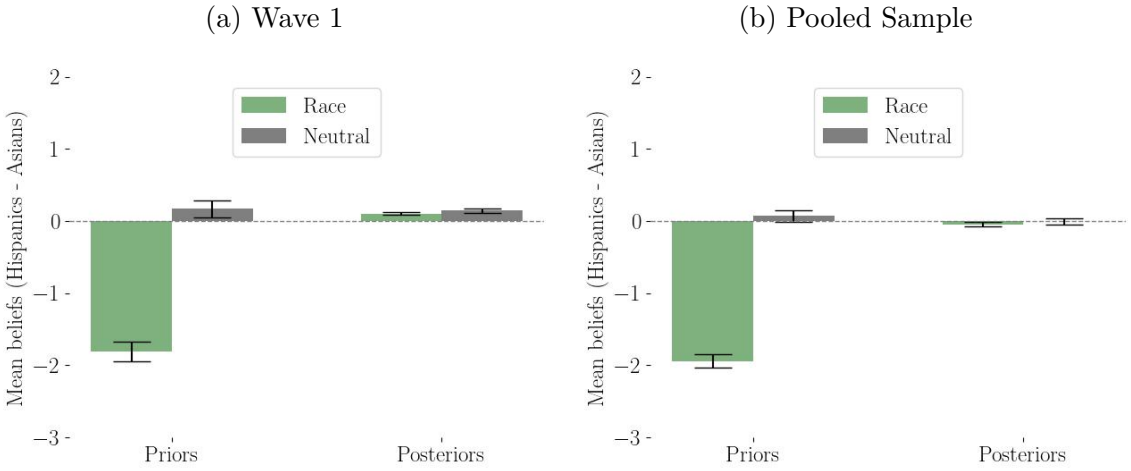
B Previous Data Collection

This sections describes the full data collection process and presents the main results for the pooled sample as well as for wave 1 only.

The hiring experiments were conducted in two waves: Wave 1 in December 2021 and Wave 2 in September 2025. Both waves were pre-registered prior to data collection. The pre-registration for Wave 1, however, was not fully specified, and the realized sample size was smaller than the pre-registered target of 2,500 employers, due to initial resource constraints. To address this, Wave 2 implemented a fully pre-registered replication based on the working paper version of this study, with all hypotheses, analyses, treatments, and exclusion criteria clearly defined. Pooling the two waves yields a combined sample of 2,533 participants.

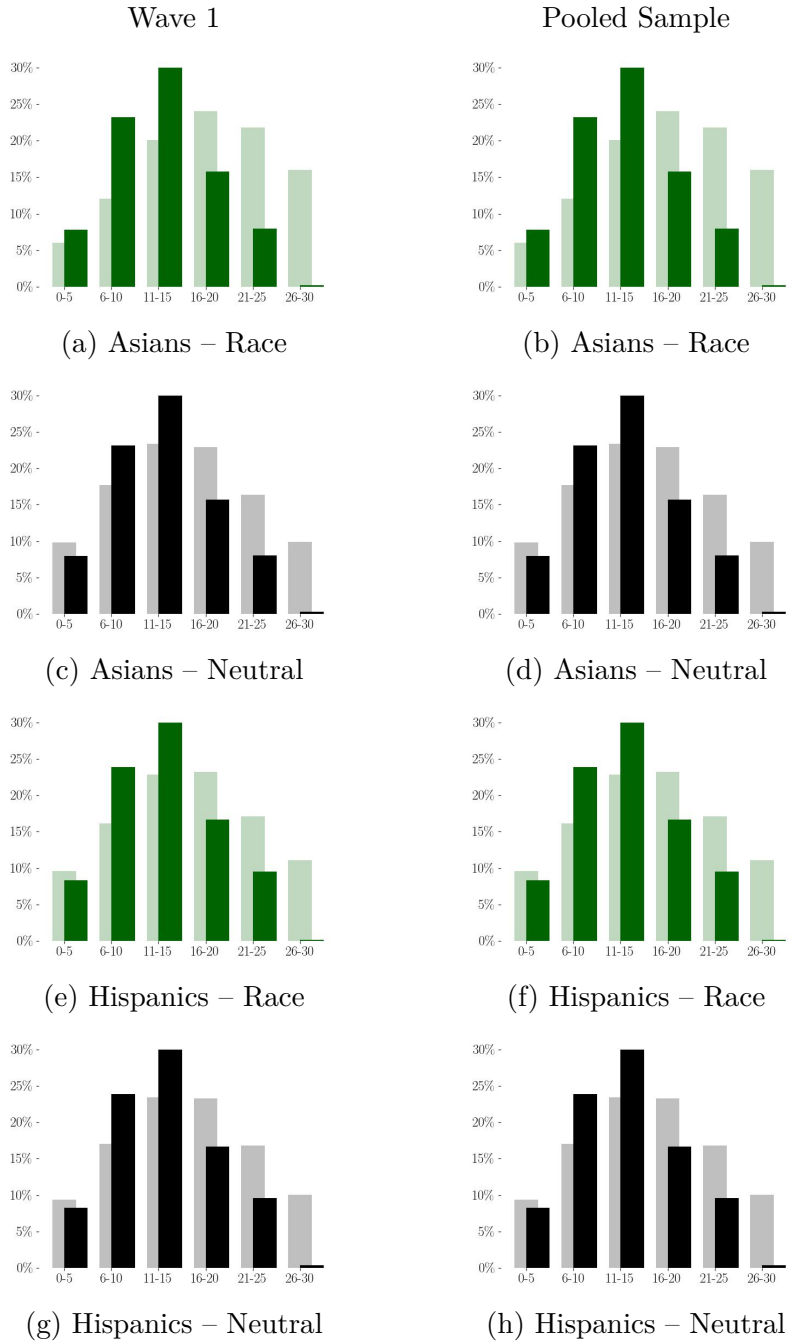
Wave 1 was pre-registered in the AEA RCT Registry under AEARCTR-0008609. In wave 1, $1,267$ participants completed the study in time. After application of the exclusion criteria (pass at least 60% of the comprehension questions), I ended up with a sample of $1,138$ participants, unevenly split across the four experiments as follows: 454 participants in the 'No Information' experiment, 234 participants in the 'Ambiguous Information' experiment 108 participants in the 'Uncertain Information' experiment, and 106 participants in the 'Tangible Information' experiment.
This Section presents the results of the main text for wave 1 as well as for the pooled sample.

Figure B1: Mean beliefs about group productivities



Notes: This figure shows the mean productivity beliefs. The first two bars in each panel show the mean productivity beliefs *before* the group-level information update of employers in treatment Race and treatment Neutral, respectively, the last two bars in each panel show the mean productivity beliefs *after* the group-level information update. Beliefs are plotted as the difference in mean beliefs about Hispanic workers and mean beliefs about Asian workers such that negative values indicate higher beliefs about the productivity of Asian workers than beliefs about the productivity of Hispanic workers, and vice versa. Panel (a) shows data from wave 1, panel (b) shows data from the pooled sample. Error bars show the standard errors of the means.

Figure B2: Prior and Posterior belief distributions



(a) Asians – Race (b) Asians – Race

(c) Asians – Neutral (d) Asians – Neutral

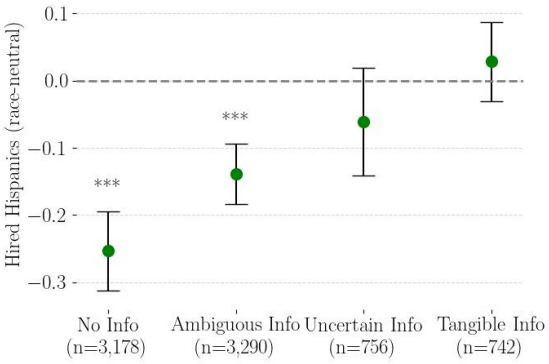
(e) Hispanics – Race (f) Hispanics – Race

(g) Hispanics – Neutral (h) Hispanics – Neutral

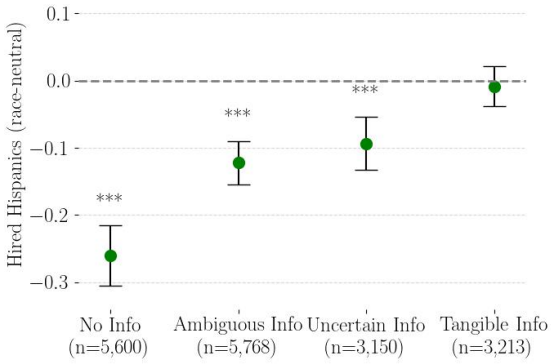
Notes: Panels (a) – (d) show employer belief distributions about the productivities of Asian workers before (low opacity) and after (high opacity) the group-level information update in treatment Race and treatment Neutral, respectively. Panels (e) – (f) show employer belief distributions about the productivities of Hispanic workers before (low opacity) and after (high opacity) the group-level information update in treatment Race and treatment Neutral, respectively. The left panels show data from wave 1, the right panels use data from the pooled sample.

Figure B3: Disparities in hiring rates across experiments

(a) Wave 1

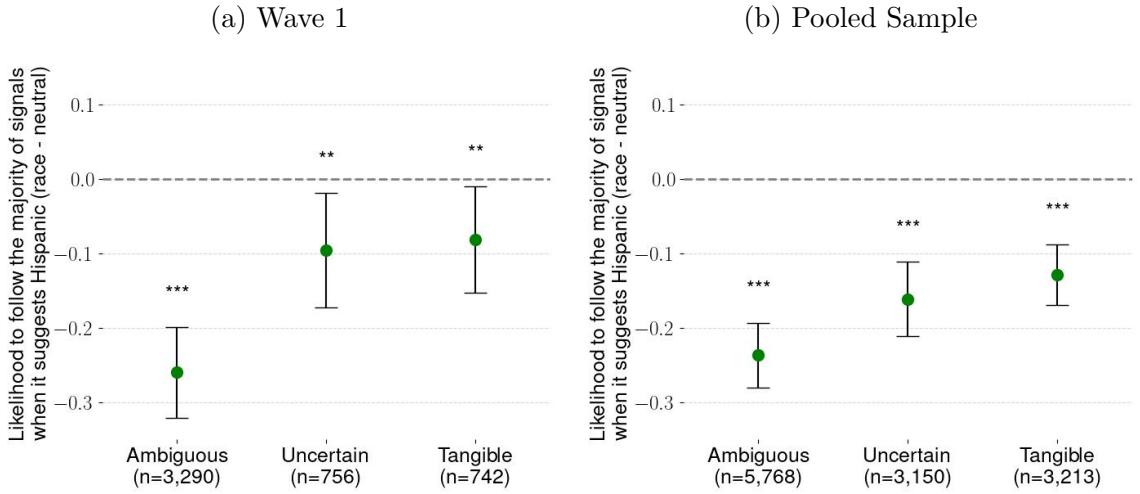


(b) Pooled Sample



Notes: This figure shows the coefficients from an OLS regression of a dummy that is 1 if the Hispanic worker is hired and 0 otherwise, on the treatment dummy (Race = 1, Neutral = 0), controlling for employer fixed effects. *Negative* coefficients indicate discrimination against Hispanic workers. The horizontal axis groups results by experiment ('No Information', 'Ambiguous Information', 'Uncertain Information', and 'Tangible Information'). Panel (a) shows data from wave 1, panel (b) shows data from the pooled sample. Standard errors are clustered at the individual level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

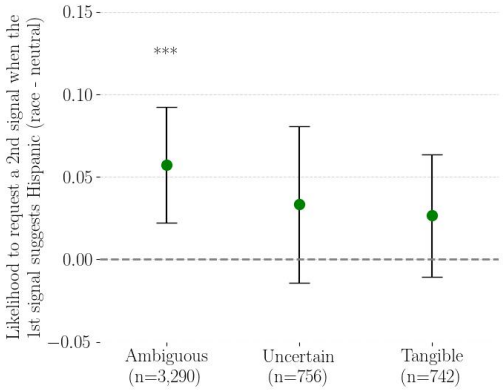
Figure B4: Systematically different information processing



Notes: This figure shows the coefficients from an OLS regression of a dummy that is 1 if the employer follows the majority of observed signals on the interaction between a treatment dummy (Race = 1, Neutral = 0) and a dummy equal to 1 if the majority of signals suggested hiring the Hispanic worker, controlling for employer fixed effects. *Negative* coefficients indicate that employers in the Race treatment were *less* likely to follow the majority of signals when they favored the Hispanic worker. The horizontal axis groups results by experiment ('Ambiguous Information', 'Uncertain Information', and 'Tangible Information'). Panel (a) shows data from wave 1, panel (b) shows data from the pooled sample. Standard errors are clustered at the individual level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

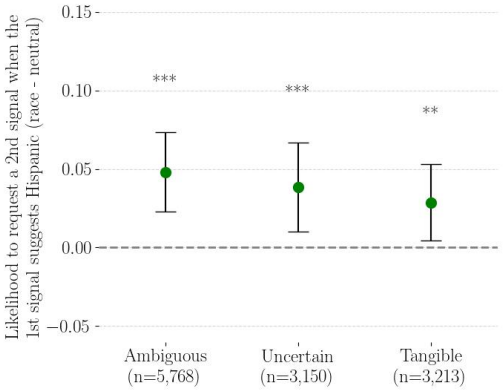
Figure B5: Systematically different information search

Wave 1

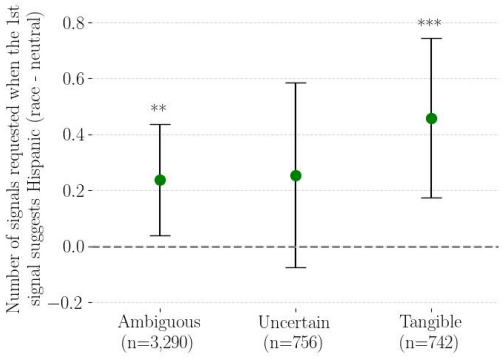


(a) Likelihood to request a 2nd signal

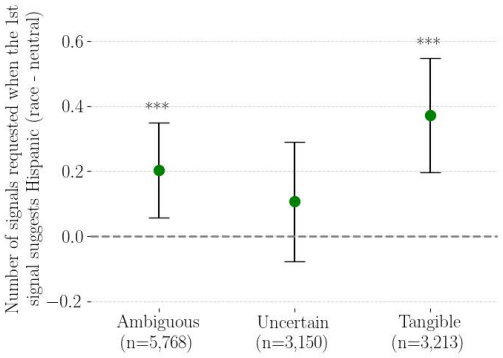
Pooled Sample



(b) Likelihood to request a 2nd signal



(c) Number of signals requested per decision



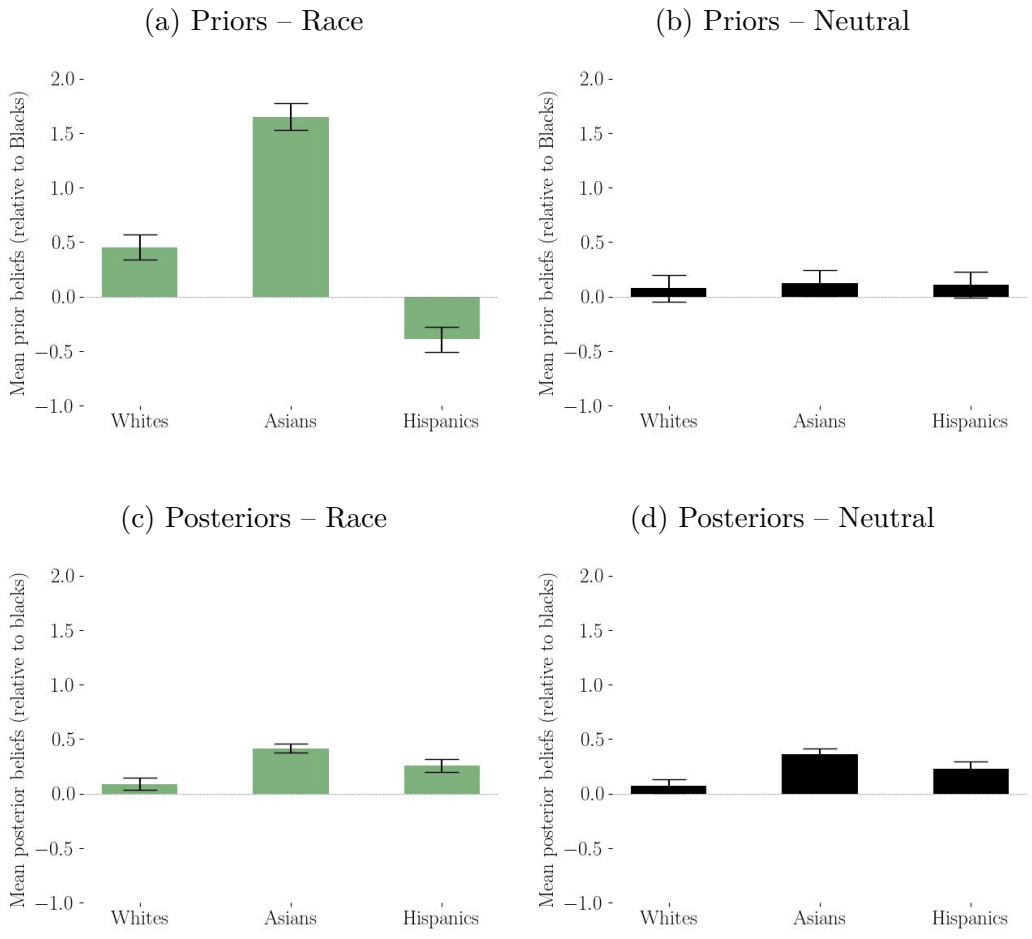
(d) Number of signals requested per decision

Notes: Panels (a) and (b) of this figure show the coefficients from an OLS regression of a dummy that is 1 if the employer requests a second signal on the interaction between a treatment dummy (Race = 1, Neutral = 0) and a dummy equal to 1 if the first signal suggested hiring the Hispanic worker, controlling for employer fixed effects. *Positive* coefficients indicate that employers in the Race treatment were *more* likely to request a second signal when the first signal favored the Hispanic worker. Panels (c) and (d) show the coefficients from an OLS regression of the number of requested signals per decision on the same interaction, again controlling for employer fixed effects. *Positive* coefficients indicate that employers in the Race treatment requested *more* signals when the first signal favored the Hispanic worker. The horizontal axes group results by experiment ('Ambiguous Information', 'Uncertain Information', and 'Tangible Information'). The left panels show data from wave 1, the right panels use data from the pooled sample. Standard errors are clustered at the individual level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

C Group-level Beliefs

Figure C1 shows the relation between mean beliefs about black, white, Asian and Hispanic workers and provides evidence, that the largest difference in prior beliefs, and hence in potential for motivated beliefs, lies in the difference between Asian and Hispanic workers.

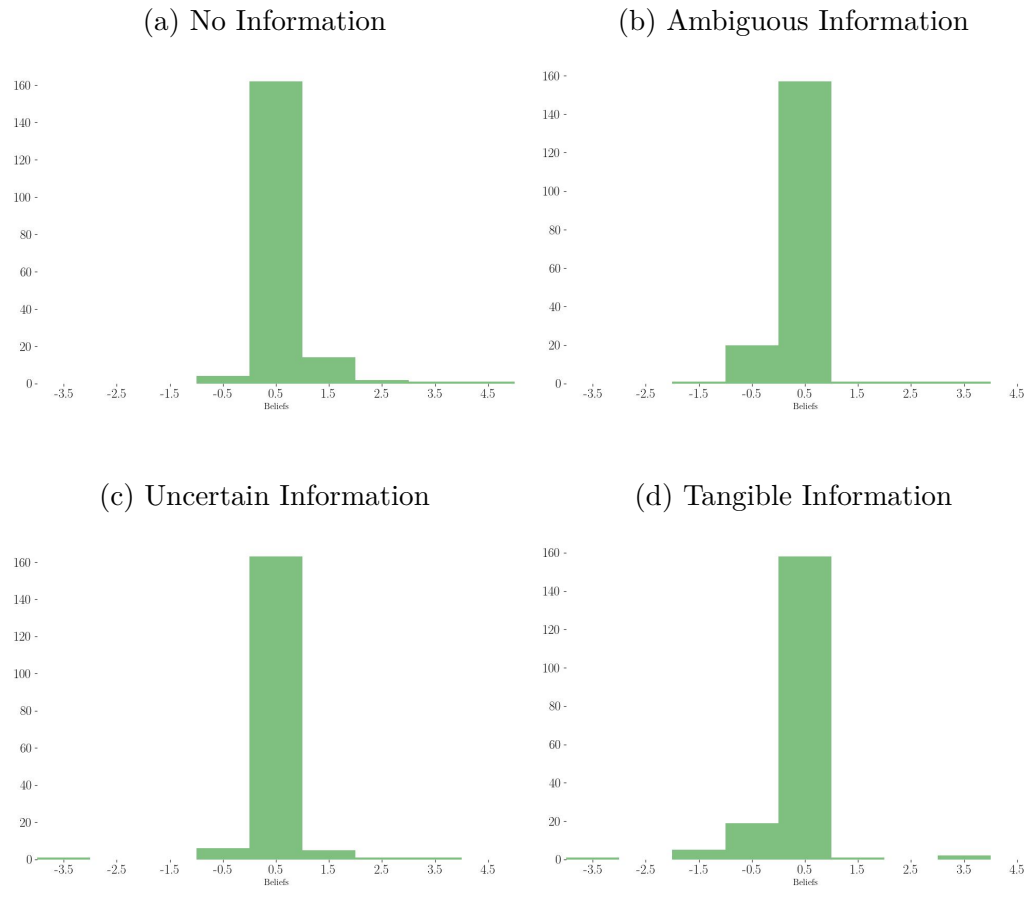
Figure C1: Mean beliefs about group productivities



Notes: Figures (a) and (b) show the mean beliefs before the group-level information update of employers in group Race and group Neutral, respectively, figures (c) and (d) show the mean beliefs after the group-level information update of employers in group Race and group Neutral, respectively. Beliefs are plotted as the difference in means compared to beliefs about blacks. Error bars show the standard errors of the means.

Figure C2 shows differences in differences in posterior beliefs about Asian and Hispanic workers between treatment Race and treatment Neutral for each experiment separately. More specifically, I calculate the differences in mean beliefs about Asian and Hispanic workers for each employer of treatment Race and subtract the difference of the mean of the mean beliefs about Asian and Hispanic workers of employers in treatment Neutral. This is a measure of heterogeneity in posterior beliefs and shows that beliefs were homogeneously updated across experiments.

Figure C2: Heterogeneity in posterior beliefs



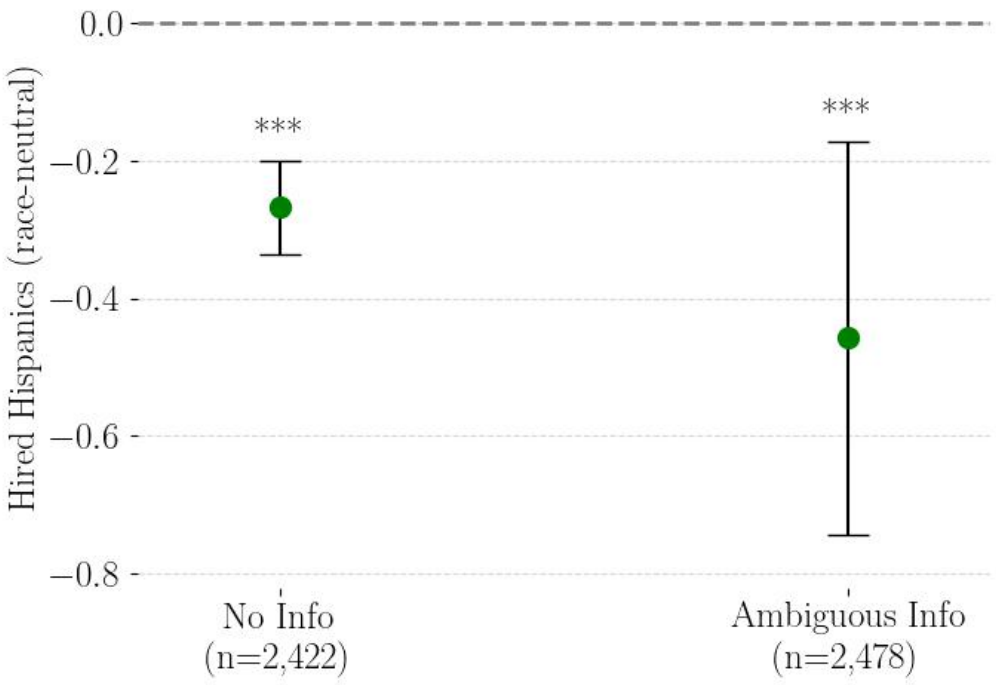
Notes: The figures show histograms of the differences in differences between employer beliefs about Asian and Hispanic workers, and between both treatments. The calculations uses the individual posteriors of treatment Race and the mean posteriors of treatment Neutral, as no employer participated in both treatments. Negative values imply higher beliefs for Asians than for Hispanics in treatment Race than in treatment Neutral.

D Additional Results

This section provides additional results. Section D.1 shows results from decisions in the 'Ambiguous Information' experiment, when individuals requested one additional signal and ended up with two conflicting signals. Section D.2 provides results of the counter-hypotheses for hypothesis 2a and hypothesis 2b. Section D.3 adds the numbers to all figure of the main text. Finally, Section D.4 provides results of some robustness exercises.

D.1 Discrimination under Conflicting Signals

Figure D1: Hiring rates of Hispanic workers under wiggle room



Notes: The vertical axis displays the fraction of hired Hispanics in treatment Race minus the fraction of hired Hispanics in treatment Neutral. The horizontal axis groups the decisions by experiments ('No Information', 'Ambiguous Information'). Decisions in the 'Ambiguous Information' experiment only include decisions in which subjects only requested one additional signal and ended up with two conflicting signals. Error bars indicate standard errors of the means.

Figure D1 provides further support to the idea that individuals misperceived the informativeness of the ambiguous signals to align with their motives to form the

belief that the Asian worker is more productive. Given that subjects believe there is a 50% chance of Fake News for each signal, in the absences of motivated reasoning, decisions in the 'Ambiguous Information' experiment with two conflicting signals (i.e., one suggesting the Hispanic worker and one suggesting the Asian worker) should be similar to decisions in the No Information experiment. Instead, Figure D1 shows that there is more discrimination in the 'Ambiguous Information' experiment in this case.

D.2 Counter-hypotheses

To give more insights into the data structure this section provides results for the following "counter-hypotheses" for hypothesis 2a and hypothesis 2b:

Counter-hypothesis 2a: *If the majority of all considered signals in a decision suggests to hire the Asian worker in the 'Ambiguous Information' experiment, employers in treatment Race are more likely to act according to the suggestion than employers in treatment Neutral.*

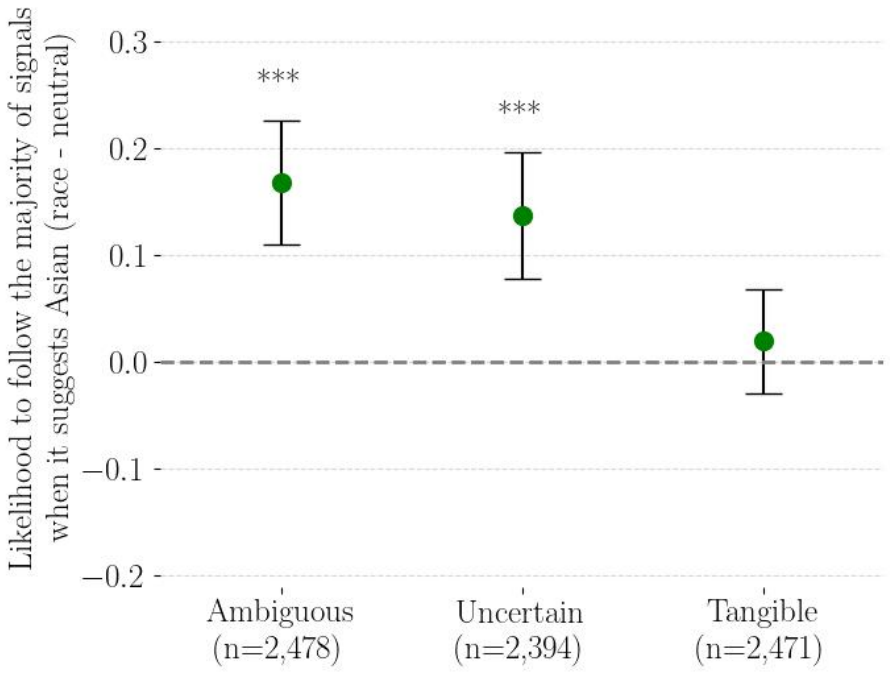
Counter-hypothesis 2b: *If the majority of all considered signals in a decision suggests to hire the Asian worker, the difference in the likelihood to follow the suggestion between treatment Race and treatment Neutral is larger in the 'Ambiguous Information' experiment than in the 'Uncertain Information' and the 'Tangible Information' experiments.*

Counter-hypothesis 3: *If the initial information signal suggests to hire the Asian worker in the 'Ambiguous Information' experiment, employers in treatment Race are less likely to request a 2nd signal and request less signals than employers in treatment Neutral.*

These counter-hypotheses compare the subjects that stopped requesting signals when the majority of the received signals suggested to hire the Asian worker (instead of the Hispanic worker).

The results in Figures D2 and D3 confirm the "counter-hypotheses" and hence lend further support for the motivated information acquisition behavior described in the main text.

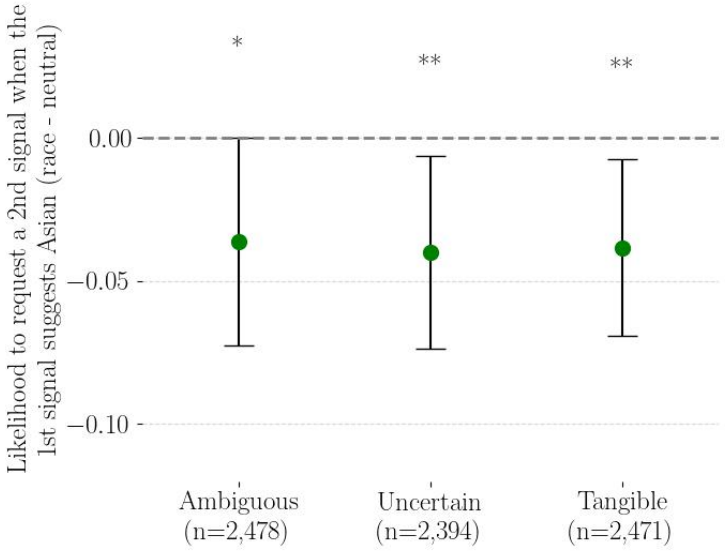
Figure D2: Results of the Counter-hypotheses of Hypotheses 2a and 2b



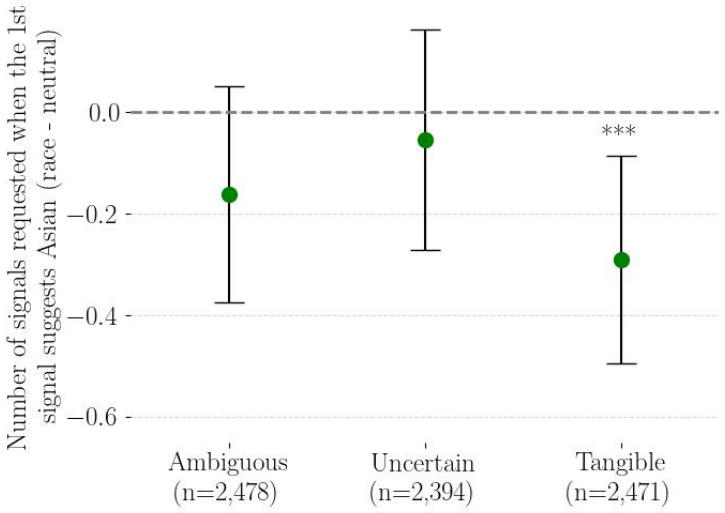
Notes: This figure shows the coefficients from an OLS regression of a dummy that is 1 if the employer follows the majority of observed signals on the interaction between a treatment dummy (Race = 1, Neutral = 0) and a dummy equal to 1 if the majority of signals suggested hiring the Asian worker, controlling for employer fixed effects. *Positive* coefficients indicate that employers in the Race treatment were *more* likely to follow the majority of signals when they favored the Asian worker. The horizontal axis groups results by experiment ('Ambiguous Information', 'Uncertain Information', and 'Tangible Information'). Standard errors are clustered at the individual level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Figure D3: Results of the Counter-hypothesis of Hypothesis 3

(a) Likelihood to request a 2nd signal



(b) Number of signals requested per decision



Notes: Panel (a) of this figure shows the coefficients from an OLS regression of a dummy that is 1 if the employer requests a second signal on the interaction between a treatment dummy (Race = 1, Neutral = 0) and a dummy equal to 1 if the first signal suggested hiring the Asian worker, controlling for employer fixed effects. *Negative* coefficients indicate that employers in the Race treatment were *less* likely to request a second signal when the first signal favored the Asian worker. Panel (b) shows the coefficients from an OLS regression of the number of requested signals per decision on the interaction between a treatment dummy (Race = 1, Neutral = 0) and a dummy equal to 1 if the first signal suggested hiring the Hispanic worker, controlling for employer fixed effects. *Negative* coefficients indicate that employers in the Race treatment requested *less* signals when the first signal favored the Asian worker. The horizontal axes group results by experiment ('Ambiguous Information', 'Uncertain Information', and 'Tangible Information'). Standard errors are clustered at the individual level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

D.3 Main Results

Table D1 shows the main text results across experiments and thereby provides the numbers for Figures 3 to 5 of the main text.

Table D1: Main results across experiments

	Information Experiment			
	(1) No	(2) Ambiguous	(3) Uncertain	(4) Tangible
Discrimination				
Race	-0.267***	-0.102***	-0.104***	-0.020
	(0.035)	*(0.024)*	*(0.023)*	*(0.017)*
Mean (neutral)	0.479	0.512	0.543	0.467
Following Signals				
Race*Hispanic (maj)		-0.209***	-0.180***	-0.143***
		(0.031)	*(0.031)*	*(0.024)*
Mean (neutral)		0.828	0.890	0.848
Request 2nd Signal				
Race*Hispanic (1st)		0.036*	0.040**	0.029*
		(0.019)	*(0.017)*	*(0.015)*
Mean (neutral)		0.593	0.761	0.808
Number of Signals				
Race*Hispanic (1st)		0.162	0.064	0.344***
		(0.108)	*(0.110)*	*(0.108)*
Mean (neutral)		4.079	5.071	5.154
N	2,422	2,478	2,394	2,471

*Notes: This table provides the OLS results underlying the presented estimated in Figures 3–5. Outcomes are presented in bold, 'Race' is a treatment dummy equal to 1 if the employer was in treatment Race, and 0 otherwise, 'Race*Hispanic (maj)' refers to the interaction of an employer in treatment Race who got a majority of signals that suggest to hire the Hispanic worker, 'Race*Hispanic (1st)' refers to the interaction of an employer in treatment Race who got a first signal that suggests to hire the Hispanic worker. Mean (neutral) show the means of the respective outcome in treatment 'Neutral'. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.*

D.4 Robustness Checks

During the Instructions, participants in all four experiments had to answer a few comprehension questions in order to make sure that instructions were understood. While the main text includes all participants that answered at least 3 out of 5 of these questions correctly, this section shows results for stricter limits. Table D2 shows results for participants who answered at least 4 out of 5 questions correctly, Table D3 for those who answered all comprehension questions correctly. The results of the main text are robust to these variations.

Table D2: Main results across experiments

	Information Experiment			
	(1) No	(2) Ambiguous	(3) Uncertain	(4) Tangible
Discrimination				
Race	-0.264***	-0.083***	-0.114***	-0.013
	(0.037)	*(0.025)*	*(0.024)*	*(0.018)*
Mean (neutral)	0.476	0.506	0.555	0.465
Following Signals				
Race*Hispanic (maj)		-0.189***	-0.180***	-0.136***
		(0.031)	*(0.032)*	*(0.025)*
Mean (neutral)		0.829	0.891	0.846
Request 2nd Signal				
Race*Hispanic (1st)		0.038*	0.038**	0.027*
		(0.020)	*(0.018)*	*(0.016)*
Mean (neutral)		0.606	0.780	0.818
Number of Signals				
Race*Hispanic (1st)		0.202*	0.046	0.346***
		(0.117)	*(0.116)*	*(0.112)*
Mean (neutral)		4.112	5.240	5.276
N	2,184	2,254	2,177	2,219

*Notes: This table provides the OLS results of the main analyses on a sample of employers who correctly answered at least 4 out of 5 comprehension questions. Outcomes are presented in bold, 'Race' is a treatment dummy equal to 1 if the employer was in treatment Race, and 0 otherwise, 'Race*Hispanic (maj)' refers to the interaction of an employer in treatment Race who got a majority of signals that suggest to hire the Hispanic worker, 'Race*Hispanic (1st)' refers to the interaction of an employer in treatment Race who got a first signal that suggests to hire the Hispanic worker. Mean (neutral) show the means of the respective outcome in treatment 'Neutral'. *, **, and *** denote significance at the* 10%, 5%, *and* 1% *levels, respectively.*

Table D3: Main results across experiments

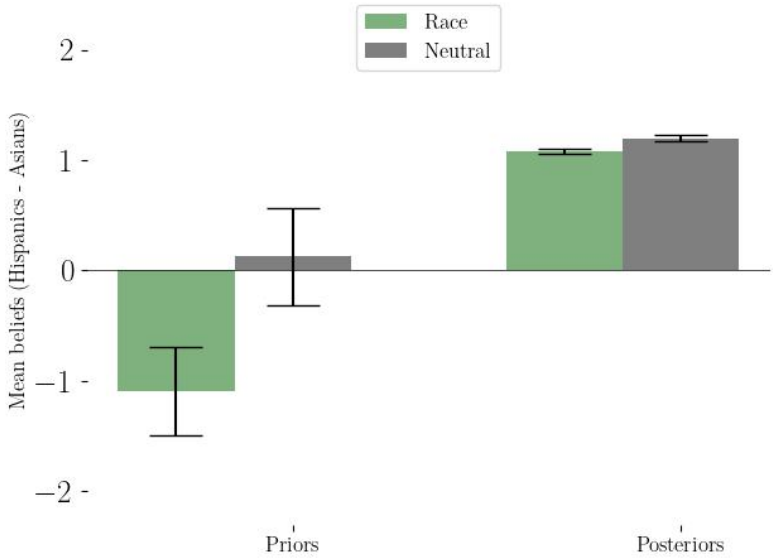
	Information Experiment			
	(1) No	(2) Ambiguous	(3) Uncertain	(4) Tangible
Discrimination				
Race	-0.260***	-0.061**	-0.134***	-0.003
	(0.045)	*(0.027)*	*(0.028)*	*(0.022)*
Mean (neutral)	0.482	0.482	0.574	0.455
Following Signals				
Race*Hispanic (maj)		-0.174***	-0.185***	-0.142***
		(0.038)	*(0.038)*	*(0.030)*
Mean (neutral)		0.840	0.892	0.839
Request 2nd Signal				
Race*Hispanic (1st)		0.049*	0.007	0.017
		(0.026)	*(0.019)*	*(0.020)*
Mean (neutral)		0.601	0.788	0.835
Number of Signals				
Race*Hispanic (1st)		0.311**	-0.092	0.302**
		(0.144)	*(0.130)*	*(0.136)*
Mean (neutral)		4.020	5.247	5.400
N	1,519	1,652	1,617	1,631

*Notes: This table provides the OLS results of the main analyses on a sample of employers who correctly answered all 5 comprehension questions. Outcomes are presented in bold, 'Race' is a treatment dummy equal to 1 if the employer was in treatment Race, and 0 otherwise, 'Race*Hispanic (maj)' refers to the interaction of an employer in treatment Race who got a majority of signals that suggest to hire the Hispanic worker, 'Race*Hispanic (1st)' refers to the interaction of an employer in treatment Race who got a first signal that suggests to hire the Hispanic worker. Mean (neutral) show the means of the respective outcome in treatment 'Neutral'. *, **, and *** denote significance at the* 10%, 5%, *and* 1% *levels, respectively.*

E Additional Experiments

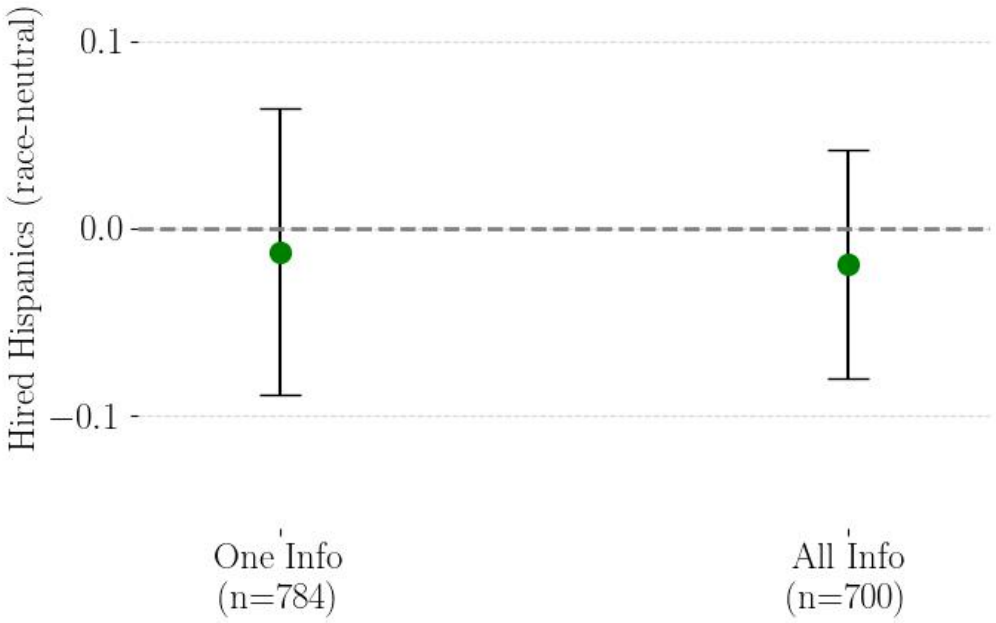
Since neither the One Information experiment nor the All Information experiment allow to study the information acquisition and processing behavior without any process tracing techniques, results of the 'One Information' and 'All Information' experiments are not included in the main text but mainly serve as additional evidence. Figure E2 confirms the prior and posterior beliefs from the other four experiments as shown in Figure 1 from the main text. Figure E2 summarizes the discrimination rates in these two experiments. Neither of the experiments shows rates of discrimination, which is consistent with the limited wiggle room in both experiments.

Figure E1: Mean beliefs about group productivities (All Info & One Info)



Notes: The first two bars show the mean productivity beliefs *before* the group-level information update of employers in treatment Race and treatment Neutral, respectively, the last two bars show the mean productivity beliefs *after* the group-level information update. Beliefs are plotted as the difference in mean beliefs about Hispanic workers and mean beliefs about Asian workers such that negative values indicate higher beliefs about the productivity of Asian workers than beliefs about the productivity of Hispanic workers, and vice versa. The data underlying this figure are pooled from the One Information and All Information experiments. Error bars show the standard errors of the means.

Figure E2: No disparities in hiring rates across additional experiments



Notes: This figure shows the coefficients from an OLS regression of a dummy that is 1 if the Hispanic worker is hired and 0 otherwise, on the treatment dummy (Race = 1, Neutral = 0), controlling for employer fixed effects. *Negative* coefficients indicate discrimination against Hispanic workers. The horizontal axis groups results by experiment ('One Information', 'All Information'). Standard errors are clustered at the individual level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.