# Seeing is Believing? How Learning Modes Shape Belief Bias and Discrimination

Christina (Che) Sun*

February 19, 2026

[Click Here for Latest Version]

**Abstract**

Can learning from accurate information entrench bias rather than correct it? This paper shows that endogenous information acquisition sustains biased beliefs despite accurate signals. Through a combination of theory and experiment, I show that even when accurate information is available, individuals who can decide how much to learn and when to stop end up with systematically biased evidence that reinforces their existing belief bias. In contrast, when people acquire similar amounts of information but without the ability to control when to stop, their initial belief bias is significantly reduced after learning. The results identify a novel channel of persistent belief bias - the endogeneity of information acquisition - and point to exogenously structured learning as a powerful tool for reducing belief-based discrimination and designing effective information interventions.

JEL codes: C91, D90, J71

*Department of Economics, University of California, Davis. I am grateful to Anujit Chakraborty, Scott Carrell, Andrés Carvajal, Michal Kurlaender, and Arman Rezaee for their continued support and guidance throughout my graduate studies. This paper has benefited from discussions with Marianne Bitler, Brad Barber, Burkhard Schipper, Kevin Dinh, Kalyani Chaudhuri, Remy Beauregard, Christoph Schlom, Paco Martorell, Noah Thoron, Paul Feldman, Tony Fan, Billur Aksoy, Anindo Sarkar, and seminar participants at the Economic Science Association, Bay Area Behavioral & Experimental Economics Workshop, Western Economic Association International Conference, AEA CSQIEP seminars, UC Davis Theory & Behavioral seminars, and UC Davis Applied Micro seminars, among many others. I gratefully acknowledge research funding from the Department of Economics at UC Davis. The experiments in this paper were approved by the Institutional Review Board at UC Davis and were pre-registered at AsPredicted. All errors are my own. Email: ucsun@ucdavis.edu, Website: https://christinasun.net/.

# 1 Introduction

We often form beliefs about others under limited and costly information, through sequential encounters, interviews, and observations. In labor markets, for example, employers decide how much to learn and candidates and when to stop, acquiring information about through resumes, interviews, or referrals. Over time, they also form beliefs about different demographic groups based on repeated exposure to applicants from those groups. Such beliefs are an important driver of labor market discrimination (e.g. Phelps, 1972; Arrow, 1973), alongside preference-based mechanisms (Becker, 1971). Recent research highlights the role of *biased beliefs* in reinforcing discrimination (Bohren et al., 2020). However, despite increasing awareness and policy efforts, such belief bias remain remarkably persistent, raising a central question: why do biased beliefs endure even when information is abundant?

In this paper, I propose and experimentally test a novel channel in the persistence of biased beliefs: the endogeneity of the information acquisition process. Through a combination of theory and experiment, I show that even when accurate information is available, individuals who can decide how much to learn and when to stop end up with systematically biased evidence that reinforces their existing belief bias. In contrast, when people acquire similar amounts of information but without the ability to control when to stop, their initial belief bias is significantly reduced after learning. This difference arises because when people have endogenous control over information search, they are more likely to stop when the information aligns with their prior beliefs, resulting in systematically distorted information sets that sustain existing bias. Unlike previous accounts that focus on misinterpretation or cognitive biases, my approach highlights how distorted beliefs can emerge upstream from the way people direct their own search for information.

A robust literature in psychology and economics has documented that the mode through which people learn about risky prospects leads to divergent choice patterns. When people learn through sequential sampling rather than explicit description, their decisions systematically diverge, a phenomenon know as the "Description-Experience Gap" (Hertwig and Erev, 2009). Recently, Oprea and Vieider (2024) proposed a unifying account rooted in cognitive noise: since noisy coding interacts differently with described vs. experienced probabilities, the two learning modes produce different degrees of Bayesian compression of posterior beliefs towards priors. A natural implication is that the structure of information acquisition may have broader implications for economic be-

havior beyond risky choice, since belief are a key driver in economic decision-making. This paper investigates one such dimension. Within experience-based learning, I ask whether endogenous control over the sampling process – the ability to decide when to stop – generates additional belief distortions relative to exogenous experience. Where the noisy cognition literature points to the noisy encoding of observed information, I identify a complementary, upstream channel: when information acquisition is endogenous, realization-based stopping skews the signals that are observed in the first place, producing distortion towards priors. Both channels generate similar distortion of posterior towards the priors, but operate at different stages of the belief formation process and call for different interventions.

Guided by a theoretical framework of sequential information search by Bayesian agents who derive utility from cognitive consistency, I design and implement a controlled laboratory experiment that isolates the causal impact of different information acquisition modes on belief formation and discriminatory behavior. I set up the experiment using a simulated hiring environment, where participants in the role of "employers" evaluate "workers" from Asian and Hispanic groups and make wage offers based on perceived productivity. This setting allows me to directly elicit employer beliefs and measure discrimination using their wage offers. To operationalize productivity, I use a 12-question math test taken by participants in the role of "workers". Using a math task induces natural belief bias while offering an objective, quantifiable measure of worker productivity. Critically, I construct the Asian and Hispanic worker groups to have identical score distributions, providing a clear "ground truth" benchmark. Any systematic differences in employer beliefs or wage offers between groups thus reflect biased beliefs or preferences, rather than any underlying differences in worker performance. I employ a fully between-subject design, where each employer is randomly assigned to evaluate only one worker group, thereby reducing potential confounds.

Subjects are randomly assigned to one of three information treatments: **Voluntary**, where employers engage in fully endogenous information search and can decide whether to stop or continue after observing each draw; **Exogenous Mean**, where endogenous control over the search process is removed, and employers must draw a fixed number of random workers equal to the average number drawn in Voluntary[1]; and **Exogenous 20**, where information search is also exogenous, but with a fixed sample size of 20 workers. In all treatments, employers sequentially draw random workers from their assigned worker group and observe their individual productivity in the form of score bins (1–3, 4–6, 7–9, or

[1] 10 draws for employers with Hispanic workers, 8 for those with Asian workers

10–12). This design mirrors real-world settings where information is incomplete and open to interpretation. By preserving some "wiggle room" in how employers read signals, it allows us to compare which modes of information acquisition yield more accurate beliefs despite that discretion. Comparing employer beliefs in Voluntary against Exogenous Mean isolates the role of the endogeneity of information search while holding average information content the same, whereas Exogenous 20 doubles the amount of information acquired in Exogenous Mean, testing the effect of a larger and more representative sample.

To measure belief bias, I elicit incentivized employer beliefs about the productivity distribution of their worker group before and after information acquisition. In the final stage of the experiment, employers make a series of incentivized wage offers to 10 randomly drawn individual workers. Because worker groups are statistically identical, any differences in wage offers reflect either biased beliefs or discriminatory preferences. In addition to the three main treatments, I implement a *Baseline* treatment where employers report prior beliefs and proceed directly to the wage offer task. This treatment measures the extent of discriminatory behavior in the absence of new information. Full details of the experimental design are provided in Section 2.

The findings from my experiment show that the mode of information acquisition substantially affects both belief bias and discrimination. First, I find that employers hold significantly biased prior beliefs against Hispanic workers. In the Baseline condition, where employers receive no information about worker performance, those evaluating the Hispanic worker group systematically underestimate the group average score relative to those evaluating the Asian worker group. Further, this biased belief directly translates into wage discrimination, with Hispanic workers receiving significantly lower average wage offers. These baseline patterns underscore the role of inaccurate beliefs as a mechanism for driving discrimination.

Starkly, belief bias persists even when employers can endogenously acquire additional information in the *Voluntary* condition. Employers revise their beliefs toward greater accuracy for both Asian and Hispanic workers after sampling, yet a substantial gap remains: updated beliefs about average Hispanic worker performance are still significantly lower than those for Asian workers. This persistence of bias is not simply due to limited information. In the *Exogenous Mean* treatment, where employers are exogenously assigned to sample the average number of workers drawn by their counterparts evaluating the same worker group in the Voluntary condition, belief bias is significantly reduced, by 56% compared to Voluntary. This shows that endogenous control over the information search process, rather than information quantity, sustains most of the belief gap. The *Ex-*

*ogenous 20* treatment provides a high-information benchmark: with a large, fixed sample of 20 workers, posterior beliefs about average performance for both Asian and Hispanic workers converge, and belief bias is effectively eliminated. These belief patterns carry through to behavior: employers in Voluntary offer significantly lower wages to Hispanic workers, while in both Exogenous Mean and Exogenous 20, wage discrimination vanishes alongside belief bias.

My findings, discussed in detail in Section 5 demonstrate that removing endogeneity of information acquisition, even while holding the amount of information constant, can correct distorted beliefs and meaningfully reduce discriminatory behavior. The next part of the paper aims disentangle the mechanisms that drive this effect.

In Section 6, I examine *why* endogenous information acquisition produces systematically biased beliefs. Two mechanisms can explain why bias persists in Voluntary but not in Exogenous Mean. The first is *biased interpretation*: employers who have endogenous control over information search may interpret the same signals in a more biased way than agents without control. The second is *biased information sets*: control over stopping skews the realized samples - agents tend to stop at sequences that disadvantage Hispanic workers - so the evidence they observe is systematically more biased than under exogenous stopping.

I design additional experimental treatments to separate these two channels. The **Matched Sample** treatment isolates biased interpretation by forcing employers to view the same signal sequences generated in Voluntary, but removing control over information search. I find that belief bias in Matched Sample remain as large as Voluntary, indicating that agency over the information search process does not significantly change how individuals interpret information. Therefore, biased interpretation does not play a major role, and it is instead the content of the information produced by endogenous search that is biased.

There are two key features of endogenous information acquisition: realization-based stopping and endogenous sample size. realization-based stopping allows employers to decide whether to continue or halt sampling after each signal, which can systematically skew the information set if search is terminated on confirmatory evidence. Endogenous sample size selection determines how much information to acquire in total; when too few signals are drawn, it can lead to high sampling variation and more scope for biased posteriors. To identify which feature leads to biased information sets, I introduce the **Commitment** treatment. In this treatment, employers commit to a sample size ex-ante before observing any signals, and they then must draw that number of workers. This

creates a middle ground between fully endogenous and fully exogenous search: realization-based stopping is removed, while agency over sample size remains.

I find that belief bias in Commitment is reduced significantly to a level similar to Exogenous Mean, even though the number of signals drawn are similar to that in Voluntary. This result supports the interpretation that biased information sets, shaped by realization-based stopping, are a key driver of belief bias. This reinforces the importance of addressing biased sampling processes as a mechanism for reducing belief-based discrimination.

This paper contributes to the growing literature on discrimination driven by biased beliefs (Bohren et al., 2019, 2020; Campos-Mercade and Mengel Friederike, 2022; Eyting, 2022; Coffman et al., 2019) by identifying a novel source of persistent belief bias: the mode of information acquisition. Classic models in economics distinguish between taste-based discrimination driven by preferences (Becker et al., 1964) and statistical discrimination driven by accurate group-level beliefs (Arrow, 1973; Phelps, 1972). However, recent work shows that inaccurate beliefs often underlie discriminatory behavior (Bohren et al., 2020), and understanding why such beliefs persist despite access to information remains a central challenge. This paper offers a new explanation: beliefs remain biased not necessarily because individuals lack access to corrective evidence, but because they endogenously shape the evidence they encounter. I show experimentally that even when individuals receive similar amounts of information about group performance, their beliefs and hiring decisions differ sharply depending on whether that information was acquired voluntarily or exogenously. Biased beliefs about Hispanic workers persist under voluntary experience, but are eliminated when the same information is provided through an exogenously assigned sample. This demonstrates that the structure of information acquisition, beyond the amount of information, fundamentally shapes belief formation and downstream decisions. My results also help explain why biased beliefs can persist even in environments rich with information: because information acquisition in everyday life is fully endogenous, individuals tend to stop prematurely when evidence aligns with their prior beliefs. As a result, even when the information source is unbiased, the endogenous nature of learning produces skewed belief updates that reinforce existing views.

I also contribute to the literature on biased belief updating and information processing. Prior research shows that individuals often interpret information in self-serving ways, overweighting confirming evidence and underweighting contradictory signals (Mobius et al., 2014; Coutts, 2019; Zimmermann, 2020). I extend this literature by showing that biases in belief formation can arise even when individuals interpret signals in a

fully Bayesian manner—simply due to the endogenous nature of information acquisition. I develop a theoretical framework in which employers update beliefs rationally based on sequential signals but choose when to stop sampling. This framework identifies realization-based stopping as a key source of belief distortion: when individuals can decide in real time whether to continue or stop, they are more likely to stop after observing signals that confirm prior expectations, resulting in posterior beliefs that are systematically skewed toward those priors. Importantly, the theory predicts that even when the sample size is held constant, eliminating control over when to stop (i.e., removing endogeneity) is sufficient to reduce belief bias. These predictions are supported by the experimental data, where treatments that restrict stopping flexibility lead to more accurate beliefs and less biased behavior.

This paper also connects to the growing literature on noisy cognition and internal processing constraints in economics, which shows that many behavioral anomalies can arise as Bayesian response to imprecise internal representations of numerical information. Foundational work by Woodford (2012) proposes that key features of prospect theory can emerge from efficient internal coding under capacity constraints, and Khaw et al. (2021) show that a single imprecision parameter can generate small-stakes risk aversion and the inverse-S pattern of probability weighting through Bayesian compression of noisy signals towards the prior. In the context of belief updating, Bohren et al. (2024) study how cognitive constraints such as limited memory and salience-channeled attention shape belief formation through sequential processing of signals, a setup directly related to evaluation contexts. These existing research identify the source of belief compression towards priors as the noisy encoding of signals. This paper identifies a complementary, upstream mechanism: distortion towards priors can arise also from information acquisition – when agents engage in endogenous search for information, realization-based stopping can lead to asymmetric stopping when signals confirms their prior, skewing the observed evidence itself. Both channels produce similar distortions of posterior belief towards the prior, but at different stages of belief formation.

Finally, this paper makes a novel contribution to the literature on discrimination by designing and empirically validating an intervention that eliminates belief-based discrimination through structural changes in the learning process. While a large body of work has tested information-based interventions aimed at reducing prejudice and discrimination (Wozniak and Macneill, 2018; Alesina et al., 2023; Pedersen and Nielsen, 2024; Kuziemko et al., 2015; Bohren et al., 2020; Alesina et al., 2024; Haaland and Roth, 2019, 2020), results are often mixed. I provide evidence that bias persists not just because of insuf-

ficient information, but because individuals control how and when to acquire it. In my experiment, when participants sequentially sample signals with full autonomy (Voluntary), belief distortions and discriminatory wage offers emerge. However, structuring the learning process through exogenous assignment of a representative sample completely eliminates both belief bias and wage discrimination, even when the information content is held constant. This identifies a powerful and underutilized intervention strategy: rather than simply increasing exposure to information, design learning environments that constrain realization-based stopping and motivated sampling.

The rest of the paper is structured as follows: I first describe the details of the experimental design in Section 2. Then I develop a simple theoretical model of belief formation under sequential information search in Section 3, in order to provide a conceptual underpinning of belief updating under different information acquisition modes. Section 4 formulates hypotheses and discusses empirical strategies to test them, and Section 5 presents my main empirical results. In Section 6, I present additional experimental treatments that test for underlying mechanisms. Finally, Section 7 concludes and discusses potential policy implications.

# 2 Experimental Design

I design my experiment to answer three research questions. First, what are the impacts of different information acquisition modes - endogenous search and exogenous search - on employer beliefs about worker productivity? Second, how do people search for and interpret information when they have full control over information seeking, and how does it affect belief updating? And third, do these modes of learning meaningfully affect discriminatory behavior against minority workers, and to what degree?

In order to answer these questions, I set up a controlled laboratory experiment to simulate a real-world labor market. I recruit two separate groups of subjects for my experiment, "workers" and "employers". The "workers" are an ethnicity-balanced sample of Asian and Hispanic subjects who take an incentivized math test with 12 questions, and receive a score equal to the number of questions they answered correctly. I also collect demographic information from the workers to construct worker "resumes" used in the employer hiring task. In order to avoid cross-group comparisons and reduce social desirability bias, I randomly match "employers" with one worker group (Asian or Hispanic). Employers are informed about the calculation of the worker test scores and shown example questions to provide a basis for prior belief formation. I further ran-

domize each employer into one of four information conditions and one control condition. Therefore, there are 2 orthogonal treatment axes (ethnicity and information mode), with 2 and 5 treatment arms in each axis, respectively. This constitutes a $2 \times 5$ full factorial between-subject design. Employers in the information conditions go through four task stages:

1. **Prior Belief Elicitation:** I elicit employer beliefs about the average score as well as the full distribution of scores in their assigned worker group. This first stage is identical for all employers, which ensures prior beliefs are equally distributed across treatments.

2. **Information Acquisition:** I randomize each employer into one of three information treatments: Voluntary, Exogenous Mean, and Exogenous 20. In all conditions, employers learn about the score distribution within their assigned worker group, but I experimentally manipulate the mode of information acquisition. In Voluntary, employers sequentially sample individual workers by clicking a button, with each draw revealing which of the four score bins (1–3, 4–6, 7–9, 10–12) this worker falls into. Information in Exogenous is acquired through the same button-click interface, but with an exogenously assigned, fixed number of draws.

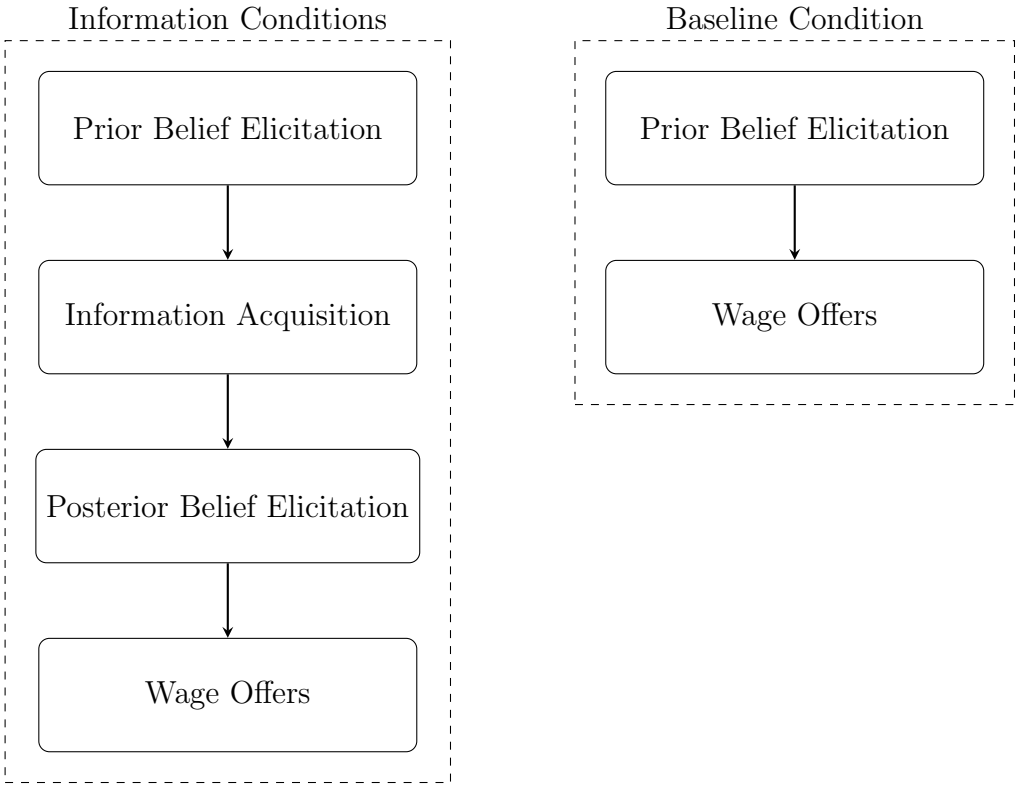
3. **Posterior Belief Elicitation:** After employer learn about their worker group, I elicit updated group-level beliefs about the worker score distribution.

4. **Wage Offers:** Finally, employers sequentially make wage offers to a series of 6 randomly selected workers from their worker group. I construct a worker profile for each worker and elicit the employers' Willingness to Pay (WTP) for each worker in an incentive compatible way using the BDM value elicitation mechanism (Becker et al., 1964).

Employers randomized into the Baseline control condition do not receive any information about their worker group and therefore do not report updated beliefs. They only go through two task stages: Prior Belief Elicitation and Wage Offers. This provides a no-information benchmark to study the baseline levels of belief bias and wage discrimination. This flow of tasks is summarized graphically in Figure 1.

With the core structure of the experiment in place, I now turn to technical details of the experimental design. In what follows, I describe the worker tasks and the employer experiment, and how each treatment condition in the employer experiment is implemented. I also detail the design of the employer task, how I measure employer

Figure 1: Order of tasks in the employer experiment



beliefs and discrimination, and the incentive structure. I conclude this section with the implementation procedures.

## 2.1 Worker Survey

The worker survey serves two primary purposes: to obtain an objective measure of productivity and to construct the worker groups used in the employer experiment. I recruit participants who self-identify as Asian or Hispanic to take on the role of "workers". Each worker completes an incentivized 12-question math test under a six-minute time limit. The questions are adapted from SAT Math questions and workers earn a \$0.2 bonus per correct answer, in addition to a \$2 participation payment. After the time expires, the test is automatically submitted, the workers are shown their score and corresponding bonus payment. Finally, they complete a demographics questionnaire in which I collect their age, gender, race, ethnicity, education, and favorite color. Select demographic characteristics are later used to construct worker resumes shown to the employers.

Utilizing a math test allows me to induce natural belief bias at baseline, since math is a task domain in which Asian workers are stereotyped as higher performing relative to Hispanic workers. A score out of 12 allows me to elicit the entire subjective belief distribution of employers without imposing excessive effort and time costs, which allows identification of the source of discrimination. After collecting data from the worker survey, I construct the two worker groups: an Asian worker group and a Hispanic worker group, each consisting of 100 workers. Importantly, I construct the two groups to have identical test score distributions. This ensures that the two worker groups are in fact equally productive, providing a clear objective benchmark to measure updated beliefs against. By holding actual performance constant across groups, I can rule out accurate statistical discrimination, and distinguish the source of any observed discrimination between taste-based discrimination and discrimination driven by biased beliefs.

## 2.2 Employer Experiment

I recruit a separate group of participants for the employer experiment. Upon entering the experiment, participants are informed that they will be playing the role of employers and make evaluation and hiring decisions about some workers. I inform the employer that they have been matched with a worker group - Asian or Hispanic - consisting of 100 Prolific workers based in the United States. This between-subject design ensures that employers never engage in explicit cross-group comparison, reducing the impact of other behavioral

factors that can alter behavior, such as social desirability bias. Specifically, employer know that each worker completed a timed SAT-level math test with 12 questions, and that each worker received a score equal to the number of correct answers. I also inform them that no worker received a score of 0. This gives a common support of integers from 1 to 12 for the employer belief distribution. The employers can view sample math questions from the worker survey which aids in their initial belief formation.

### 2.2.1 Prior Belief Elicitation

After employers learn about the experiment setup, I elicit their initial beliefs about the score distribution in their worker group. Specifically, I collect three incentivized belief measures: the employer's entire subjective belief distribution of worker score (the probability mass function), and their belief of the average worker score in the group. Since belief-based discrimination may be influenced by both the mean and the variance of worker productivity, collecting the entire belief distribution allows me to precisely pin down employer beliefs and disentangle taste-based and belief-driven discrimination. I separately elicit belief on the average worker score because this provides an intuitive belief measure that is easy for employers to understand, and it serves as a robustness check for the validity of the belief measures.

To elicit the employer's subjective beliefs, I ask them to report how many workers they think have each possible score in the group. This involves 12 questions: they report how many workers they think received a score of 1, how many workers received a score of 2, etc. Their answers for these 12 questions must add up to 100. To elicit employer belief on the average worker score in the group, I use a text box where they can enter a positive number less than or equal to 12, up to two decimal points. Figure B.1 and Figure B.2 display the survey interface for belief elicitation, which is identical for both prior and posterior beliefs.

Belief elicitations are incentivized to ensure truthful reporting. At the end of the experiment, one question is randomly selected to be paid a bonus. If a belief question is selected, the employer receives a \$1 bonus [2] if their answer is within $\pm 2$ of the correct answer. To ensure participants understood the structure of these incentives, the experiment included detailed written instructions, annotated screenshots of the elicitation interfaces, and a mandatory comprehension quiz with four questions. Participants could only proceed if they answered all questions correctly within two attempts. Section 2.3

[2] Bonus is calculated in experimental tokens with an exchange rate of 10 tokens to \$1 USD. The bonus for belief elicitation is 10 tokens.

provides details on the incentive structure in this experiment.

### 2.2.2 Information Acquisition

After reporting their prior beliefs about group-level productivity, each employer is randomly assigned to one of three information conditions: Voluntary, Exogenous Mean, Exogenous 20, and a no-information benchmark, the Baseline condition. In the information treatments, all employers receive data about the distribution of worker scores within their assigned group, but the mode of information delivery varies systematically across conditions. This variation allows me to isolate the effects of different information environments on belief updating and hiring behavior. I describe each condition in detail below.

**Voluntary.** In the Voluntary condition, I study the impact of experience-based learning when employers can endogenously seek out information about workers. Employers in this treatment are presented with a button they can click to sequentially sample individual workers from their assigned group, without replacement. Each click triggers the computer to randomly select a worker and display that worker's score range (1–3, 4–6, 7–9, and 10–12), rather than the exact score. Data on worker productivity distribution in all information conditions are delivered in the same binned format. This choice serves two purposes. First, it introduces a degree of ambiguity into the information, giving employers wiggle room in forming their beliefs. Rather than seeing precise averages or raw scores, they must mentally interpolate and reason about what the data implies—closer to how information is processed in real-world decision-making. Second, binning mimics how data is often summarized in practice. Outside of lab settings, individuals rarely encounter complete information about underlying distributions; instead, they work with aggregated, approximate data.

To encourage reflection and preserve the sequential structure of experience-based information acquisition, I implement a 3-second cool-down between clicks. During this period, the draw button is temporarily disabled, mimicking the temporal cost of acquiring real-world information and discouraging rapid, thoughtless clicking to accumulate the sample and viewing them simultaneously. In addition, each new worker's score range is appended below the previous draws, creating a cumulative list on screen. This design feature minimizes memory constraints and ensures that employers have access to the full history of their observations throughout the task. Figure B.3 shows a screenshot of the button interface, which is identical through all three experience conditions.

Employers can draw as many workers as they like (including 0 and up to 100), allowing me to observe not only how they update their beliefs in response to experience, but also how they choose to seek out information when the process is completely endogenous. This condition captures the dynamics of endogenous information acquisition, where individuals may selectively seek out certain outcomes—potentially reinforcing prior beliefs or biases.

**Exogenous Mean.** In the Exogenous Mean condition, employers sequentially sample workers without replacement from their assigned group by clicking a button. The interface is identical to the Voluntary treatment arm: each click reveals a randomly selected worker's score bin (1–3, 4–6, 7–9, or 10–12), and the new observation is appended below the previous ones to create a visible sampling history. Unlike in the Voluntary condition, however, employers in this treatment are required to draw exactly the average number of workers drawn by the employers with the same worker group in the Voluntary condition. That is, employers in the Exogenous Mean-Asian treatment draw the mean sample size from Voluntary-Asian employers (8 workers), and Exogenous Mean-Hispanic employers drawn the mean sample size from Voluntary-Hispanic employers (10 workers). This eliminates the endogeneity of information acquisition by fixing the sample size, making it comparable to the voluntary condition while holding equal the information format and average information content. All other aspects of the task—including the interface, timing constraints, and binned feedback format—are identical to the Voluntary treatment.

**Exogenous 20 (20 Draws).** The Exogenous 20 condition follows the same structure as the Exogenous Mean treatment, with one key difference: employers are required to sample 20 workers, a larger and more representative sample. Sampling is again done without replacement, and the information is displayed using the same button-click interface and binned score format as in all other experience conditions. Each new draw is appended below the previous ones, allowing employers to view their full sampling history and eliminating memory constraints. All other aspects of the design remain identical to the Voluntary and Forced Benchmark conditions. This treatment allows me to examine whether reducing sampling bias by providing a larger sample of exogenously assigned experiential data leads to more accurate belief updating and less discriminatory behavior.

**Baseline.** The Baseline condition serves as a control group in which employers receive no information about the productivity distribution of the worker group they are assigned to evaluate. After reporting their prior beliefs, they proceed directly to the wage offer tasks without any opportunity to learn about the workers. This condition allows me to measure the extent of initial belief bias and discriminatory behavior in the absence of any information provision.

### 2.2.3 Posterior Belief Elicitation

After the information acquisition phase, I elicit posterior beliefs from employers in all information treatments. The belief elicitation mirrors the prior belief stage and uses the same set of questions, allowing for direct measurement of belief updating. To capture each employer's subjective belief distribution, I ask them to estimate how many workers in the group received each possible score from 1 to 12. This consists of 12 separate questions—one for each score value—and employers are instructed that their responses must sum to 100, reflecting the total number of workers in the group. In addition to the full distribution, I elicit a point estimate of the average worker score. Employers report this value using a text input box, where they can enter a number between 0 and 12, rounded to two decimal places.

To minimize memory constraints, employers have full access to the information they received during the learning phase while reporting their posterior beliefs. The complete history of sampled draws is displayed in order, allowing employers to refer back to the observed data. The Baseline condition skips this step, as employers receive no new information prior to making hiring decisions.

### 2.2.4 Wage Offers

After employers report their updated group-level beliefs, they enter the wage offer stage, where I elicit their willingness to pay for individual workers. Each employer sees 10 randomly selected workers from their matched group of 100, presented sequentially. I present a short "resume" for each worker, which includes an avatar, anonymized nickname, country of residence (United States), gender, age (18–45 or 46+), and favorite color. Ethnicity is conveyed subtly through the avatar and nickname to minimize social desirability bias. Each worker group includes 90 ethnically representative names (e.g., Mateo for Hispanic workers, Ming for Asian workers) and 10 white-sounding names (e.g., Mark), reflecting the real-world practice of ethnic minorities sometimes adopting Anglicized names. Figure B.4 shows example profiles for a Hispanic and an Asian worker, respectively. These profiles also allow me to control for observable worker characteristics in the analysis of wage offers.

To elicit willingness to pay in an incentive compatible manner, I implement the BDM elicitation mechanism (Becker et al., 1964), which is incentive compatible for arbitrary risk preferences. For each worker, employers are endowed with 12 experimental tokens as their hiring budget, and they are asked to make an integer wage offer between 0

and 12. After they make the wage offer, the computer randomly generates an asking wage for the worker, also an integer between 0 and 12. If the asking wage is less than or equal to the wage offer, the employer hires the worker at the asking wage, and earns (12 − asking wage+actual worker score tokens. If the asking wage is higher than the wage offer, the employer does not hire the worker and keep the 12 tokens. Importantly, employers are informed that their decisions do not affect the compensation of the workers they evaluate. This removes the influence of interpersonal preferences or prosocial considerations from the decision environment, and ensures that wage offers reflect only beliefs about worker productivity and preference-driven bias. Figure B.5 displays the interface for the wage offer.

I implement several design features to ensure comprehension and data quality. First, the instructions for the BDM mechanism is framed in an intuitive, easily understandable manner, using the comparison between asking wage and wage offer to determine hiring outcomes. Second, employers receive detailed instructions and must correctly calculate their payment in 2 hypothetical scenarios in the comprehension quiz in order to proceed with the experiment. Because worker groups have identical productivity distributions, any systematic difference in wage offers between the Asian and Hispanic groups reflects discriminatory behavior, rather than differences in actual performance.

After completing the wage offer stage, employers answer a brief demographics questionnaire. I include an open-ended question—"What do you think this experiment is trying to study?"—to assess potential social desirability bias and the degree to which participants infer the study's purpose. I utilize supervised machine learning to code this question and classify respondents by whether they correctly inferred the true study purpose. As a robustness check, I exclude participants who were able to infer the true purpose of the experiment and test whether results are robust to this sample restriction.

## 2.3 Incentives

The experiment was carefully designed to incentivize both accurate beliefs and realistic hiring behavior. At the end of the study, one decision was randomly selected for bonus payment to ensure that participants took each decision seriously. This Random Problem Selection (RPS) incentive scheme has been shown to incentive compatible for experiments with multiple tasks (Azrieli et al., 2018).

If a belief elicitation question was selected, participants received 10 experimental tokens for bonus (the experimental tokens have a conversion rate of 10 tokens = \$1

USD) if their reported belief about the average score or distribution was within a $\pm 2$ margin of the correct value. This scoring rule is incentive-compatible and encourages truthful reporting of beliefs under risk neutrality, following standard practice in belief elicitation.

If a wage offer decision was selected, the bonus depended on the realized hiring outcome and the true productivity score of the matched worker, implemented using the Becker–DeGroot–Marschak (BDM) mechanism. For each worker, employers are endowed with 12 experimental tokens as their hiring budget, and they are asked to make an integer wage offer between 0 and 12. After they make the wage offer, the computer randomly generates an asking wage for the worker, also an integer between 0 and 12. If the asking wage is less than or equal to the wage offer, the employer hires the worker at the asking wage, and earns (12 − asking wage + actual worker score tokens. If the asking wage is higher than the wage offer, the employer does not hire the worker and keep the 12 tokens (i.e. the possible amounts of bonus an employer can earn in a given wage offer ranges from \$0.1 to \$2.3).

To ensure participants understood the structure of these incentives, the experiment included detailed written instructions, annotated screenshots of the elicitation interfaces, and a mandatory comprehension quiz with four questions. Participants could only proceed if they answered all questions correctly within two attempts. To further promote careful reading, participants received a \$1 bonus for answering all questions correctly on the first try. Approximately 66% of participants earned this bonus, indicating a high level of engagement and comprehension. This supports the validity of the responses and high quality of the experimental data.

## 2.4 Implementation

The experiment was approved by the UC Davis Institutional Review Board and pre-registered on AsPredicted (#213,758). The Exogenous Mean condition, which uses the average number of draws from the Voluntary condition as a design parameter, was conducted later and pre-registered separately as a follow-up study (#237,319).

All surveys were programmed in Qualtrics using custom JavaScript to implement dynamic interfaces. I recruited participants for both the worker survey and employer experiment via the online platform Prolific, which has been shown to produce data quality comparable to traditional in-person lab experiments and superior to other online platforms such as Amazon Mechanical Turk (Gupta et al., 2021; Peer et al., 2022). Pro-

lific enforces strict privacy protections and prohibits the collection of personally identifiable information, helping reduce potential experimenter demand and social desirability bias—especially important in studies of sensitive topics like discrimination. Additionally, recent evidence suggests that demand effects in online experiments are relatively small, and even when participants infer the purpose of a study, their behavior is largely unaffected (De Quidt et al., 2018; Mummolo and Peterson, 2019).

All participants were at least 18 years old and residents of the United States. For the worker survey, I used Prolific's screening tools to recruit participants who self-identify as either Asian or Hispanic. Workers received a $2 base participation payment and earned a bonus of $0.20 for each correct answer on the math test. Employers were paid $2 for completing the experiment, with an additional bonus ranging from $0 to $3.4. The average time taken to complete the survey for employers was 15 minutes, with an average total compensation of around $3.72. resulting in an hourly compensation of approximately $15. The final sample includes approximately 400 workers and 2,400 employers, generating rich data on belief formation, sampling behavior, and wage offers.

# 3 Theoretical Framework

This section presents a model of belief updating under sequential information acquisition, in which agents weigh two objectives: instrumental utility—making better decisions by forming more accurate beliefs—and belief utility—a psychological benefit from maintaining cognitive consistency. While both exogenous and endogenous experience provide signals drawn from the same distribution, the key distinction lies in the agent's control over the information search process. In the endogenous regime, agents can choose when to stop acquiring information based on their evolving beliefs, creating scope for distorted inference due to belief utility.

To capture this, I formalize three distinct modes of learning that vary in the degree of control agents have over the information search process:

1. Sequential Search with Optional Stopping (Voluntary): Agents observe signals and choose at each step whether to continue.

2. Exogenous Sample Size Assignment (Exogenous): Agents receive a fixed, exogenously assigned number of signals, independent of their beliefs or preferences.

3. Ex Ante Commitment to Sample Size: Agents choose the number of signals in

advance and are then required to draw that number of signals.

In the Voluntary learning mode, information acquisition is fully endogenous. Agents can not only decide how many signals to draw, but also make stopping decisions based on signal observations. In contrast, Exogenous learning removes all endogenous control over the process. Ex-Ante Commitment represents a middle ground: agents can decide on the number of signals, but are not able to condition stopping decisions on sample realizations. In particular, Voluntary and Ex-Ante Commitment are equivalent for a fully Bayesian agent but yield different predictions when belief utility is present.

I begin with a baseline model of a fully Bayesian agent, characterizing optimal stopping and belief updating under instrumental utility alone. This benchmark illustrates how information acquisition and belief evolution proceed when agents update objectively and terminate search based solely on marginal gains from information. I then extend the model to incorporate belief utility in the spirit of cognitive consistency preferences (e.g. Yariv, 2001), where agents experience disutility from shifting too far from their current beliefs. Under belief utility, search decisions become dependent on the realized signal path, leading to earlier stopping times, producing skewed information sets and stickier beliefs.

This framework explains how two agents with access to the same average amount of information may arrive at systematically different posteriors depending on whether they chose that information themselves. The model yields testable predictions about belief bias and behavior across the three learning regimes, which guide the empirical analysis that follows.

## 3.1 Setup

I consider a world with a continuum of true states $\theta \in \mathbb{R}$ and discrete time. In the setting of my experiment, we can think of $\theta$ as the actual average score of a worker group. The state of the world is drawn from a known prior distribution $f(\theta)$ at time $t = 0$. A decision maker faces a sequential information acquisition problem in discrete time, indexed by $t = 1, 2, 3, ....$ While she does not observe the true state $\theta$, she knows that the true state is draw from the prior distribution $f(\theta)$.

In the each period, the agent may draw a signal $x_t \sim h(x|\theta)$, where signals $x_t$ are independently distributed across time conditional on the true state $\theta$. After observing any history of signals $\{x_1, x_2, \ldots, x_t\}$, the agent updates her beliefs using Bayes' Rule to form posterior belief distribution $f_t(\theta|x_1, x_2, ..., x_t)$. Each draw incurs a fixed cost $c \geq 0$.

To simplify the problem, I consider a standard model with true state drawn from a normal distribution $\mathcal{N}(\mu_0, \sigma_0^2)$. This means that the agent's prior belief distribution at time $t = 0$ is:

$$p_0(\theta) \sim \mathcal{N}(\mu_0, \sigma_0^2).$$

Conditional on the true state, the signal also follows a normal distribution: $x_t \sim \mathcal{N}(\theta, \sigma_d^2)$, where $\sigma_d^2$ denotes the variance of the signal distribution centered around the true state.

## 3.2 Belief Updating

After observing a sequence of $t$ signals, the agent updates her beliefs using Bayes' Rule. Denote the dataset of $t$ signals as $D_t = \{x_1, x_2, ..., x_t\}$, then we can calculate the posterior belief in period $t$ as

$$p_t(\theta|D_t) = \frac{p(D_t|\theta)p(\theta)}{p(D_t)}$$

After algebra, it can be shown that the agent's posterior belief distribution is also normally distributed up to a normalization constant $p(D_t)$. That is,

$$p_t(\theta|D_t) \propto \mathcal{N}(\mu_t, \sigma_t^2) \tag{1}$$

where

$$\sigma_t^2 = \frac{1}{\frac{t}{\sigma_d^2} + \frac{1}{\sigma_0^2}} \tag{2}$$

and

$$\mu_t = \sigma_t^2 \left( \frac{t\bar{x}}{\sigma_d^2} + \frac{\mu_0}{\sigma_0^2} \right). \tag{3}$$

In the expression for $\mu_t$, $\bar{x}$ is the sample mean of the signals, i.e. $\bar{x} = \frac{\sum_{i=1}^{t} x_i}{t}$. The mean of the posterior distribution can also be written as a weighted combination of the sample mean and the prior mean:

$$\mu_t = \omega_t \bar{x} + (1 - \omega_t)\mu_0 \tag{4}$$

where $\omega_t = \frac{t\sigma_t^2}{\sigma_d^2}$. It is straightforward to show that $\lim_{t\to\infty} \omega_t = 1$, meaning as the sample size approaches infinity, the posterior mean approaches the sample mean. This reflects the fact that as the sample size gets larger, the agent places more weight on the sample mean relative to the prior mean, which is intuitive.

We also can see that[3]

$$\mathbb{E}(\mu_{t+1}|\mu_t) = \mu_t, \tag{5}$$

i.e. the posterior mean is a martingale.

## 3.3 The Rational Employer

Here I describe the decision problem of a fully rational employer, who only derives instrumental utility from beliefs, and therefore only cares about belief accuracy. I then discuss the rational employer's optimal information search under the three distinct learning modes: fully endogenous sequential search, exogenously assigned sample size, and ex-ante commitment to sample size.

### 3.3.1 The Sequential Problem

Under fully endogenous sequential information search, the agent faces a discrete time optimal stopping problem. In each period, the agent draws a signal, updates her beliefs, and decides whether to stop or continue. If she stops, she cannot draw another signal. Suppose the agent decides to stop in period $\tau$, then she reports posterior mean $\mu_\tau$ and earns instantaneous instrumental utility $u$:

$$u(\mu_\tau, \theta) = \begin{cases} 1 & \text{if } |\mu_\tau - \theta| \leq b \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The instrumental utility of beliefs comes from guessing the true state $\theta$ sufficiently accurately: the agent earns 1 util if her guess $\mu_\tau$ is within an error bound $b > 0$ of the true state. If the agent decides to continue, she enters the next period, draws another signal, updates her beliefs, and faces the same stop/continue decision problem again. The problem is finite horizon: $t \leq T$, which captures the constraint in the experiment that subjects cannot sample beyond 100 workers. There is no discounting.

In period $t$, the agent's decision problem is captured by the following Bellman Equation:

$$V(\mu_t) = \max \left\{ \underbrace{u(\mu_t, \theta)}_{\text{Stop}}, \underbrace{\mathbb{E}\big[V(\mu_{t+1}) \,\big|\, \mu_t\big] - c}_{\text{Continue}} \right\} \tag{7}$$

Since the value function has no closed form solution, I will discuss the intuition behind the existence of a stopping time. Essentially, this involves showing that the stopping

[3]Derivation in Appendix A

value is monotonically increasing in t and the continuation value is monotonically decreasing in t. In the Bellman equation, stopping value is the instrumental utility of beliefs represented by Equation 6, which is the probability that the true state $\theta$ falls within $\pm b$ of her current posterior mean $\mu_t$. It can be shown that[4]

$$u(\mu_t, \theta) = 2\Phi\left(\frac{b}{\sigma_t}\right) - 1$$

It is immediately clear that

$$\frac{\partial u}{\partial b} > 0 \tag{8}$$

and

$$\frac{\partial u}{\partial \sigma_t^2} < 0. \tag{9}$$

This is intuitive, since a wider error window increases the probability of $\mu_t$ landing within the error bounds, while a larger posterior variance means more residual uncertainty, thus decreasing the probability of guessing within the error bounds. Notice that from Equation 2, we have $\frac{\partial \sigma_t^2}{\partial t} < 0$, that is, the posterior variance shrinks deterministically with the number of signals. Therefore, the stopping value is increasing with the number of signal draws $t$:

$$\frac{\partial u}{\partial t} > 0 \tag{10}$$

Next, I consider the continuation value. Equation 25 gives the evolution of the state variable, posterior mean $\mu_t$, and we can re-write the continuation value as $C_t = \mathbb{E}[V(\mu_t + \omega_{t+1}\delta_{t+1})|\mu_t] - c$. Since both $\omega_{t+1}$ and the variance of $\delta_{t+1}$ shrink deterministically as t increases, the belief drift $\omega_{t+1}\delta_{t+1}$ decreases as the number of signals $t$ increases, while the belief variance $\sigma_{t+1}$ also shrinks. This means that the marginal value of information is decreasing, and each successive signal moves the belief less. As $t$ increases, the posterior belief distribution becomes tighter, and its mean moves less, thus the expectation $\mathbb{E}[V(\mu_{t+1})|\mu_t]$ approaches $V(\mu_t)$, i.e. the expected gain from drawing another signal decreases over time. Since $V$ is concave[5], we have $V(\mu_t) = V(\mathbb{E}[\mu_{t+1}|\mu_t]) \geq \mathbb{E}[V(\mu_{t+1})|\mu_t]$ by Jensen's Inequality. Further, since $V$ is not affine in $\mu$, and $\mu_{t+1}$ is nondegenerate, the strict inequality holds: $V(\mu_t) > \mathbb{E}[V(\mu_{t+1})|\mu_t]$. Now, by definition, $C_t = \mathbb{E}[V(\mu_{t+1})|\mu_t] - c$ and $C_{t+1} = \mathbb{E}[V(\mu_{t+2})|\mu_t] - c$, so $C_t > C_{t+1}$. Therefore, the continuation value is strictly monotonically decreasing in $t$.

Now we have a strictly increasing sequence $\{u_t\}_{t=1}^T$ and a strictly decreasing sequence $\{C_t\}_{t=1}^T$. With reasonable parameter assumptions such that $u_0 < C_0$ and $u_t > C_t$ as

[4]Full derivation in Appendix A

[5]See Appendix A for a proof

$t \to T$, there exists a unique $t^*$ where the two sequences cross each other. Therefore, the existence of a unique optimal stopping time is guaranteed: there exists a unique

$$t^* = \min\{t \geq 0 : u(\mu_t, \theta) \geq C_t\}.$$

Since the value function is concave, we can attempt to characterize an optimal stopping threshold rule. We re-write the stopping rule as:

$$\textit{Stop when } \Delta_t := \mathbb{E}[V(\mu_{t+1})|\mu_t] - V(\mu_t) \leq c.$$

Notice that the marginal value of an additional signal $\Delta_t$ is the Jensen gap, since $V$ is concave. Therefore, as $\sigma_t^2$ is decreasing in $t$, $\Delta_t$ is also decreasing in $t$. This means that there exists a threshold value of posterior variance $\bar{\sigma}$ such that the agent continues if $\sigma_t^2 > \bar{\sigma}$, and stops if $\sigma_t^2 \leq \bar{\sigma}$.

**Comparative Statics.** Here I discuss the comparative statics of the optimal stopping time with respect to select model parameters.

**Proposition 3.3.1.** *Optimal stopping time of the rational agent increases in prior variance.*

From Equation 2, it follows directly that $\frac{\partial \sigma_t^2}{\partial \sigma_0^2} > 0$, holding the number of signals $t$ fixed. That is, higher prior variance leads to slower learning, and posterior variance decreases more gradually over time. Since the value of stopping, $u(\mu_t)$, is decreasing in posterior variance (i.e., $\frac{\partial u}{\partial \sigma_t^2} < 0$), the agent requires more signals to reach the same expected utility threshold. Therefore, the optimal stopping time $t^*$ increases with $\sigma_0^2$.

**Proposition 3.3.2.** *Optimal stopping time of the rational agent increases with signal noise.*

Higher signal noise reduces the informativeness of each draw, causing posterior variance to decline more slowly. As a result, the agent needs more signals to reach the stopping threshold of posterior variance, so the stopping time increases.

**Proposition 3.3.3.** *Optimal stopping time of the rational agent decreases with signal cost.*

When the cost of drawing a signal increases, the marginal benefit of continued sampling must exceed a higher threshold to justify continuation. Thus, the agent is willing to stop with greater posterior uncertainty, and the stopping threshold is reached sooner. This reduces the optimal stopping time.

### 3.3.2 Ex-Ante Commitment Problem

In the Ex-Ante problem, an agent must first decide on the number of signals $\tau$ she will draw at time $t = 0$, before seeing any signal realizations. Then she must draw $\tau$ signals, that is, she must stop at $t = \tau$. The agent's problem is:

$$V^{commit}(\mu_0) = \max_{\tau \in \mathcal{T}} \mathbb{E}\big[\, u(\mu_\tau) - c\tau \big]. \tag{11}$$

Intuitively, with constant costs and no discounting, the agent's preferences are time consistent. Moreover, the agent's optimal policy does not depend on the actual observed signals, since the rewards are a function of only posterior variance $\sigma_t^2$, which evolves deterministically with $t$. Therefore, the revelation of new information will not shift the optimal policy, and ex-ante commitment and sequential re-optimization yields the same optimal stopping time[6].

### 3.3.3 Exogenously Assigned Sample Size

When the employer is exogenously assigned a sample size, there is no active decision to make. The rational employer simply observes the realization of signals and updates her beliefs according to Bayes' Rule. It is trivially true that for a given sample size, a rational employer's posterior beliefs are equal in expectation under the sequential problem and exogenously assigned sample size.

## 3.4 Cognitive Consistency

Here I extend the model to incorporate an intuitive form of belief-based utility: a desire for cognitive consistency. Specifically, this means the employer will incur a cost from updating away from her prior beliefs. Thus, the employer's utility function now includes two types of indices, instrumental utility and belief utility:

$$U = \underbrace{u(\mu_t, \theta)}_{\text{Instrumental Utility}} + \underbrace{\gamma v(\mu_0, \mu_t)}_{\text{Belief Utility}} \tag{12}$$

where belief utility takes the form

$$v(\mu_0, \mu_t) = -(\mu_0 - \mu_t)^2. \tag{13}$$

The parameter $\gamma \geq 0$ captures the weight the employer places on belief utility relative to instrumental utility. For example, if $\gamma = 0$, the employer is fully rational and care

[6] Proof in Appendix A

only instrumental utility, whereas if $\gamma \to \infty$, the employer will care only about cognitive consistency. The form of instrumental utility is the same, i.e.

$$u(\mu_\tau, \theta) = \begin{cases} 1 & \text{if } |\mu_\tau - \theta| \leq b \\ 0 & \text{otherwise} \end{cases}$$

Intuitively, incorporating cognitive consistency introduces path dependence for the employer's decision-making, since now the utility depends not only on the variance of her belief, but also on its location. This breaks time-consistency of the sequential problem and leads to the divergence of beliefs between the three different learning modes. I now discuss the predictions for these three types of problems.

### 3.4.1 The Sequential Problem Under Cognitive Consistency

In the sequential problem with full endogenous control over information search, the employer observes a signal realization in each period, derives belief utility (cost) from the belief drift, and decides whether to stop or continue. The employer's problem is captured by the following Bellman Equation:

$$\tilde{V}(\mu_t) = \max \left\{ \underbrace{u(\mu_t, \theta)}_{\text{Stop}}, \underbrace{\mathbb{E}[\tilde{V}(\mu_{t+1}) + \gamma v(\mu_0, \mu_{t+1}) | \mu_t] - c}_{\text{Continue}} \right\} \tag{14}$$

This Bellman Equation is similar to that of the rational employer (Equation 7, with one key difference: the continuation value now includes an additional term, the belief utility $v$, reflecting the expected cost from belief drift from continued search. One direct implication is that the agent with belief utility stops earlier in expectation than the fully rational agent, because her continuation value is now lower.

**Proposition 3.4.1.** *In expectation, the agent who desires cognitive consistency stops search earlier than the fully rational agent in the sequential problem.*

For a fully rational employer who only derives instrumental utility from beliefs, her continuation value at time $t$ is

$$C_R(\mu_t) = \mathbb{E}[V(\mu_{t+1}) | \mu_t] - c.$$

For an employer who desires cognitive consistency, her continuation value is

$$C_{Cons}(\mu_t) = \mathbb{E}[\tilde{V}(\mu_{t+1}) - \gamma(\mu_{t+1} - \mu_0)^2 | \mu_t] - c.$$

Since $(\mu_{t+1} - \mu_0)^2 \geq 0$ and $\gamma \geq 0$, then $C_{Cons}(\mu_t) \leq \mathbb{E}[\tilde{V}(\mu_{t+1}] - c \leq C_R(\mu_t)$. Intuitively, when we introduce belief utility, all other aspects of the employer's reward function are identical, with the only difference being she now incurs a cost from belief drift. Therefore continuing to search for information is less attractive at every single period. Now, given the continuation value is lower, it is easy to see that the stopping region is larger, i.e.

$$\mathcal{S}_{Cons} \supseteq \mathcal{S}_R,$$

where $\mathcal{S}_{Cons} = \{\mu : u(\mu, \theta) \geq C_{Cons}(\mu_t)\}$ is the stopping region under belief utility, and $\mathcal{S}_R = \{\mu : u(\mu, \theta) \geq C_R(\mu_t)\}$ is the stopping region under the fully rational model. This means that the agent with belief utility will stop in more belief states, including some of those where the fully rational agent would continue. Therefore, the stopping time $\tau_{Cons}$ is earlier in expectation: $\mathbb{E}[\tau_{Cons}] \leq \mathbb{E}[\tau_R]$, i.e. employers who desire cognitive consistency will acquire less information than the fully Bayesian employer, leading to posterior beliefs that are closer to the priors, i.e. their prior beliefs will be stickier.

### 3.4.2 Ex-Ante Commitment Under Cognitive Consistency

In the Ex-Ante problem, the agent decides on the number of signals to draw before observing any realizations. She solves the following optimization problem:

$$\max_{\tau \leq T} \mathbb{E}_0\big[\, u(\mu_\tau, \theta) + \sum_{t=1}^{\tau} \gamma v(\mu_t, \mu_0) - c\tau\big] \tag{15}$$

This means that ex-ante, the agent must balance the expected instrumental value with the expected total belief cost and sampling cost.

Notice that under the quadratic loss form for belief utility $v(\mu_t, \mu_0) = -(\mu_t - \mu_0)^2$, we can write the total expected future belief (dis)utility as:

$$\mathbb{E}_0 \sum_{t=1}^{\tau} [-\gamma(\mu_t - \mu_0)^2] = -\gamma \sum_{t=1}^{\tau} Var_0(\mu_t) \tag{16}$$

We can thus re-write the ex-ante problem as:

$$\max_{\tau \leq T} \; u(\mu_\tau, \theta) - \gamma \sum_{t=1}^{\tau} Var_0(\mu_t) - \tau c.$$

Comparing this objective function against that of the fully rational agent in the ex-ante problem described in Equation 11, we can see that the extra term for belief utility will

lead to earlier stopping times, since the total psychological cost increases with search time, i.e. the agent incurs more cost from belief drift the longer she searches.

Now compare this with the sequential problem: since the reward function of the agent in the sequential problem is now path dependent, the two problems are no longer equivalent. The key difference between the two problems is that the sequential agent can decide whether to stop based on signal realizations, whereas the commitment agent does not enjoy this flexibility. Therefore, the commitment agent should in expectation plan conservatively and sample less than the sequential agent: $\mathbb{E}[\tau^{commit}] \leq \mathbb{E}[\tau^{sequential}]$.

Further, since the sequential agent can stop based on observed signal realizations, she can stop earlier when belief drift is large early on. This creates an asymmetric stopping behavior: the agent stops when her beliefs are close to the prior, which introduces selection of posterior beliefs on the observed signal direction. However, the commitment agent does not have this flexibility: even when she encounter signal realizations that are particularly dis-confirming, she is constrained by her commitment to the sample size and must update beliefs and incur that extra belief cost. This means that the commitment agent's posterior beliefs are not selected on signal realizations. Therefore, on average, we should expect posterior beliefs that are "stickier" under sequential decision making, compared to the commitment case.

**Proposition 3.4.2.** *Under identical sample sizes, posterior beliefs will be more biased towards prior for the sequential agent than the agent with ex-ante commitment to sample size.*

### 3.4.3 Exogenously Assigned Sample Size Under Cognitive Consistency

When an agent that desires cognitive consistency is exogenously assigned a sample size, she is constrained in her choices as she no longer has the option to make stopping decisions based on signal realizations or set the sample size. Even though she still incurs belief disutility, she must update her beliefs based on observed signals, and thus arrive at posterior beliefs that are identical as a Bayesian agent with the same sequence of realized signals.

A particularly interesting prediction is that even if the sample sizes are held the same, an agent in the exogenous problem will still arrive at more unbiased posteriors than the sequential agent. This is because of selection on observed signal paths in the sequential problem. Even if a sequential agent ends up with the same number of signals as an exogenous agent, her realized signal sequence will on average be skewed towards

her priors, therefore leading to more bias in posterior beliefs.

**Proposition 3.4.3.** *Conditional on the same sample size, an agent in the sequential problem will arrive at more biased posteriors than an agent with exogenously assigned sample size.*

# 4 Hypotheses and Empirical Strategy

In this section, I formulate hypotheses about two categories of employer behavior: beliefs and belief updating, and wage discrimination. I also discuss the empirical strategies I use to test these hypotheses.

In order to study the evolution of belief bias, we must first induce bias in prior beliefs. I accomplish this using the math task coupled with random assignment of Asian vs. Hispanic workers. Math is a task domain in which Asians are stereotyped to perform better, therefore worker group assignment should induce employers to form differential priors. This leads me to formulate the first hypothesis:

**Hypothesis 1** (Bias in Prior Beliefs)**.** *Employers hold higher prior beliefs for the average Asian worker productivity compared to Hispanic workers.*

To test for belief bias in the priors, I compare the subjective beliefs about worker scores of employers assigned to Hispanic versus Asian worker groups before any additional performance information is provided. I estimate the following Ordinary Least Squares (OLS) regression of prior belief on worker group:

$$Prior_i = \beta_0 + \beta_1 Asian_i + \beta_2 X_i + \epsilon_i \tag{17}$$

where $Prior_i$ is the prior belief of employer $i$ on the average performance of the assigned worker group, $Asian_i$ is a binary indicator that equals 1 if employer $i$ is randomly assigned the Asian worker group, and $X_i$ is a vector of employer demographic controls. $\beta_1 > 0$ would provide evidence on prior belief bias in favor of Asian workers. I estimate this regression on the pooled dataset by worker group assignment.

In the experiment, I also collect the subjective probability mass functions of employers, where they report how many workers they think had each possible score. I test for belief bias in this alternative measure of beliefs by conducting a multivariate comparison of the subjective PMFs using the Energy Distance Test (Székely and Rizzo, 2004), a nonparametric test similar to a multivariate Kolmogorov-Smirnov test.

A key advantage of my experimental design is that the prior elicitation stage is identical for employers in the same Worker Ethnicity treatment arm, regardless of information conditions. This design feature ensures that prior beliefs follow the same distribution across information conditions, allowing for clean comparisons of posterior beliefs. I test for the equality of prior belief distributions using a Kruskal-Wallis test, a non-parametric test for equality of distributions in more than two samples.

The central question I ask is how seeking information endogenously vs. acquiring information exogenously can impact belief updating and the persistence of belief bias. My theoretical framework predicts that conditional on the same sample size, individuals who maintain endogenous control over the information search process will arrive at more biased posteriors than those with exogenously assigned search (Proposition 3.4.3). This leads me to formulate the next hypothesis on treatment differences in belief bias.

**Hypothesis 2** (Bias in Updated Beliefs)**.** *The Asian-Hispanic gap in updated beliefs is smaller in the Exogenous Mean condition compared to Voluntary condition. Belief bias is the smallest in the Exogenous 20 treatment.*

To test Hypothesis 2, I compare the differences in the Asian-Hispanic belief gap in updated employer beliefs between information conditions. Specifically, I estimate the following Differences-in-Differences regression for employer $i$:

$$\begin{aligned} Posterior_i = \beta_0 &+ \beta_1 Asian_i + \rho_1 ExogenousMean_i + \rho_2 Exogenous20_i \\ &+ \phi_1 Asian_i \times ExogenousMean_i + \phi_2 Asian_i \times Exogenous20_i \\ &+ \beta_2 X_i + \epsilon_i \end{aligned} \tag{18}$$

where $Posterior_i$ is the employer's updated belief about the average score of their assigned worker group. $Asian_i$ is a binary indicator that equals 1 if employer $i$'s assigned worker group is Asian, and $ExogenousMean_i$ and $Exogenous20_i$ are indicators for whether employer $i$ is randomly assigned to the Exogenous Mean or Exogenous 20 treatment, respectively. The omitted reference group is the Voluntary treatment. $X_i$ is a vector of employer $i$'s demographic characteristics.

In this regression, coefficient $\beta_1$ represents the Asian-Hispanic belief gap in the Voluntary treatment, which serves as the reference group. Coefficients $\rho_1$ and $\rho_2$ measure the effects of other treatments - Exogenous Mean and Exogenous 20, respectively - on updated beliefs about Hispanic workers, relative to Voluntary. Coefficients for the interaction terms $\phi_1$ and $\phi_2$ represent difference-in-differences estimates, capturing how the belief gap between Asian and Hispanic workers changes under each alternative treatment

compared to Voluntary. For instance, $\phi_1$ quantifies the how the belief gap differs in Exogenous Mean relative to Voluntary, i.e. the treatment effect of removing endogeneity of information search on belief bias, while $\phi_2$ capture the analogous effects for Exogenous 20.

Next I assess discriminatory behavior. I first examine wage offers in the Baseline condition, where employers receive no information about the productivity distribution of the assigned worker group.

**Hypothesis 3** (Baseline Wage Discrimination)**.** *The wage offers for Hispanic workers are lower than those for Asian workers in the Baseline condition.*

To formally test this hypothesis, I estimate the following specification for the Baseline condition:

$$Wage_{i,j} = \beta_0 + \beta_1 Asian_i + \delta W_{i,j} + \gamma X_i + \epsilon_i \tag{19}$$

where $Wage_{i,j}$ is the wage offer by employer $i$ to worker $j$, and $Asian_i$ is a binary indicator for worker group assignment for employer $i$. The vector $W_{i,j}$ includes worker characteristics presented in the worker profile, such as age category, gender, favorite color, and $X_i$ is a vector of employer demographic characteristics. $\beta_1$ here captures the average wage gap between Asian workers and Hispanic workers within the Baseline condition. Standard errors are clustered at the employer level to account for within-employer correlation in errors across multiple workers. $\beta_1 > 0$ indicates that employers discriminate against Hispanic workers in their wage offers.

Next, I examine treatment differences in the Asian-Hispanic wage gap across information acquisition modes. If different modes of information acquisition lead to systematically different beliefs, then those beliefs should also predict subsequent wage offers. I formulate the following hypothesis based on predictions in Hypothesis 2.

**Hypothesis 4** (Wage Discrimination After Learning)**.** *The Asian-Hispanic wage gap in the Exogenous Mean treatment is smaller than the wage gap in Voluntary. The wage gap is the smallest in Exogenous 20.*

To formally test for differences in wage discrimination across conditions, I estimate the following difference-in-differences specification:

$$\begin{aligned} Wage_{i,j} = {} & \beta_0 + \beta_1 Asian_i + \rho_1 Voluntary_i + \rho_2 ExogenousMean_i \\ & + \rho_3 Exogenous20_i + \phi_1 Asian_i \times Voluntary_i \\ & + \phi_2 Asian_i \times ExogenousMean_i + \phi_3 Asian_i \times Exogenous20_i \\ & + \delta W_{i,j} + \gamma X_i + \epsilon_i \end{aligned} \tag{20}$$

In this regression, $Wage_{i,j}$ is the wage offer by employer $i$ to worker $j$, and $Asian_i$ is a binary indicator for worker group assignment for employer $i$. The dummy variables $Voluntary_i$, $ExogenousMean_i$, and $Exogenous20_i$ are indicators for information treatment assignment, with Baseline serving as the omitted reference group. Coefficient $\beta_1$ reflects the average wage gap between Asian and Hispanic workers in the Baseline condition. The coefficients $\phi_1$ through $\phi_3$ for the interactions terms represent treatment-specific differences in the wage gap relative to the Baseline. For example, $\phi_1$ measures the change in the wage gap under Voluntary relative to Baseline (i.e., the treatment effect on discrimination), and $\phi_3$ captures the same effect for the Exogenous 20 treatment. To compare wage discrimination across treatments directly, I test for statistical differences between the interaction terms $\phi_1$ through $\phi_3$. These comparisons reveal whether certain information environments, in particular exogenously assigned experiential learning, can more effectively reduce discriminatory behavior. I cluster standard errors at the employer level to account for unobserved within-employer correlations in how they make wage offers.

# 5 Results

In this section, I present results from the employer experiment. Section 5.1 reports descriptive statistics of the employer sample, and Section 5.2 reports results on bias in employer prior beliefs before learning. Section 5.3 discusses results on the employers' posterior belief bias after information search, and Section 5.4 reports results on wage discrimination.

## 5.1 Descriptive Statistics

Table 1 and Table 2 report demographic characteristics of the employer sample by information treatment for employers assigned the Hispanic worker group and Asian worker group, respectively. I collect the age, gender, and race/ethnicity of the employers directly at the end of the experiment, and augment this with information from Prolific on student status, employment status, and place of birth. We see that overall employer demographics are well balanced across treatments, with slightly more female participants. Employers have an average age of around 40, and the majority of them are White (around 70%). Asian and Hispanic employers comprise around 7% and 10% of the sample each, while the percentage of Black employers have somewhat more variance across treatments. Fur-

ther, there are no observable differences in employer demographics between those with assigned the Asian worker group and those assigned the Hispanic worker group. With the relatively small percentage of Asian and Hispanic employers, it is unlikely that results would be significantly influenced by in-group biases.

## 5.2 Employer Prior Belief Bias

I begin by examining employer prior beliefs across different information conditions in each worker ethnicity treatment arm. Importantly, my experimental design standardizes the prior elicitation stage: all employers assigned to the same worker group (Asian or Hispanic) respond to identical belief questions before receiving any information, regardless of their treatment condition. This design ensures that any differences in prior belief distributions across treatments arise solely from random variation rather than information treatment assignment. If this condition holds, it enables clean comparisons of posterior beliefs across treatments, since differences in posteriors can then be attributed to the effects of information acquisition rather than prior belief differences.

Figure 2a displays the empirical cumulative distribution functions (CDFs) of prior beliefs about the average worker score, plotted separately for each information condition for employers assigned to the Hispanic worker group. Figure 2b displays the same graph for employers assigned to the Asian worker group. These plots show that prior beliefs indeed following very similar distributions across the information treatments. To formally test for distributional differences in prior beliefs across treatments, I conduct Kruskal-Wallis tests, a multi-sample extension of the Mann-Whitney U Test, on the reported average score beliefs within each worker group across information treatments. The results fail to reject the null hypothesis of equal distribution across treatments ($p = 0.42$ for employers with Hispanic workers, and $p = 0.60$ for employers with Asian workers). This supports the internal validity of the design and confirms that post-learning differences are not solely driven by prior beliefs.

Next, I examine bias in prior beliefs—specifically, whether employers hold systematically different expectations about worker performance before receiving any information about the group. Figure 3a displays employers' prior beliefs about the average productivity of their assigned worker group, pooled across all information treatments. The results show a clear pattern: employers assigned to the Asian worker group tend to overestimate average performance, while those assigned to the Hispanic group tend to underestimate it. Both deviate from the true average productivity (indicated by the dotted horizontal

Table 1: Employer Demographics by Information Condition, Hispanic Workers

| | Baseline | Voluntary | ExogenousMean | Exogenous20 |
|---|---|---|---|---|
| Age | 40.369 | 40.620 | 36.132 | 39.768 |
| | (13.444) | (13.769) | (11.896) | (13.590) |
| Male | 0.353 | 0.413 | 0.305 | 0.415 |
| | (0.479) | (0.493) | (0.461) | (0.494) |
| Female | 0.627 | 0.562 | 0.655 | 0.564 |
| | (0.485) | (0.497) | (0.477) | (0.497) |
| Non-Binary | 0.016 | 0.025 | 0.036 | 0.021 |
| | (0.126) | (0.156) | (0.188) | (0.144) |
| White | 0.735 | 0.773 | 0.718 | 0.758 |
| | (0.442) | (0.420) | (0.451) | (0.429) |
| Black | 0.133 | 0.116 | 0.145 | 0.131 |
| | (0.340) | (0.321) | (0.353) | (0.339) |
| Asian | 0.104 | 0.087 | 0.123 | 0.097 |
| | (0.306) | (0.282) | (0.329) | (0.297) |
| Hispanic | 0.100 | 0.087 | 0.109 | 0.102 |
| | (0.301) | (0.282) | (0.312) | (0.303) |
| Education | 4.482 | 4.541 | 4.482 | 4.394 |
| | (1.258) | (1.249) | (1.244) | (1.375) |
| Student | 0.184 | 0.144 | 0.200 | 0.170 |
| | (0.389) | (0.352) | (0.402) | (0.377) |
| Employed Full-Time | 0.000 | 0.004 | 0.009 | 0.008 |
| | (0.000) | (0.064) | (0.095) | (0.092) |
| Employed Part-Time | 0.329 | 0.368 | 0.282 | 0.331 |
| | (0.471) | (0.483) | (0.451) | (0.471) |
| Unemployed | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Born in the US | 0.934 | 0.920 | 0.932 | 0.965 |
| | (0.249) | (0.272) | (0.253) | (0.185) |
| Subjects | 249 | 242 | 220 | 236 |

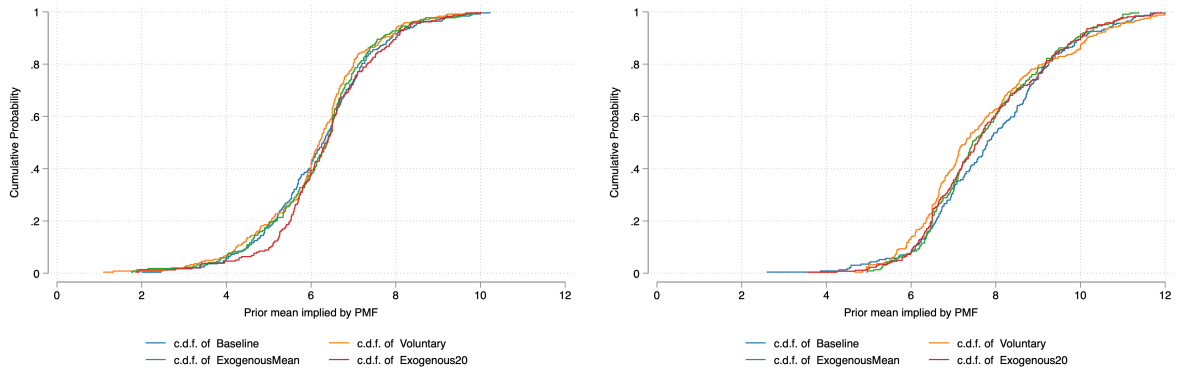
*Notes:* This table reports the mean of each demographic variable across different information treatments (Baseline, Voluntary, Exogenous Mean, Exogenous 20) for employers randomly assigned the Hispanic worker group. The corresponding standard deviation is reported in parentheses. Age is a numeric variable with integer values. The rest of the demographic variables are binary indicators, with 0 being "No" and 1 being "Yes". Age, student status, employment status, and country of birth were provided by Prolific. Gender and Ethnicity were recoded from categorical variables. Education level is a categorical variable with values 0 (Prefer not to say), 1 (Some high school or less), 2 (High school diploma or GED), 3 (Some college, but no degree), 4 (Associates or technical degree), 5 (Bachelor's degree), and 6 (Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)).

Table 2: Employer Demographics by Information Condition, Asian Workers

| | Baseline | Voluntary | ExogenousMean | Exogenous20 |
|---|---|---|---|---|
| Age | 40.598 | 40.554 | 38.764 | 39.395 |
| | (12.962) | (13.485) | (12.604) | (13.616) |
| Male | 0.384 | 0.415 | 0.387 | 0.373 |
| | (0.487) | (0.494) | (0.488) | (0.484) |
| Female | 0.585 | 0.557 | 0.582 | 0.612 |
| | (0.494) | (0.498) | (0.494) | (0.488) |
| Non-Binary | 0.026 | 0.028 | 0.031 | 0.015 |
| | (0.160) | (0.167) | (0.174) | (0.123) |
| White | 0.769 | 0.752 | 0.787 | 0.734 |
| | (0.423) | (0.433) | (0.411) | (0.443) |
| Black | 0.144 | 0.134 | 0.111 | 0.156 |
| | (0.352) | (0.342) | (0.315) | (0.363) |
| Asian | 0.074 | 0.098 | 0.089 | 0.099 |
| | (0.263) | (0.297) | (0.285) | (0.299) |
| Hispanic | 0.070 | 0.073 | 0.076 | 0.099 |
| | (0.255) | (0.261) | (0.265) | (0.299) |
| Education | 4.301 | 4.415 | 4.396 | 4.612 |
| | (1.402) | (1.284) | (1.312) | (1.163) |
| Student | 0.182 | 0.176 | 0.205 | 0.200 |
| | (0.388) | (0.382) | (0.406) | (0.401) |
| Employed Full-Time | 0.000 | 0.008 | 0.004 | 0.000 |
| | (0.000) | (0.090) | (0.067) | (0.000) |
| Employed Part-Time | 0.345 | 0.325 | 0.338 | 0.399 |
| | (0.476) | (0.469) | (0.474) | (0.491) |
| Unemployed | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Born in the US | 0.968 | 0.929 | 0.964 | 0.949 |
| | (0.176) | (0.257) | (0.186) | (0.220) |
| Subjects | 229 | 246 | 225 | 263 |

*Notes:* This table reports the mean of each demographic variable across different information treatments (Baseline, Voluntary, Exogenous Mean, Exogenous 20) for employers randomly assigned the Asian worker group. The corresponding standard deviation is reported in parentheses. Age is a numeric variable with integer values. The rest of the demographic variables are binary indicators, with 0 being "No" and 1 being "Yes". Age, student status, employment status, and country of birth were provided by Prolific. Gender and Ethnicity were recoded from categorical variables. Education level is a categorical variable with values 0 (Prefer not to say), 1 (Some high school or less), 2 (High school diploma or GED), 3 (Some college, but no degree), 4 (Associates or technical degree), 5 (Bachelor's degree), and 6 (Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)).

Figure 2: Empirical Distributions of Prior Belief by Information Condition




(a) Left: Employers assigned Hispanic workers (b) Right: Employers assigned Asian workers

*Notes:* $N = 1186$ employers who are assigned the Hispanic worker group (Left Panel), and $N = 1203$ employers who are assigned the Asian worker group (Right Panel). The empirical distributions are plotted separately for each information condition within each worker group assignment treatment arm.

line at 7.1), which is identical for both groups by design.

To formally test the hypothesis that prior beliefs are biased against Hispanic workers (Hypothesis 1), I estimate Equation 17, displayed again below:

$$Prior_i = \beta_0 + \beta_1 Asian_i + \beta_2 X_i + \epsilon_i.$$

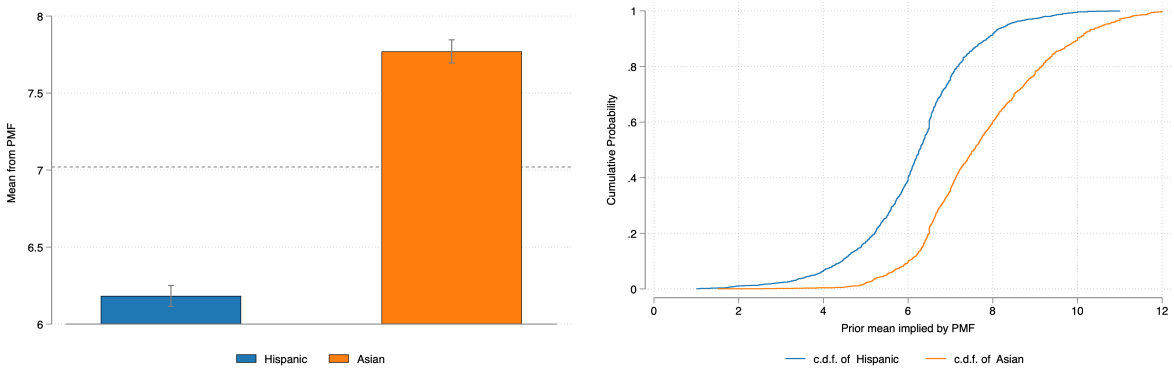
The results show a sharp divergence between the prior belief on average group productivity between Asian workers and Hispanic workers. On average, employers believe that Asian workers scored 1.53 points higher than Hispanic workers, before receiving additional performance information. This belief gap is around 90% of a standard deviation (1.69), representing a large and significant bias against Hispanic workers in employer prior beliefs.

This pattern is further confirmed in Figure 3b, which plots the empirical cumulative distribution functions (CDFs) of prior beliefs. The CDF for the Asian group first-order stochastically dominates that of the Hispanic group, indicating that employers systematically expect Asian workers to perform better. A Kolmogorov–Smirnov test rejects the null hypothesis of equal distributions ($p < 0.001$), providing strong evidence of biased prior beliefs based on worker group assignment. The next result follows:

**Result 1** (Initial Belief Bias)**.** *Before receiving performance information, employers believe that Hispanic workers have lower scores than Asian workers on average.*

In addition to examining beliefs about average worker score, I also analyze the full

Figure 3: Prior Beliefs, Pooled Across Information Conditions




(a) Mean (b) Empirical Distribution

*Notes:* $N = 2389$. Pooled across information treatments within each worker group ethnicity treatment arm.

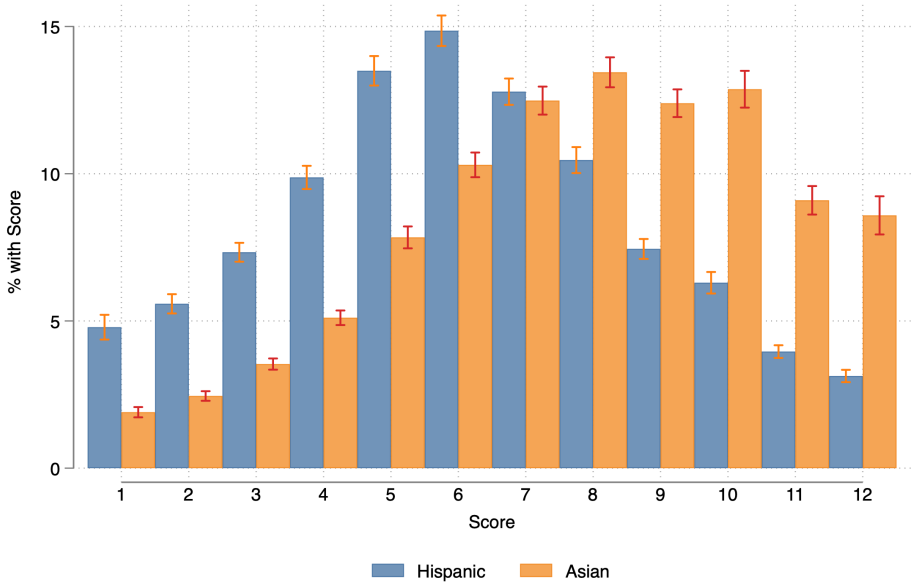
Table 3: Prior Beliefs about Worker Score

| | Prior Beliefs |
|---|---|
| Asian Worker Group | 1.526*** |
| | (0.067) |
| Constant | 5.760*** |
| | (1.041) |
| Hispanic Mean Priors | 6.216 |
| Asian Mean Priors | 7.731 |
| Observations | 1911 |
| Employer Controls | Yes |

*Notes:* This table reports OLS estimates from a cross-sectional dataset with 1911 observations, one observation per employer. The dataset consists of employers pooled over the Voluntary, Exogenous Mean, and Exogenous 20 information treatments. The dependent variable is the initial belief about the average worker group productivity, calculated from the reported subjective probability mass functions. Asian Worker Group is a binary indicator equal to 1 if the assigned worker group is Asian and 0 if the assigned group is Hispanic. The Hispanic Mean Priors and Asian Mean Priors report the raw average prior belief for average productivity in the Hispanic worker group and Asian worker group, respectively. Employer controls include employer age, gender, race/ethnicity, and education level.

subjective probability mass function over worker scores reported by employers. While average estimates offer a summary measure of perceived group performance, they may obscure important difference in the shape of the underlying distribution-particularly the expected proportions of high- and low-performing workers. To capture this, I elicit employers' beliefs about the number of worker in each possible score category in their assigned group. Figure 4 presents the results by worker group assignment. A striking pattern emerges: the subjective belief distribution for Asian workers is strongly right-skewed, with employers expecting a high concentration of top scores, whereas the belief distribution for Hispanic workers is relatively symmetric, with employers expecting similar proportions of high- and low-performers. To test for statistical differences in these belief distributions, I conduct an Energy Distance Test (Székely and Rizzo, 2004), a nonparametric test for equality of multivariate distributions that is sensitive to both location and shape. The results indicate a statistically significant difference between the two distributions ($p < 0.001$), showing that bias in prior beliefs stems from the differential expectations about the prevalence of high-performing workers in each group.

Figure 4: Prior Subjective Belief Distributions




*Notes:* Prior subjective belief distributions of employers by worker group race, pooled across all treatments.

## 5.3 Employer Posterior Belief Bias

After the information acquisition stage, employers report their updated beliefs about worker performance in their assigned group. Here I present results on the posterior belief bias in the information treatments: Voluntary, Exogenous Mean, and Exogenous 20.

Formally, I estimate the difference-in-differences specification in Equation 18, reproduced from Section 4:

$$\begin{aligned} Posterior_i = \beta_0 + \beta_1 Asian_i + \rho_1 ExogenousMean_i + \rho_2 Exogenous20_i \\ + \phi_1 Asian_i \times ExogenousMean_i + \phi_2 Asian_i \times Exogenous20_i \\ + \beta_2 X_i + \epsilon_i \end{aligned}$$

In this regression, estimated coefficient $\beta_1$ captures the belief bias against Hispanic workers (i.e. the Asian-Hispanic belief gap) in the Voluntary treatment. Coefficients $\rho_1$ and $\rho_2$ capture the treatment effects of Exogenous Mean and Exogenous 20 on the beliefs about Hispanic worker productivity. The coefficients for the interaction terms $\phi_1$ through $\phi_3$ are the difference-in-differences estimates, capturing the treatment effects on the Asian-Hispanic belief gap. I report the results for this regression in Table 4.

A striking pattern emerges in the Voluntary treatment: despite having the opportunity to sample performance information by drawing workers from the group, employers remain significantly biased against Hispanic workers. On average, employers believe Asian workers scored 0.55 points higher than Hispanic workers, a disparity equivalent to 52% of a standard deviation (1.06). While this represents a reduction in bias relative to prior beliefs, it remains substantial. In contrast, both Exogenous treatments substantially reduce this bias. In the Exogenous Mean condition, where employers are assigned the average number of draws observed in the Voluntary treatment, the belief gap narrows by 0.31 points, a 56% reduction relative to Voluntary. Notably, this reduction in belief bias is driven by the removal of endogenous control over information search rather than simply having more information, since the quantity of information is identical on average in both treatments. The Exogenous 20 condition, which provides a larger and fixed sample of 20 random workers, achieves an even greater reduction, eliminating 82% of the bias. These results suggest that the structure of information acquisition plays a critical role in shaping beliefs, with endogenous information search leading to sustained belief bias.

**Result 2** (Bias in Posterior Beliefs)**.** *Removing endogenous control over the information acquisition process in Exogenous Mean significantly reduces belief bias compared to the*

Table 4: Treatment Effects on Posterior Belief Bias

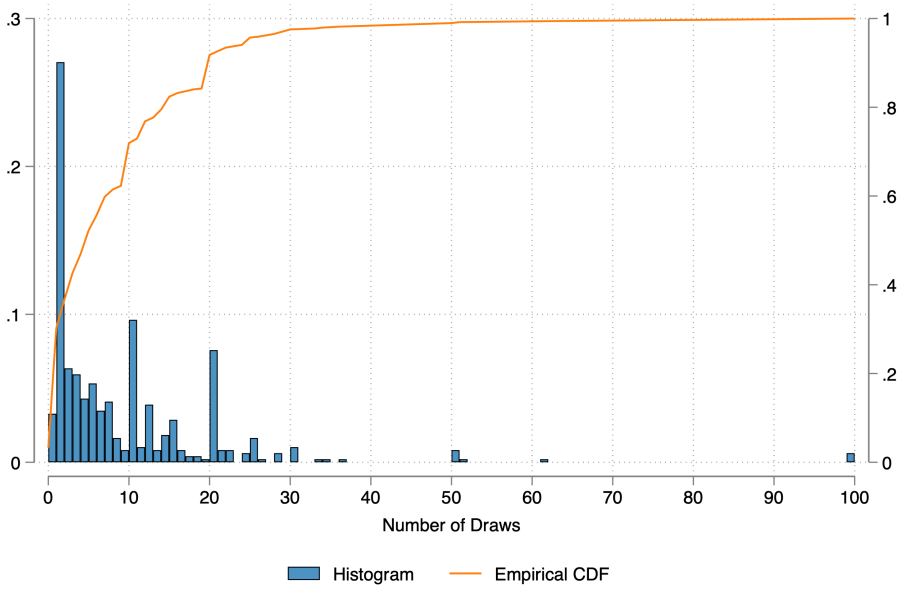
| | |
|---|---|
| Asian Worker Group | 0.549*** |
| | (0.093) |
| ExogenousMean | 0.196** |
| | (0.096) |
| Exogenous20 | 0.254*** |
| | (0.094) |
| Asian Worker Group × ExogenousMean | -0.306** |
| | (0.135) |
| Asian Worker Group × Exogenous20 | -0.455*** |
| | (0.131) |
| SD of Posterior Beliefs | 1.066 |
| $H_0 : Asian \times ExogenousMean = Asian \times Exogenous20$ | 0.267 |
| Observations | 1432 |
| Employer Controls | Yes |

*Notes:* This table reports OLS estimates of Equation 18, and the data consists of cross-sectional data from employers in the Voluntary, Exogenous Mean, and Exogenous 20 information treatments, one observation per employer. Asian Worker Group is a binary indicator equal 1 if the employer's assigned worker group is Asian. ExogenousMean, and Exogenous20 are indicators for information treatment assignment. $H_0 : Asian \times ExogenousMean = Asian \times Exogenous20$ reports the p-value from testing the null hypothesis that the treatment effects on the Asian-Hispanic bias between ExogenousMean and Exogenous20 are different. Employer controls include age, race/ethnicity, education level, and gender. Standard errors in parentheses. * ($p < 0.10$) ** ($p < 0.05$) *** ($p < 0.01$).

*Voluntary treatment, despite employers in both treatments acquiring the same amount of information on average.*

I now examine how much information employers actually gather when they have full control over the information search process. Figure 5 displays both the histogram and the empirical cumulative distribution of the number of draws in the Voluntary treatment, pooled across employers assigned to Asian and Hispanic worker groups. On average, employers make 7.8 draws—relatively low given the potential to sample up to 100 workers. The median number of draws is even lower, at 5, indicating a skewed distribution with a long right tail. There is clear clustering at salient round numbers such as 10 and 20, suggesting the use of intuitive or heuristic stopping rules. Over a quarter of employers (28%) stop after just a single draw, highlighting how limited information acquisition is for a substantial share of the sample.

Figure 5: Number of Draws in Voluntary



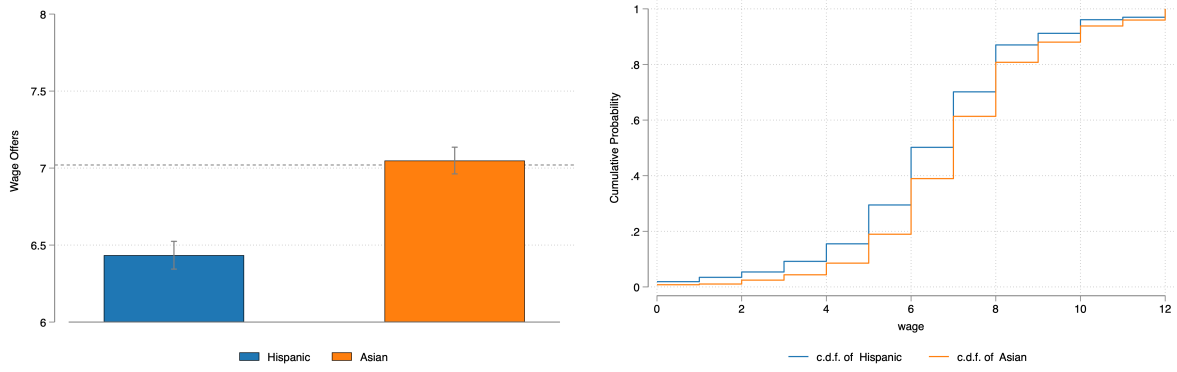

When disaggregated by worker group, we see an interesting pattern: employers assigned to evaluate Hispanic workers make slightly more draws on average (mean = 10) than those assigned to Asian workers (mean = 8). This is shown in Figure B.6, which plots the kernel density functions of number of draws by worker group assignment. The difference is mainly driven by the larger share of employers who make very few draws for Asian workers. This suggests that employers may feel more certain about their prior beliefs for Asian workers, as stronger priors will lead to less intensive search.

## 5.4 Discriminatory Behavior

Having established how different modes of information acquisition affect belief formation, I now turn to the consequences of these beliefs for decision-making. In particular, I examine how biased beliefs translate into wage discrimination. Since worker groups are designed to be objectively identical in productivity, any systematic difference in wage offers between the Asian and Hispanic groups reflects discrimination, whether rooted in biased beliefs or preference. This allows me to assess not only the behavioral relevance of belief distortions, but also whether improved belief accuracy can meaningfully reduce discriminatory behavior.

I begin by examining wage offers in the Baseline condition, where employers make hiring decisions without receiving any additional information about the productivity distribution of their assigned worker group beyond what they were told in the instructions. In this setting, employers evaluating Hispanic workers offer significantly lower wages on average than those evaluating Asian workers, despite the fact that both groups have identical underlying performance. As shown in Figure 6, the average wage gap is substantial, and the full distribution of wage offers is also first-order stochastically dominated for Hispanic workers. This pattern of behavior is consistent with belief-based discrimination: in the absence of objective information, employers rely on prior expectations that favor Asian workers over Hispanic ones. The baseline findings highlight the central role that biased prior beliefs can play in driving discriminatory outcomes, providing a benchmark against which the effects of subsequent information interventions can be evaluated.

Figure 6: Wage Offers in Baseline Condition (without Learning)




(a) Average Wage Offers

(b) Empirical Distribution of Wage Offers

To formally assess whether employers discriminate against Hispanic workers in the

Baseline condition, I estimate Equation 19 in Section 4 for employers in the Baseline condition, reproduced below:

$$Wage_{i,j} = \beta_0 + \beta_1 Asian_i + \delta W_{i,j} + \gamma X_i + \epsilon_i$$

This specification controls for both the observable characteristics of worker $j$ (captured by $W_{i,j}$) and the demographic characteristics of employer $i$ (captured by $X_i$). Standard errors are clustered at the employer level to account for within-employer correlation, since each employer evaluates and makes wage offers to ten workers. If the coefficient $\beta_1 > 0$, this would indicate that Asian workers are offered systematically higher wages than Hispanic workers with equivalent observable profiles, which is evidence of wage discrimination against Hispanics.

Table 5: Wage Discrimination in Baseline

| | Wage Offer in Baseline |
|---|---|
| Asian Worker Group | 0.662*** |
| | (0.174) |
| Female Worker | 0.354*** |
| | (0.115) |
| 45+ years old Worker | -0.133 |
| | (0.086) |
| Constant | 5.701*** |
| | (0.400) |
| Hispanic Mean Wage | 6.734 |
| Asian Mean Wage | 6.915 |
| SD of Wage | 1.925 |
| Observations | 4780 |
| Unique Employers | 478 |
| Employer Controls | Yes |
| Worker Controls | Yes |
| Clustered SE | Yes |

*Notes:* This table reports OLS estimates of Equation 19, and the data consists of panel data from employers in the Baseline condition, with 10 observations per employer. Asian Worker Group is a binary indicator equal 1 if the employer's assigned worker group is Asian. Employer controls include age, race/ethnicity, education level, and gender. Worker controls include the worker's age (18-45 or 45+), gender (Male or Female), race of the name (white-sounding name or Hispanic/Asian name), and favorite color. Coefficient estimates for worker age and gender are displayed in the table, the rest are omitted. Standard errors, clustered at the employer level, are displayed in parentheses. * ($p < 0.10$) ** ($p < 0.05$) *** ($p < 0.01$).

**Result 3** (Wage Discrimination in Baseline)**.** *Employers in the Baseline condition offer significantly lower wages to Hispanic workers compared to Asian workers with similar profiles.*

The results, presented in Table 5, show a statistically significant wage gap of approximately 0.66 experimental tokens, or about 34% of a standard deviation (1.93), favoring Asian workers. Notably, the estimates also suggest that female workers receive slightly higher wage offers on average, pointing to a potential *wage premium* for women in this setting.

This wage gap is consistent with the distorted prior beliefs documented earlier. In the Baseline condition, employers hold systematically lower beliefs about the average productivity of Hispanic workers, despite both groups having identical performance distributions. Since employers rely on these biased priors in the absence of additional information, the observed wage discrimination likely stems from incorrect beliefs rather than taste-based preferences. This interpretation is supported by the finding that belief gaps in Baseline are large and favor Asian workers. Thus, the evidence points to a key mechanism through which biased beliefs translate directly into discriminatory behavior in the labor market.

Next, I investigate how different modes of information acquisition affect wage discrimination by estimating the difference-in-differences specification in Equation 20. Table 6 reports the results. The regression includes both employer demographic controls and worker profile controls, with standard errors clustered at the employer level. The wage gap between Asian and Hispanic workers in the Baseline group is large and significant—employers offer Asian workers approximately 0.63 tokens more than Hispanic workers, as we saw from earlier results. Interestingly, in the Voluntary condition, the interaction term is negative but not statistically significant, suggesting that allowing employers to endogenously sample information about workers leads to only a modest and statistically insignificant reduction in wage discrimination. In contrast, when endogenous control over information search is eliminated (Exogenous Mean and Exogenous 20), the wage gap is significantly reduced. What is especially striking is that in the Exogenous Mean condition, where employers are exogenously assigned a number of draws equal to the average sample size in Voluntary, wage discrimination is effectively eliminated. This result underscores that it is not only the quantity of information acquired that matters, but also the mode of acquisition: removing endogenous control over search, even while holding sample size constant, dramatically reduces discriminatory behavior.

Importantly, formal F-tests confirm that the wage gap in Voluntary is significantly larger than those in Exogenous Mean and Exogenous 20, while the gaps between Exogenous treatments themselves are statistically indistinguishable. These findings closely mirror the patterns observed in belief formation: voluntary information acquisition fails to fully eliminate biased beliefs, whereas exogenous information leads to more accurate posterior beliefs and less biased wage offers.

Taken together, the results demonstrate that endogeneity in information acquisition plays a central role in the persistence of belief biases and wage discrimination. When employers are free to choose both the quantity and content of information, as in the Voluntary treatment, their posterior beliefs remain biased and their wage offers continue to discriminate against Hispanic workers. In contrast, when the learning process is externally constrained, as in the Exogenous Mean and Exogenous 20 treatments, employers place more weight on performance data, form more accurate beliefs, and offer fairer wages. The next section investigates the mechanisms that drive the differences in belief bias and discriminatory outcomes between Voluntary and Exogenous information acquisition modes.

# 6 Mechanisms

The preceding results reveal a striking pattern: employers in the Voluntary treatment exhibit the most biased posterior beliefs, despite having access to informative signals through their own sampling. In contrast, belief bias is substantially reduced in the Exogenous Mean, even when the average quantity of information is held equal to that of Voluntary. This raises the question: what features of endogenous information acquisition generate these distortions?

Broadly, belief distortions can arise in two ways: through **biased information sets**, or through **biased interpretation**. Biased information sets occur when the signals employers observe are systematically skewed toward their priors, for example, because they stop sampling once they encounter confirmatory evidence. Biased interpretation, by contrast, implies that employers in different information treatments "read" the same signals differently, such as over-weighting high scores for Asian workers and low scores for Hispanic workers. In this section, I show that biased interpretation plays little role, while realization-based stopping, i.e., the ability to stop search conditional on signal realizations, is the key driver of biased information sets.

Table 6: Treatment Effects on Wage Discrimination

| | |
|---|---|
| Asian Worker Group | 0.631*** |
| | (0.171) |
| Voluntary | 0.195 |
| | (0.148) |
| ExogenousMean | 0.536*** |
| | (0.151) |
| Exogenous20 | 0.529*** |
| | (0.154) |
| Asian Worker Group × Voluntary | -0.300 |
| | (0.213) |
| Asian Worker Group × ExogenousMean | -0.769*** |
| | (0.218) |
| Asian Worker Group × Exogenous20 | -0.757*** |
| | (0.220) |
| SD of Wage | 1.925 |
| $H_0 : Asian \times Voluntary = Asian \times ExogenousMean$ | 0.012** |
| $H_0 : Asian \times Voluntary = Asian \times Exogenous20$ | 0.015** |
| $H_0 : Asian \times ExogenousMean = Asian \times Exogenous20$ | 0.952 |
| Observations | 19100 |
| Unique Employers | 1910 |
| Employer Controls | Yes |
| Worker Controls | Yes |

*Notes:* This table reports OLS estimates of Equation 20, and the data consists of panel data from employers in the Baseline, Voluntary, Exogenous Mean, and Exogenous 20 information treatments, ten observations per employer. Asian Worker Group is a binary indicator equal 1 if the employer's assigned worker group is Asian. Voluntary, ExogenousMean, and Exogenous20 are indicators for information treatment assignment. The omitted reference group is the Baseline condition. $H_0 : Asian \times Voluntary = Asian \times ExogenousMean$ reports the p-value from testing the null hypothesis that the treatment effects on the Asian-Hispanic wage gap between Voluntary and ExogenousMean are different. $H_0 : Asian \times Voluntary = Asian \times Exogenous20$, and $H_0 : Asian \times ExogenousMean = Asian \times Exogenous20$ report the p-value from testing the analogous null hypotheses between Voluntary and Exogenous20, and ExogenousMean and Exogenous20, respectively. Employer controls include age, race/ethnicity, education level, and gender. Worker controls include the worker profile's age (18-45 or 45+ years old), gender (male or female), race (white-sounding names or Hispanic/Asian name), and favorite color. Standard errors are clustered at the employer level and reported in parentheses. * ($p < 0.10$) ** ($p < 0.05$) *** ($p < 0.01$).

## 6.1 Biased Interpretation

Previous work has documented belief biases arising through biased interpretation of information, such as motivated reasoning and confirmation bias (e.g. Zimmermann, 2020; Coutts, 2019; Eil and Rao, 2011; Mobius et al., 2014). My experimental design, however, makes this explanation less likely: random assignment ensures that employers in different information treatments should "read" signals in similar ways. Nonetheless, it is possible that agency itself, i.e., having control over information search, induces more biased interpretation in the Voluntary condition. The **Matched Sample** treatment is designed to test precisely this possibility.

In this treatment, employers are randomly assigned to observe the exact *same* sequence of signal realizations drawn by a counterpart in the Voluntary treatment with the same worker group. All other aspects of the treatment design are identical to Voluntary. In the Matched Sample treatment, all agency is removed: employers can neither control how much information to acquire, nor can they decide when to stop based on signal realizations. As a result, employers in Matched Sample receive equally biased information as employers in the Voluntary treatment. Because both the sample size and signal content are held constant across these two treatments, any differences in posterior beliefs between Matched Sample and Voluntary must arise from differences in interpretation rather than differences in information.

By eliminating all agency while holding information identical as Voluntary, the Matched Sample treatment provides a clean test of the role of biased interpretation. If belief bias observed in the Matched Sample treatment is similar to that in the Voluntary treatment, then agency itself does not affect how employers interpret signals, and biased interpretation is not a driver of observed belief distortions. If, on the contrary, belief bias is reduced in Matched Sample, then it would suggest that agency leads to more biased interpretation of information for employers in Voluntary.

I empirically test the treatment effect on belief bias in the Matched Sample treatment by estimating the following regression, extending Equation 18:

$$\begin{aligned} Posterior_i = \beta_0 + \beta_1 Asian_i + \rho_1 ExogenousMean_i + \rho_2 MatchedSample_i \\ + \phi_1 Asian_i \times ExogenousMean_i + \phi_2 Asian_i \times MatchedSample_i + \beta_2 X_i + \epsilon_i \end{aligned} \tag{21}$$

In this regression, $MatchedSample_i$ is a dummy for the Matched Sample treatment, and $Asian_i$ is a dummy for the worker group assignment of employer $i$ being Asian. The coefficient of interest is $\phi_2$, the difference-in-differences estimate for the Asian-Hispanic

belief gap in Matched Sample compared to the Voluntary treatment. $X_i$ is a vector of employer demographics.

Table 7: Treatment Differences in Belief Bias, Commitment and Matched Sample

| | (1) | (2) |
|---|---|---|
| Asian Worker Group | 0.524*** | 0.527*** |
| | (0.128) | (0.125) |
| ExogenousMean | 0.113 | 0.128 |
| | (0.133) | (0.130) |
| Match | -0.099 | |
| | (0.132) | |
| Asian Worker Group × ExogenousMean | -0.401** | -0.402** |
| | (0.186) | (0.181) |
| Asian Worker Group × MatchedSample | 0.026 | |
| | (0.186) | |
| Commitment | | 0.078 |
| | | (0.126) |
| Asian Worker Group × Commitment | | -0.397** |
| | | (0.177) |
| SD of Posterior Beliefs | 1.066 | 1.066 |
| $H_0 : Asian \times ExogenousMean = Asian \times Commitment$ | | 0.979 |
| Observations | 1376 | 1417 |
| Employer Controls | Yes | Yes |

*Notes:* This table reports OLS estimates of Equation 21, and the data consists of cross-sectional data from employers in the Voluntary, Exogenous Mean, Matched Sample, and Commitment treatments, one observation per employer. Asian Worker Group is a binary indicator equal 1 if the employer's assigned worker group is Asian. ExogenousMean, Commitment, and MatchedSample are indicators for treatment assignment. The omitted reference group is the Voluntary condition. $H_0 : Asian \times ExogenousMean = Asian \times Commitment$ reports the p-value from testing the null hypothesis that the treatment effects on the Asian-Hispanic belief gap between ExogenousMean and Commitment are different. Employer controls include age, race/ethnicity, education level, and gender. Standard errors in parentheses. * ($p < 0.10$) ** ($p < 0.05$) *** ($p < 0.01$).

The regression results are reported in Column 1 of Table 7. Belief bias in Matched Sample is not statistically different from that in the Voluntary treatment, indicating that agency over information acquisition does not have a significant impact on how employers interpret information. This finding points instead to biased information sets, acquired through endogenous information search, as the central driver of belief distortions. Even when employers are forced to passively view a biased sequence of signals, they arrive at

posterior beliefs that are similarly biased. In other words, once the information is skewed, belief distortion persists regardless of whether the employer actively chose to acquire it. This reinforces the importance of addressing the biased sampling processes itself, rather than solely focusing on post hoc interpretation, as a mechanism for reducing belief-based discrimination.

## 6.2 Biased Information from realization-based stopping

Having ruled out biased interpretation as the primary source of belief distortions, I now turn to the alternative channel: biased information sets. In the Voluntary treatment, employers have full endogenous control over the sampling process, which manifests in two key dimensions:

- *realization-based stopping*: agents decide whether to stop or continue after observing signal realizations.
- *Endogenous sample size*: agents can decide how many signals to draw.

While endogenously sample size only affects the *amount* of information, realization-based stopping can alter the *content* of the information set itself. Because employers may terminate search once they encounter signals consistent with their priors, or continue until they do so, the resulting evidence can become systematically skewed towards prior beliefs. Thus, even when the underlying information source is unbiased, realization-based stopping can produce information sets that disproportionately reinforce prior beliefs, leading to posteriors that are biased in the same direction.

To isolate the role of realization-based stopping, I introduce the **Commitment** treatment, which removes employers' ability to condition stopping decisions on the sequence of observed signals. In this treatment, employers are asked to commit ex ante to the number of workers they would like to draw, before observing any signals realizations. They then must draw that number of workers sequentially. All other features of the treatment remain identical to the Voluntary treatment. By eliminating the ability to make stopping decisions based on signal realizations, while preserving agency over sample size, the Commitment treatment creates a middle ground between fully endogenous search (Voluntary) and fully exogenous search (Exogenous Mean). If belief bias is reduced in the Commitment treatment, this would provide direct evidence that *realization-based stopping* plays a meaningful role in sustaining distorted beliefs.

The theoretical framework in Section 3 yields clear predictions across these three different information acquisition regimes. When agents commit ex-ante to a fixed sample size, the posterior belief will be a convex combination of the prior and the true average, i.e. $\mathbb{E}[\mu_n^{commit}] = \omega\mu_0 + (1-\omega)\theta$ for some weight $\omega = \frac{\sigma_d^2}{n\sigma_0^2+\sigma_d^2}$. As a result, we should expect posterior beliefs in the Commitment condition to exhibit less bias than in Voluntary. Intuitively, the Commitment treatment removes endogenous signal path but preserves endogenous sample size, therefore reducing potential for bias.

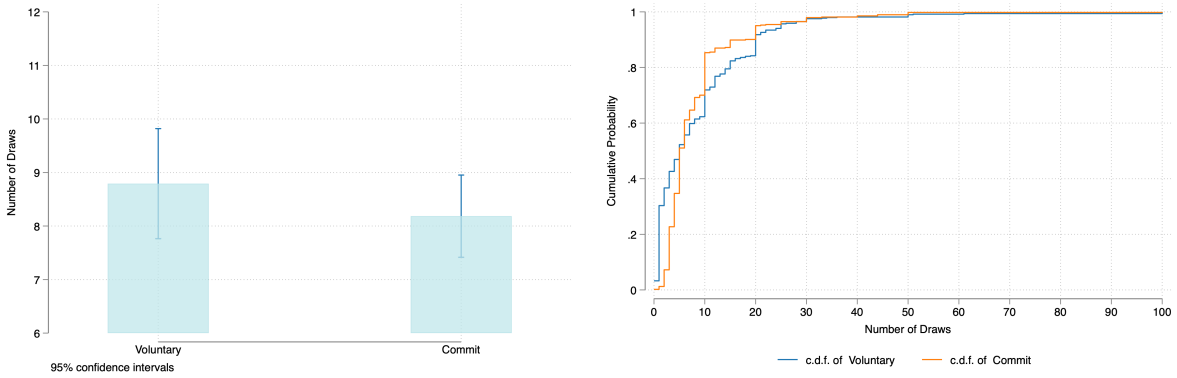
Additionally, since the sample size $n$ in Commitment is decided before observing any information, agents may choose sub-optimally small sample sizes compared to the fully Bayesian agent, because agents who desire cognitive consistency will engage in ex-ante information avoidance in order to preserve belief consistency. Further, sample sizes in Commitment will be weakly smaller than Voluntary, because now employers must choose sample size under more uncertainty and without the strategic flexibility of optional stopping, i.e. $\mathbb{E}[n_{ommit}] \leq \mathbb{E}[n_{voluntary}]$.

I begin by examining the sampling behaviors of employers from the Commitment treatment. Empirical results, shown in Figure 7, are broadly consistent with theoretical predictions. Employers in the Commitment condition draw slightly fewer workers on average, though the difference is not statistically significant. More revealing is the distributional pattern: the Commitment treatment reduces the proportion of both very small and very large sample sizes. This suggests that the flexibility to stop based on signal realizations in Voluntary introduces greater dispersion: some employers stop early when confronted with contradictory evidence in order to avoid incurring the psychological cost of revising beliefs, while others continue searching when realized signals align with their priors. In contrast, Commitment imposes ex ante discipline on the decision process, leading to a tighter, more centered distribution of sample sizes.

Next, I estimate the treatment effect on belief bias using the analogous version of Equation 21. The regression results on posterior belief bias are presented in Column 2 of Table 7. Relative to the Voluntary treatment, belief bias is substantially smaller when employers must commit ex-ante to a sample size, even though average information quantity remains comparable. In fact, the magnitude of this reduction is very similar to that observed in the Exogenous Mean treatment. This finding points to realization-based stopping, i.e., the ability to stop search after confirmatory signals, as the key mechanism producing distorted information sets. Once this channel is removed, whether through exogenous assignment of sample size or ex-ante commitment, belief bias declines sharply.

Taken together, the results across treatments suggest that realization-based stop-

Figure 7: Sample Sizes in Voluntary and Commitment Treatments




(a) Average Sample Sizes

(b) Empirical Distributions

ping is the key mechanism driving belief distortions. The Matched Sample treatment shows that biased interpretation alone cannot explain the persistence of belief gaps: once employers are exposed to a skewed set of signals, their posterior beliefs remain biased regardless of whether they actively acquired the information. By contrast, the Commitment treatment demonstrates that when realization-based stopping is removed, belief bias falls significantly, to a level comparable with Exogenous Mean. Therefore, biased posteriors in endogenous information acquisition arise not from how signals are interpreted, nor from the overall amount of information, but from the flexibility to stop selectively in response to signal realizations. Eliminating this option substantially reduces belief bias.

## 6.3 Discrimination: Biased Beliefs or Preferences?

Finally, I seek to answer the question of whether biased beliefs or preferences are driving observed discrimination in this setting. In this experiment, since the Asian and Hispanic worker groups have equal productivity distributions, I can rule out discrimination driven by accurate beliefs. Therefore, any wage differences not explained by employer beliefs are due to preferences or taste. To test the source of discrimination in this setting, I estimate the following regression separately for each information treatment:

$$Wage_{i,j,T} = \beta_0 + \beta_1 Posterior_{i,T} + \beta_2 Asian_{i,T} + +\delta W_{i,j,T} + \gamma X_{i,T} + \epsilon_{i,T} \tag{22}$$

Here, $Posterior$ is the posterior belief of employer $i$ in information condition $T$, $Asian$ is an indicator for employer $i$'s worker group assignment, $W$ is a vector of worker $j$'s characteristics shown on the resume, and $X$ is a vector of employer specific demographic

characteristics. If wage discrimination is completely driven by beliefs, the estimated coefficient $\beta_2$ should equal to 0. However, if $\beta_2 \neq 0$, it would indicate the existence of taste-based discrimination, driven by employer preferences for worker groups.

Table 8 presents the regression results. Posterior beliefs are a strong predictor of wage offers, and once we control for them, the wage gap between Asian and Hispanic workers largely disappears. This pattern suggests that most of the observed discrimination in this setting is mediated through employers' (potentially biased) beliefs about worker ability. However, some residual gap remains, indicating that a component of the wage disparity may be attributable to taste-based discrimination or other non-belief-related factors.

Taken together, results in this section highlight that discrimination in this setting is primarily driven by belief based channels. The persistent belief bias under endogenous experience stems not only from how individuals interpret signals, but also from the biased content of the information they generate. While removing strategic stopping (Commitment) reduces bias, belief distortions persist when the underlying signals remain skewed, as seen in the Matched Sample treatment. These findings clarify why the Exogenous treatments are most effective: by removing realization-based stopping, they directly address the root of biased learning. Next, I conclude by discussing the broader implications of these findings for reducing discrimination in belief formation and decision-making.

# 7 Conclusion

> "We can be blind to the obvious, and we are also blind to our blindness."
>
> – Daniel Kahneman, *Thinking, Fast and Slow*

This paper demonstrates that the mode through which individuals acquire information plays a critical role in shaping belief formation and subsequent behavior. Using a theory-guided experiment in the context of hiring, I show that even when individuals receive similar amounts of information, the endogeneity of the learning process critically shapes how beliefs evolve. In particular, when employers engage in voluntary, sequential information search, they tend to stop after observing signals that confirm their priors. This realization-based stopping leads to skewed posterior beliefs and persistent discrimination. In contrast, when information is acquired through exogenous assignment or pre-committed sample sizes, beliefs become more accurate and wage discrimination disappears. Crucially, these differences arise despite holding the amount of information constant.

Table 8: Source of Wage Discrimination

| | Voluntary | ExogenousMean | Exogenous20 | Commitment | MatchedSample |
|---|---|---|---|---|---|
| Posterior Belief | 0.179*** | 0.373*** | 0.365*** | 0.259*** | 0.232*** |
| | (0.055) | (0.082) | (0.070) | (0.064) | (0.057) |
| Asian Worker Group | 0.250* | -0.261* | -0.172 | 0.326** | -0.038 |
| | (0.130) | (0.133) | (0.135) | (0.140) | (0.142) |
| Observations | 4880 | 4450 | 4990 | 4840 | 4430 |
| Employer Controls | Yes | Yes | Yes | Yes | Yes |
| Worker Controls | Yes | Yes | Yes | Yes | Yes |
| Clustered SE | Yes | Yes | Yes | Yes | Yes |

*Notes:* This table reports OLS estimates of Equation 22, and the data consists of panel data from employers in the Voluntary, Exogenous Mean, Exogenous 20, Commitment, and Matched Sample information treatments, 10 observations per employer. Asian Worker Group is a binary indicator equal 1 if the employer's assigned worker group is Asian. ExogenousMean, Exogenous20, Commitment, and MatchedSample are indicators for information treatment assignment. The omitted reference group is the Voluntary condition. Employer controls include age, race/ethnicity, education level, and gender. Worker controls include the worker profile's age (18-45 or 45+ years old), gender (male or female), race (white-sounding names or Hispanic/Asian name), and favorite color. Standard errors are clustered at the employer level and reported in parentheses. * ($p < 0.10$) ** ($p < 0.05$) *** ($p < 0.01$).

These findings clarify why biased beliefs survive despite informative signals: not because decision makers ignore evidence, but because they control when and how to seek it. By identifying realization-based stopping as a key mechanism, the paper complements explanations centered on non-Bayesian updating, motivated reasoning, and cognitive imprecision in the encoding of signals.

These findings carry important implications for policymakers and practitioners seeking to combat bias. Many existing efforts to reduce discrimination focus on exposure to information, but my results show that exposure alone is insufficient if people retain discretion over stopping; the remedy is to structure the learning process. In hiring, for instance, organizations can set fixed, pre-committed review budgets — evaluate a predetermined number of resumes and complete a standardized set of interview questions before any decisions. Moving from rolling, real-time decisions to batched evaluation windows with a common decision date further limits opportunistic stopping. Similar logic applies to promotion and performance review, where fixed-horizon evidence packets, pre-specified time windows, and standardized requirements can replace ad hoc "stop-when-satisfied" sampling. And in evidence-generating contexts such as program pilots and clinical research, pre-specified sample sizes or stopping boundaries prevent realization-based stopping from biasing conclusions. These are low-cost, scalable designs that improve belief accuracy and decision quality without changing content.

This paper also lays the foundation for future work exploring belief formation in value-neutral settings. In an upcoming project, I study how endogenous and exogenous modes of information acquisition affect beliefs in value-neutral domains. By removing potential motivated reasoning tied to group identity, this follow-up study will help isolate the pure cognitive channels through which information structure shapes belief accuracy. Together, these lines of research deepen our understanding of how people learn about the world, and how thoughtful design of learning environments can support better judgments and more equitable outcomes.

# References


Alesina, A., M. Carlana, E. L. Ferrara, and P. Pinotti (2024). Revealing Stereotypes: Evidence from Immigrants in Schools.

Alesina, A., A. Miano, and S. Stantcheva (2023, 1). Immigration and Redistribution. *Review of Economic Studies 90*(1), 1–39.

Arrow, K. J. (1973). The Theory of Dscrimination. In O. Ashenfelter and A. Rees (Eds.), *Discrimination in Labor Markets*, pp. 3–33. Princeton University Press.

Azrieli, Y., C. P. Chambers, and P. J. Healy (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy 126*(4), 1472–1503.

Becker, G. M., M. H. DeGroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science 9*(3), 226–232.

Becker, G. S. (1971). *The Economics of Discrimination* (2nd Editio ed.). Chicago, IL: The University of Chicago Press.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2020). Inaccurate Statistical Discrimination: An Identification Problem.

Bohren, J. A., J. Hascher, A. Imas, M. Ungeheuer, M. Weber, and J. Aislinn Bohren (2024). NBER WORKING PAPER SERIES A COGNITIVE FOUNDATION FOR PERCEIVING UNCERTAINTY A Cognitive Foundation for Perceiving Uncertainty. Technical report.

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

Campos-Mercade and Mengel Friederike (2022). Non-Bayesian Statistical Discrimination. *134*.

Coffman, K. B., C. L. Exley, and M. Niederle (2019). The role of beliefs in driving gender discrimination.

Coutts, A. (2019). Good news and bad news are still news: experimental evidence on belief updating. *Experimental Economics 22*(2), 369–395.

De Quidt, J., J. Haushofer, and C. Roth (2018). Measuring and bounding experimenter demand. *American Economic Review 108*(11), 3266–3302.

Eil, D. and J. M. Rao (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics 3*(2), 114–138.

Eyting, M. (2022). Why do we Discriminate? The Role of Motivated Reasoning.

Gupta, N., L. Rigotti, and A. Wilson (2021). The Experimenters' Dilemma: Inferential Preferences over Populations.

Haaland, I. and C. Roth (2019). Beliefs About Racial Discrimination and Support for Pro-Black Policies.

Haaland, I. and C. Roth (2020, 11). Labor market concerns and support for immigration. *Journal of Public Economics 191*.

Hertwig, R. and I. Erev (2009, 12). The description-experience gap in risky choice.

Khaw, M. W., Z. Li, and M. Woodford (2021, 7). Cognitive Imprecision and Small-Stakes Risk Aversion. *Review of Economic Studies 88*(4), 1979–2013.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015, 4). How elastic are preferences for redistribution? Evidence from randomized survey experiments.

Mobius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2014). Managing Self-Confidence.

Mummolo, J. and E. Peterson (2019). Demand effects in survey experiments: An empirical assessment. *American Political Science Review 113*(2), 517–529.

Oprea, R. and F. M. Vieider (2024). Minding the Gap: On the Origins of Probability Weighting and the Description-Experience Gap *. Technical report.

Pedersen, M. J. and V. L. Nielsen (2024, 2). Understanding Discrimination: Outcome-Relevant Information Does Not Mitigate Discrimination. *Social Problems 71*(1), 77–105.

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

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

Székely, G. J. and M. L. Rizzo (2004). Testing for equal distributions in high dimension. *InterStat 5*(16.10), 1249–1272.

Woodford, M. (2012). Prospect theory as efficient perceptual distortion. In *American Economic Review*, Volume 102, pp. 41–46.

Wozniak, D. and T. Macneill (2018). Ethnic Discrimination in the Lab: Evidence of Statistical and Taste-Based Discrimination 1. Technical report.

Yariv, L. (2001). I'll See It When I Believe It — A Simple Model of Cognitive Consistency.

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review 110*(2), 337–363.

# Appendix A Additional Details for Theoretical Framework

## A.1 The Rational Employer

**Belief Evolution** Consider the evolution of beliefs and write the posterior mean in period $t+1$ as a function of the posterior mean in period $t$:

$$\mu_{t+1} = \omega_{t+1} x_{t+1} + (1 - \omega_{t+1})\mu_t \tag{23}$$

where $\omega_{t+1} = \frac{\sigma_{t+1}^2}{\sigma_d^2} = \frac{1}{1+t+\frac{\sigma_d^2}{\sigma_0^2}}$.

Since the posterior predictive of $x_{t+1}$ is centered around $\mu_t$ with variance $\sigma_{t+1}^2 + \sigma_d^2$, we can write $x_{t+1}$ as

$$x_{t+1} = \mu_t + \delta_{t+1},\ \delta_{t+1} \sim \mathcal{N}(0, \sigma_{t+1}^2 + \sigma_d^2). \tag{24}$$

Therefore, the posterior mean evolves stochastically as

$$\mu_{t+1} = \mu_t + \omega_{t+1}\delta_{t+1},\ \delta_{t+1} \sim \mathcal{N}(0, \sigma_{t+1}^2 + \sigma_d^2). \tag{25}$$

It is easy to see that

$$\mathbb{E}(\mu_{t+1}|\mu_t) = \mu_t, \tag{26}$$

**Instrumental Utility.** The instrumental utility is the probability that the true state $\theta$ falls within $\pm b$ of her current posterior mean $\mu_t$. That is,

$$\begin{aligned} u(\mu_t, \theta) &= Pr\left(|\mu_t - \theta| \leq b\right) \\ &= Pr\left(\mu_t - b \leq \theta \leq \mu_t + b\right) \\ &= Pr\left(\frac{-b}{\sigma_t} \leq Z \leq \frac{b}{\sigma_t}\right) \\ &= 1 - 2\Phi\left(\frac{-b}{\sigma_t}\right) = 1 - 2\left(1 - \Phi\left(\frac{b}{\sigma_t}\right)\right) \\ &= 2\Phi\left(\frac{b}{\sigma_t}\right) - 1 \end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$ and $\Phi$ is the cumulative distribution function of the standard normal.

**Concavity of the Value Function.** The concavity of the value function can be shown using backward induction. Since the problem is finite horizon, the agent must stop in the last period $T$, so:

$$V_t(\mu) = u(\mu) = \mathbb{P}(|\mu - \theta| \leq b)$$

which is trivially concave. This is our base case.

Now for the induction step, assume $V_{t+1}(\mu)$ is concave. We consider:

$$V_t(\mu) = \max\{u(\mu), -c + \mathbb{E}[V_{t+1}(\mu')]\}$$

We already showed $u$ is concave from the base case, and the continuation value is an expectation over concave functions, therefore it is also concave. So $V_t(\mu)$ is the max over two concave functions, which is necessarily concave. Therefore, the value function $V$ is concave.

**Equivalence between the Sequential Problem and the Ex-Ante Commitment Problem.** More formally, we can show this equivalence using backward induction. In the ex–ante (commitment) formulation, the agent chooses a stopping time $\tau \in \{0, 1, \ldots, T\}$ at $t = 0$ to maximize

$$V^{commit}(\mu_0) = \max_{0 \leq \tau \leq T} \mathbb{E}\big[u(\mu_\tau) - c\tau \mid \mu_0\big].$$

Equivalently, define the value functions

$$V_T(\mu) = u(\mu), \qquad V_t(\mu) = \max\Big\{\, u(\mu),\ -c + \mathbb{E}\big[V_{t+1}(\mu_{t+1}) \mid \mu_t = \mu\big] \Big\}, \quad t = T-1, \ldots, 0.$$

We now show by backward induction that for each $t$ and belief $\mu_t$,

$$V_t(\mu_t) = \max_{\tau \geq t} \mathbb{E}\big[u(\mu_\tau) - c(\tau - t)\big|\mu_t\big].$$

**Base case ($t = T$).** Since no draws remain, the only stopping time $\tau \geq T$ is $\tau = T$. Thus

$$\max_{\tau \geq T} \mathbb{E}\big[u(\mu_\tau) - c(\tau - T) \mid \mu_T = \mu\big] = \mathbb{E}\big[u(\mu_T) \mid \mu_T = \mu\big] = u(\mu) = V_t(\mu).$$

**Inductive step.** Assume for some $t + 1 \leq T$ that

$$V_{t+1}(\mu_{t+1}) = \max_{\tau \geq t+1} \mathbb{E}\big[u(\mu_\tau) - c(\tau - (t+1)) \mid \mu_{t+1}\big].$$

Then at date $t$, by definition,

$$\begin{aligned}
V_t(\mu_t) &= \max\{u(\mu_t), -c + \mathbb{E}[V_{t+1}(\mu_{t+1})|\mu_t]\} \\
&= \max\{u(\mu_t), -c + \mathbb{E}[\max_{\tau \geq t+1} \mathbb{E}\big[u(\mu_\tau) - c(\tau - (t+1)) \mid \mu_{t+1}\big]|\mu_t]\} \\
&= \max\{u(\mu_t), \max_{\tau \geq t+1} \mathbb{E}[u(\mu_\tau) - c(\tau - t)|\mu_t]\} \\
&= \max\{u(\mu_t) - c(t-t), \max_{\tau \geq t+1} \mathbb{E}[u(\mu_\tau) - c(\tau - t)|\mu_t]\} \\
&= \max_{\tau \geq t} \mathbb{E}\big[u(\mu_\tau) - c(\tau - t)\big|\mu_t\big].
\end{aligned}$$

Hence we have shown that

$$V_t(\mu_t) = \max_{\tau \geq t} \mathbb{E}\big[u(\mu_\tau) - c(\tau - t)\big|\mu_t\big]$$

for all $t$ and $\mu_t$. Setting $t = 0$ shows

$$V_0(\mu_0) = \max_{0 \leq \tau \leq T} \mathbb{E}\big[u(\mu_\tau) - c\tau \mid \mu_0\big] = V^{commit}(\mu_0).$$

Furthermore, the greedy "stop-if" rule

$$\tau^* = \min\big\{\, t \geq 0 :\ u(\mu_t) \geq -c + \mathbb{E}_[V_{t+1}(\mu_{t+1}) \mid \mu_t]\big\}$$

attains this value, so the sequential (time-consistent) policy coincides exactly with the ex–ante optimal stopping time.

# Appendix B Additional Figures

Figure B.1: Elicitation for Belief Distribution

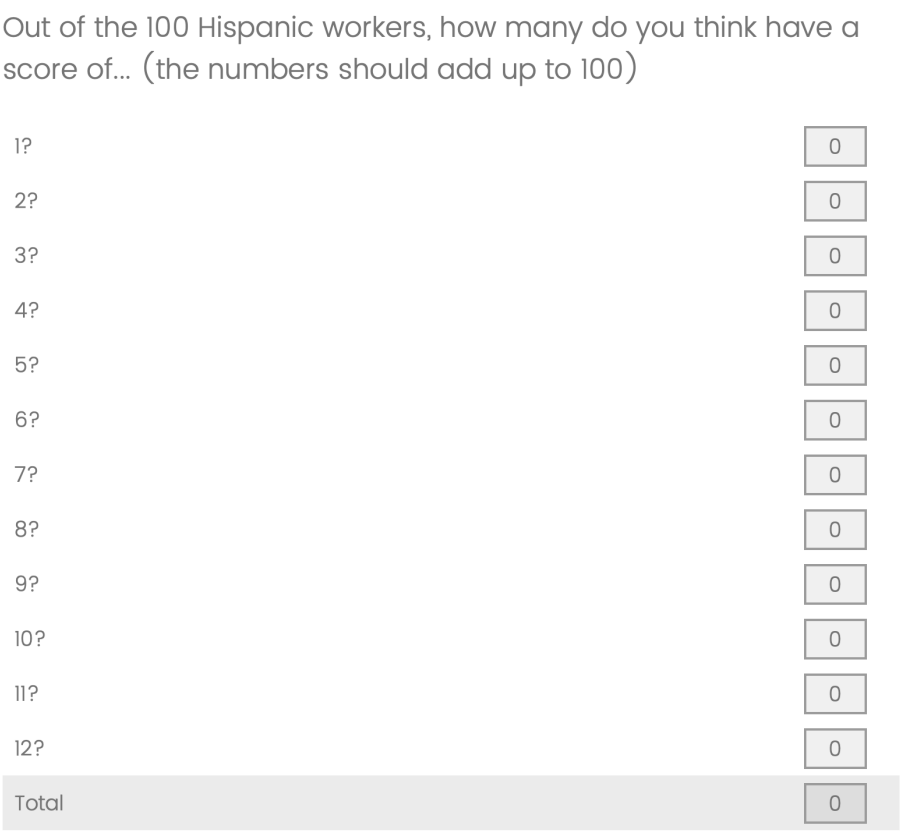



Figure B.2: Elicitation for Belief on Average Worker Score

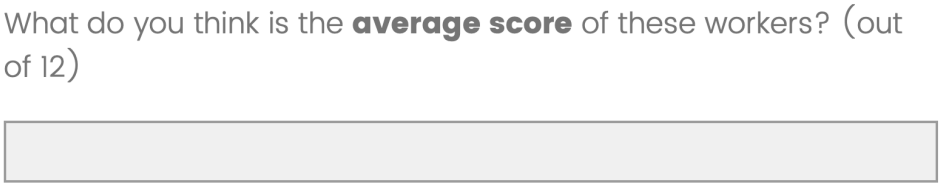


Figure B.3: Sampling Interface

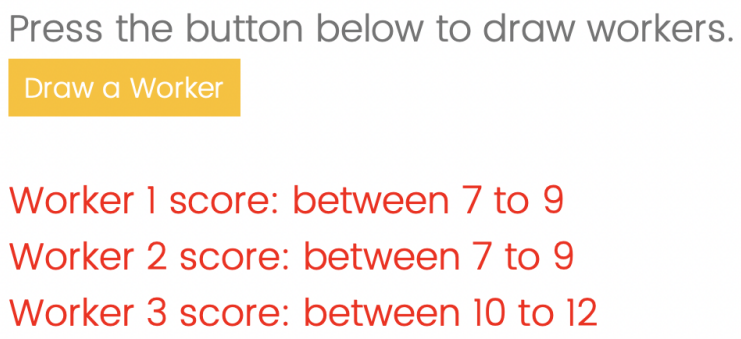
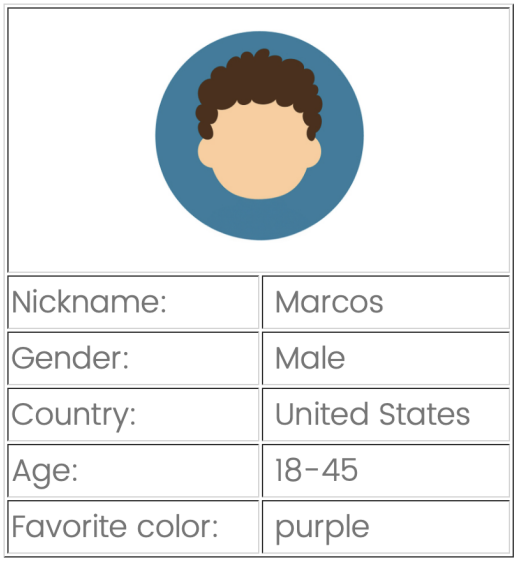


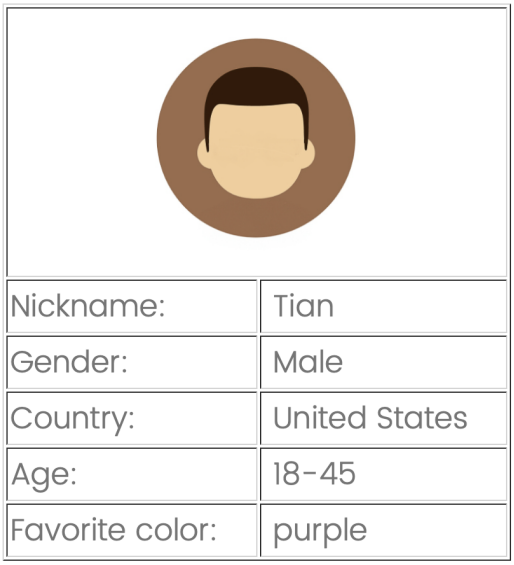

Figure B.4: Example Worker Profiles

(a) Example Hispanic Profile



| | |
|---|---|
| Nickname: | Marcos |
| Gender: | Male |
| Country: | United States |
| Age: | 18-45 |
| Favorite color: | purple |

(b) Example Asian Profile



| | |
|---|---|
| Nickname: | Tian |
| Gender: | Male |
| Country: | United States |
| Age: | 18-45 |
| Favorite color: | purple |

Figure B.5: Wage Offer Interface

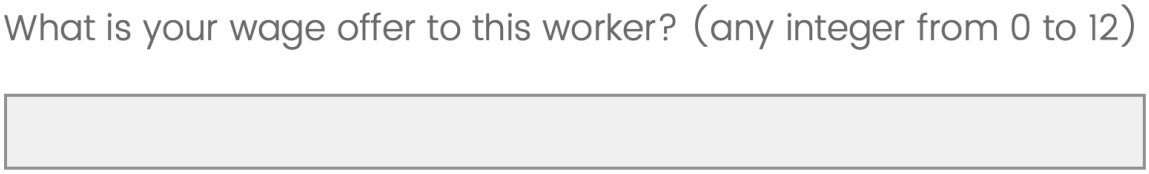


Figure B.6: Number of Draws in Voluntary, by Worker Group