The Effects of a Foreign Accent on Hiring Discrimination

Elisa Taveras
Ozlem Tonguc
Maria Zhu
Nicola Miller*

January 30, 2025

Abstract

This study investigates whether employers display accent bias in their beliefs about workers' productivity. To assess accent bias, we implement a hiring lab experiment in which employers are randomly assigned to a treatment that either reveals the worker's accent ("Accent Revealed") or does not ("Accent Blind"). We elicit employers' beliefs regarding randomly assigned workers' productivity based on their resumes and accent characteristics. We study worker productivity in skills: mathematics, verbal reasoning, and knowledge of social norms. We find that evaluators assess accented workers as performing worse than non-accented counterparts on all three skills and that this gap persists even after showing employers a signal of the worker's performance. Employers do not display differences in willingness to pay or beliefs regarding workers' willingness to accept by worker accent status.

Keywords: Foreign accent discrimination, Labor markets, Laboratory experiments

JEL Classification Codes: C91, D90, J01, J71.

Acknowledgement: This project was reviewed by Binghamton University's Institutional Review Board and determined exempt under protocol STUDY00005035. The authors declare no conflicts of interest.

*Taveras: University of Texas, Rio Grande Valley (elisa.taveraspena@utrgv.edu); Tonguc: Binghamton University (otonguc@binghamton.edu); Zhu: Syracuse University (mzhu33@syr.edu); Miller: Binghamton University (nmille26@binghamton.edu). All mistakes are our own.

1 Introduction

Many developed economies, including the United States, rely heavily on the labor of foreign-born workers. In the US, the proportion of foreign-born workers in the civilian labor force reached a record high of 18.1 percent in 2022 (Bureau of Labor Statistics, 2023). Furthermore, a significant number of these workers are non-native English speakers: 80 percent of the foreign-born civilian labor force in the US speak a language other than English at home, and 41 percent of the foreign-born civilian labor force self-report not being fully proficient in English (Ruggles *et al.*, 2024).[1] As the share of foreign-born workers in the US labor force continues to grow, it is important to understand their experiences in the labor market. This paper investigates one potentially relevant dimension of labor market discrimination for this group: accent bias.

Speech is a critical determinant of perceived competence and employability, and one's accent certainly makes up a necessary component of how others receive speech. The salience of accent in communication makes it especially vulnerable to potential bias. Indeed, studies find that people evaluate those with non-native accents differently than their counterparts with native accents when it comes to trustworthiness (Munro *et al.*, 2006; Burns & Keswell, 2015; Caballero & Pell, 2020; Spence *et al.*, 2024). Furthermore, research has shown that accents can affect the degree to which listeners understand a message, leading to communication delays (van Wijngaarden *et al.*, 2001; Floccia *et al.*, 2009).[2] The presence of non-native accents has also been linked to perceptions of lower competence and trust, even when there are no measurable costs to communication (Rubin, 1992; Adank *et al.*, 2009; Shuck, 2004, 2006). These perceptions potentially contribute to wage disparities, as (Isphording & Sinning, 2012) find that higher English ability is associated with increased wages in the US. Moreover, language skills are widely regarded as a form of human capital, yielding significant labor market returns across countries such as Australia, the United States, Canada, and Israel (Chiswick & Miller, 1995).

This paper studies whether employers display bias towards workers who have a non- North American English accent who are otherwise comparable. Analyzing the presence of accent discrimination in the labor market requires separating accent status from factors correlated with accent, such as race and ethnicity, education level, language proficiency, and ability to

[1] These values were calculated using American Community Survey (ACS) data from 2022. The ACS measures English proficiency on a five-point scale: 1: Does not speak English, 2: Yes, but not well, 3: Yes, speaks well, 4: Yes, speaks very well, 5: Yes, speaks only English. We define individuals as not completely proficient in English if they report their proficiency as 1, 2, or 3.

[2] This remains true even though studies find that strong accents are often still intelligible, and listeners are quick in adapting to non-native accents, mitigating such delays over time (Munro & Derwing, 1995; Derwing & Munro, 2009; Bradlow & Bent, 2008; Baese-Berk *et al.*, 2013).

communicate with other workers.[3] To overcome the challenge of disentangling the effect of accent from the effect of confounding factors, we conduct an online lab hiring experiment on participants residing in the US. We explore whether the accent status of workers influences employers' assessments of workers' abilities and whether these assessments change in response to an informative performance signal. Second, we explore whether employers offer differential compensation to workers with a foreign accent. Lastly, we explore whether employer beliefs about workers' bargaining power can explain compensation differentials.

Implementing a lab experiment allows us to circumvent some of the limitations of correspondence studies, where inferences about discrimination are based on callback rates for a job posting based on fictitious resumes. By conducting a lab experiment, we construct a richer dataset that includes beliefs about worker productivity and bargaining power, in addition to wage offers. This richer dataset allows us to explore the sources of accent discrimination across various dimensions. Importantly, we can present audio recordings of actual participants in our experiment. In a real-world labor market, the employer would only speak to the worker after callbacks have been made. Note that the latter interaction may also be achieved in audit studies, where fictitious workers undergo interviews. However, such methods are subject to the researcher being able to control for workers' accents and speech content fully.

The experiment consists of two steps. In the first step, we create a pool of workers for whom we collect information on demographics and preferences, as well as a voice recording. The audio clip consists of the worker reading standardized content we generate based on their preferences, allowing us to hold constant English proficiency (i.e., vocabulary, grammar, sentence formation). In addition, we collect data on performance on three skills: mathematics, verbal, and knowledge of social norms by having workers take a 15-question test on each skill.

In the second step, we implement an information intervention that allows us to compare the evaluator (hereafter, employer) decisions of workers who are comparable on all dimensions except accent status. Each employer evaluates workers on one of the three skills tested. In the spirit of a correspondence study, we present employers with a worker's profile card ("CV") that includes worker demographic information and educational attainment. As part of the profile, the employer also listens to an audio clip where the voice characteristics vary based on the information treatment the employer is randomly assigned. All audio clips presented to employers are created using artificial intelligence (AI) software. We present employers assigned to the *Accent Revealed* treatment with audio clips generated us-

[3]In this paper, we refer to "accents" to mean non-North American English accents, rather than domestic regional English accents.

ing speech-to-speech transformation software to voiceover the workers' original recordings. These generated recordings have a standardized (gender-specific) AI voice but still preserve the workers' accent-related speech characteristics (e.g., pronunciation, intonation) while standardizing non-accent-related characteristics (e.g., timbre, volume). If instead the employer is assigned to the *Accent Blind* treatment, we present them with an audio clip generated using text-to-speech software. In this treatment, the written transcript of a worker's audio clip is then generated using an AI voice that masks the worker's accent with a neutral North American English accent. These audio clips do not preserve workers' accent-related speech characteristics. Using information on workers' CVs and audio clips, employers are asked to assess how they believe different workers performed on skills tests. These elicitations are incentivized to ensure truthful reporting.

We find that employers in the *Accent Revealed* treatment believe that, on average, workers with an accent perform worse in all three skill categories compared to workers without accents. Precisely, the accent gap in beliefs over workers' performance is -0.47 points for Math, -0.74 for Verbal, and -0.53 points for Social Norms, corresponding to 6.7 percent, 7.4 percent, and 7.2 percent of the average score in these skills, respectively. Interestingly, this bias is present even when very little English knowledge is needed, as the Math test consists of arithmetic problems that do not require significant language comprehension, and employers are informed of such. Furthermore, we find that these accent gaps persist even after controlling for the worker's actual test score, meaning they do not reflect accurate beliefs about accent gaps on the part of employers.

One potential reason for the accent gap in employer evaluations could be that employers do not have enough information about each worker's ability. We explore this mechanism with a second information intervention, in which employers evaluate additional workers after observing a signal of the worker's performance in the task. We find that while employers adjust their overall beliefs about performance in response to the signal, the signal does not reduce the accent gap in beliefs about worker performance.

Furthermore, we collect information regarding employers' willingness to pay to hire a given worker for their skill test performance. To do so, we conduct a hiring task in which employers decide whether to hire given workers and employer compensation is a function of workers' test performance. We find that worker accent does not affect employers' willingness to pay, suggesting that employers are willing to close the accent gap at a cost to themselves when making an offer that affects the worker's payoff.

Finally, we explore a novel channel through which beliefs about worker bargaining power may create wage differentials. Intuitively, employers may offer certain types of workers lower wages if they perceive these workers have fewer outside options and are thus more likely

to accept lower wage offers. To do so, we collected workers' willingness to accept values in an Ultimatum Game in the worker study. We then collect employer beliefs over workers' willingness to accept and find no accent-based differences in these beliefs.

This paper contributes to multiple areas of research. First, our work builds on a growing body of experimental work, both in a lab and field setting, seeking to understand labor market discrimination better. Prior experimental work has looked at labor market discrimination along factors such as race, immigration status, and physical disability status (Bertrand & Mullainathan, 2004; Oreopoulos, 2011; Ge & Wu, 2024; Bellemare *et al.*, 2023; Tomlin, 2024; Eames, 2024). Second, this study speaks to existing research on perceptions of foreign accents. Taste-based discrimination against speakers with foreign accents has been well-documented, with such individuals rated as less sonically pleasant, intelligent, and competent, and also of lower social status than speakers of the majority language with a standard accent (Gluszek & Dovidio, 2010; Fuertes *et al.*, 2012). Even native speakers are ascribed lower comprehensibility when assumed to be non-native (Rubin, 1992). Additionally, studies have documented belief-based discrimination in situations where comprehensibility issues are perceived as communication barriers, in which non-native speech are viewed as imposing higher cognitive processing costs (Adank *et al.*, 2009; Shuck, 2004).

Existing evidence confirms that biases against accents manifest in diverse contexts. In Germany, language cues are more relevant than applicants' names in shaping employers' responses, disadvantaging those with foreign accents (Schmaus & Kristen, 2022). Similarly, in New Zealand, non-native English-speaking professional migrants face "ethnic penalties," with perceived insufficient language competence often cited as grounds for non-employment (Li & Campbell, 2009). Non-native accents have also been shown to lower hiring recommendations (Deprez-Sims & Morris, 2010), reduce purchase intentions toward customer service workers (Tombs & Rao Hill, 2014), diminish funding for entrepreneurs (Huang *et al.*, 2013), and undermine perceptions of competence and trust in multinational teams (Tenzer *et al.*, 2014).

The extensive literature on accents of non-native speakers of English documents that accents evoke stereotypes and influence evaluative judgments unconsciously (Barrett *et al.*, 2022; Munro *et al.*, 2006; Spence *et al.*, 2024), finding that this may be exhibited as early as in childhood (Spence *et al.*, 2021). Despite the extensive demonstrations of the importance of speech in evaluations, quantifying speech is non-trivial and, as such, studies on speech characteristics and social outcomes have been limited (e.g., Antonakis *et al.*, 2022; Dietrich *et al.*, 2019).

Our experiment further contributes to existing work by evaluating various accent types and skills, offering nuanced insights into how accent discrimination manifests. Lastly, to our

knowledge, this is the first study to leverage recent innovations in AI tools to explore the role of accents in research. Our use of AI significantly strengthens our research design by allowing us to control for all aspects of an audio clip's voice, except for the accent itself.

The rest of the paper proceeds as follows. Section 2 describes the experiment and describes the workers' pool and employers' samples. Section 3 introduces our empirical specifications and Section 4 presents the estimation results. Section 5 concludes the paper.

2 Experimental Design

We implemented a two-stage online experiment with participants recruited through the Prolific platform.[4] In the first stage, hereafter "Worker Study," we recruited participants who reside in the United States to play the role of a "worker." We collected information on workers' demographics and preferences to construct a profile and measured their performance in three different knowledge tests focusing on separate skills: mathematics, verbal, and knowledge of social norms. In the second stage, hereafter "Employer Study," we recruited new participants residing in the US ("employers") to evaluate the workers. After observing the profiles of the workers, employers were tasked with reporting their incentivized beliefs on workers' performances in the tests, beliefs on workers' willingness to accept in an Ultimatum Game, and employers' own willingness to pay to hire the workers for their performance in the tests. In the following, we describe each part of the experiment in detail.

2.1 The Worker Study

A total of 522 participants completed the worker study, which consisted of four parts. In Part 1, workers self-reported their demographic information, which we later used to construct a simple profile for them, which we refer to as their Curriculum Vitae (CV). In Part 2, following recent online experiments on discrimination (e.g., Bohren *et al.*, 2023), we collect workers' preferences on "filler items": characteristics that are unlikely to be related to worker performance, but help in making worker profiles distinct from each other. In our design, these items consisted of information on workers' preferences for beverages, desserts, and colors. Once workers completed a multiple choice survey on the preferences questions, they were presented with a paragraph summarizing their answers and tasked to record themselves reciting it. Thus, the recording of the filler items in our design plays the additional role of showcasing workers' accents within the following standardized context:[5]

[4] https://prolific.com/

[5] Brackets reflect text that varies based on each individual worker's reported preferences.

"I prefer ***[coffee/tea]*** *over* ***[coffee/tea]*** ***[and/but do not]*** *like to have dessert with it. As an ice cream flavor, I prefer* ***[chocolate/vanilla]*** *over* ***[chocolate/vanilla]****. I* ***[/do not]*** *like the color yellow* ***[and/but]*** *prefer* ***[yellow gold/bright yellow]*** *over* ***[yellow gold/bright yellow]****."*

In Part 3, workers completed three 15-question tests that measure the worker's knowledge across the following topics: (1) Mathematics Skills, (2) Verbal Skills, and (3) Knowledge of Social Norms about Tipping in the US:[6]

- The **Mathematics Skills** test consisted of 15 open-ended questions measuring a worker's arithmetic and algebra knowledge. Questions were presented with minimal verbal expressions to avoid differences in performance that could be attributed to knowledge of English. Workers had a 20-second-time limit to answer each question.

- The **Verbal Skills** test consisted of 15 multiple choice questions measuring knowledge of vocabulary, grammar, and logic. As in the math test, workers had a time limit of 20 seconds per question.

- The **Social Norms Knowledge** test consisted of 15 multiple choice questions, each about a tipping rate in a given scenario (e.g., leaving a 10 percent tip after dining in a restaurant). For each question, workers submitted their evaluation of the action's social appropriateness in the given scenario. Following Krupka & Weber (2013), the possible evaluations ranged from 1:"Very Socially Inappropriate" to 4:"Very Socially Appropriate," and workers were informed that the correct answer to each question would be the evaluation that *most* worker in the agree with (i.e., the median evaluation of all workers enrolled in our study).

Finally, in Part 4, we elicited workers' willingness to accept (WTA) as the "Responder" in an Ultimatum Game (UG) that they played asynchronously with an anonymous, random non-worker participant ("Proposer"), who may see their demographic characteristics and listen to their audio recording. In the UG, the proposer was tasked to propose the allocation of a $1.50 endowment between him/herself and the worker, while the responder's task was to either accept or reject the proposer's offer.[7] If the responder accepted, the endowment was allocated according to the offer, but if the responder rejected, both players earned $0. First experimentally tested by Güth *et al.* (1982), the UG is a standard game used in laboratory experiments that study norms or fairness. In this game, we utilized the strategy method

[6]Hereafter, we refer to all as skills for brevity.
[7]To avoid possible confounds, proposers in the UG are recruited on the Prolific platform separately.

(Selten, 1967): each worker submitted a threshold, WTA_w, below which they would reject the proposer's offer.

The median completion time of the worker study was 17 minutes. Participants were paid a fixed amount of \$2.75 for completion and a bonus of up to \$1.50. The bonus payment was randomly determined based on one of the knowledge tests (10 cents for each correct answer) or the worker's payoff in the UG. The average total payment was \$3.48. In Section 2.3, we describe attributes of the "worker pool" which is the subsample of workers who were ultimately evaluated by the employers.

2.2 The Employer Study

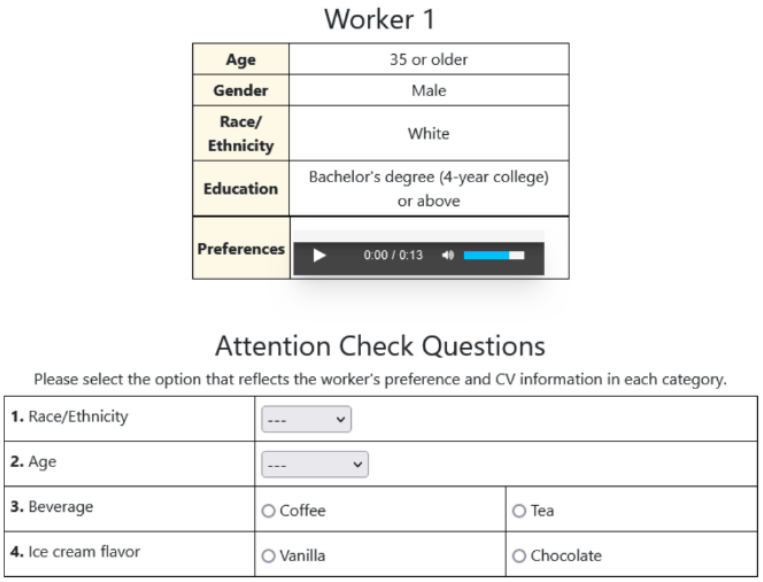
A total of 576 Prolific participants completed the employer study, in which each participant was randomly assigned to a single treatment cell in a 3×2 design (skill × information):

- **Skills:** Employers were randomized into evaluating workers based on only one of the three tests ("skills"): *Mathematics*, *Verbal*, or *Social*. To facilitate their decision making, employers were provided a sample test question (see Figure A.1-Figure A.3) as well as the average test score of the worker pool for the skill they were assigned to.

- **Information:** Employers were also randomized to listening to a version of *audio clips*: either the workers' speech characteristics related to accent were masked (*Accent Blind*) or revealed (*Accent Revealed*).

Each employer was randomly matched to 30 workers, and they were shown each profile one at a time. For each worker, the employer saw a simple profile, which we refer to as their "CV", which contained information on the worker's age, gender, race/ethnicity and education level. Employers also listened to an audio clip about the worker's preferences on filler information (beverages, desserts and colors). To ensure that employers paid attention to both the worker's CV and the audio clip, they were asked attention check questions regarding two random CV items and two statements from the audio clip for each worker. Figure 2.1 shows an example worker CV and the random attention check questions attached to it. Note that the version of the audio clip depends on the information treatment.

Figure 2.1: Example Worker CV and Attention Check Questions



In the *Accent Revealed* treatment arm, to provide information on the worker's accent, we used an Artificial Intelligence (AI) voice-transformed version of each worker's own recording collected as part of the Worker Study. The transformed voice preserved the pronunciation and intonation of the original recordings.[8] Thus, in the Accent Revealed treatment, the audio clip conveys information on whether a worker has a foreign accent or not to the employer. On the other hand, in the *Accent Blind* treatment arm, the audio clip, while still accurately stating the worker's preferences, is generated by a commercial text-to-speech AI tool using a US-native-accented voice. Thus, in this treatment we *mask* accents of workers.

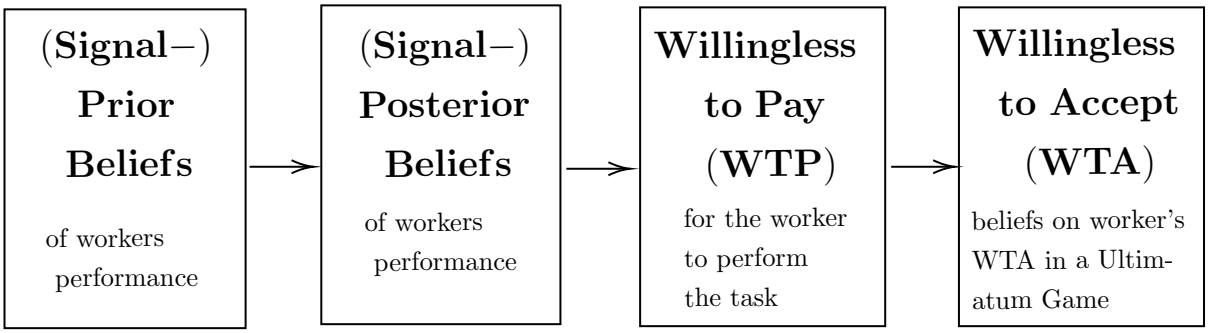
We collected demographic information as well as incentivized decisions from the employers in the study. Specifically, as shown in Figure 2.2, employers completed three incentivized tasks after observing a worker's profile and listening to an audio clip. : (1) reporting beliefs about workers' test performance, (2) reporting willingness to pay (WTP) to hire workers based on their test performance and (3) reporting beliefs about workers' willingness to accept (WTA) in the Worker Study UG.

To measure employers' baseline beliefs about workers' test performance ("*prior beliefs*"), we first asked employers, *"What is the number of questions this worker answered correctly in the* ***[Math/Verbal/Social]*** *skills test?"* Each employer responded to this question for 12 separate workers that differed from each other based on gender, race/ethnicity and accent characteristics.[9] To incentivize truthful responses, payoffs were based on Binarized Scoring

[8]Details of the voice transformation process is explained in Section 2.3.
[9]Section 2.3 provides details on how workers were classified into different categories.

Rule (BSR) (Hossain & Okui, 2013) with a maximum allowed error of 3 points.[10]

Figure 2.2: Sequence of Incentivized Tasks in the Employer Study



Note. The figure shows the timeline for information collected for each of the six treatment arms. For each information treatment, *Accent Resealed* and *Accent Blind*, employers are assigned to one skill type: *Mathematics*, *Verbal*, *Social*.

Next, to explore whether and to what extent employers' beliefs respond to information on workers' quality, we added a test performance-relevant individual **quality signal** to each worker's profile. A worker's signal was his/her score in a random subset of 5 questions in the 15-question-test ("Random-5"; see Figure A.4), which is *positively correlated* with the overall performance of the workers in the Worker Pool. The employers were informed that the test questions selected for the "Random-5" were constant and the same across all workers.

After the introduction of the signal, we first elicited employers' *posterior beliefs* on 6 additional workers where truthful reporting was incentivized by using the same payoff rules in the "Prior Beliefs" part described above. This was followed by the hiring task, in which employers decided whether to hire the additional 6 workers assigned to them in this part by submitting the highest wage they were willing to pay (hereafter WTP) to each worker in exchange for earning a revenue that increased based on the worker's test performance. Following Bohren *et al.* (2023), the presence of a hiring task in addition to the elicitation of beliefs on performance in the experiment sequence allows us to differentiate statistical from taste-based discrimination in the aggregate. Thus, in the hiring task, employers submit their WTP after observing the worker's CV, audio clip, and the "Random-5" signal. To incentivize truthful reports, the employer's payoff in this part increases in the worker's test score and the wage for hiring was based on the Becker-DeGroot-Marschak mechanism (Becker *et al.*, 1964; Healy, 2017). Note that the employers were informed that the hiring decisions were not hypothetical: if hired, the wage could be an additional bonus to the worker.[11]

[10] With this scoring rule, the employer's probability of winning the fixed payoff of \$4.50 increases quadratically as the difference between employer's prediction the worker's true score diminishes.

[11] The payoff was calculated as follows: a wage between \$0 and \$2.10 was randomly drawn and compared to

Finally, we considered the existence of an additional channel through which wage differentials may arise: beliefs about workers' bargaining power. Specifically, if these beliefs are biased, then it is possible that some wage differentials attributed to taste-based discrimination were misclassified. In order to explore this possibility, we collected employers' beliefs on workers' *willingness to accept* (WTA) in the Worker Study UG for 6 additional randomly assigned workers. To facilitate employers' understanding of the Worker Study UG, we provided the instructions as they were presented to the workers. Employers observed the worker's CV, the audio clip and the "Random-5" signal, after which they submitted their incentivized predictions by responding to the question, *"What is the lowest amount this worker accepted in the interaction? Please choose an amount between \$0 and \$1.50 (1-cent increments)."* The employer's payoff was determined based on BSR with the maximum allowed error of 15 cents.

The participants of the Employer Study were paid a fixed amount of \$4.50 for completing the study and received a bonus payment between \$0 and \$4.20, with the amount being determined based on the payoff of the decision attached to a randomly chosen worker (out of the 30 workers the employer made decisions about).

In the next section, we discuss the Worker Pool and characteristics of employers.

2.3 Worker Pool

There are 72 workers in the Worker Pool, with each worker belonging to one of 12 possible categories based on demographic and accent characteristics. The demographic characteristics are based on self-reported gender (male or female) and race/ethnicity (White, Asian, or Hispanic). The accent is based on whether the worker is classified as having a foreign accent (with respect to North American English) or not. Accent status is determined based on evaluation of the worker's voice recording using an AI tool, described in detail below.

Audio Clips

We implement a state-of-the-art voice generator and transformation software that uses AI technology to create the audio clips presented to employers.[12] We use this technology in two ways: First, we perform **speech-to-speech transformation** to voiceover the workers'

the employer's WTP. If the employer's WTP was at least as large as the randomly drawn wage, the employer hired the worker, earning a payoff of $\$2.10 + \$0.14 \times WorkerTestScore - RandomWage$. Otherwise, both the employer and the worker's payoff was \$0. The employer received the payoff as a bonus if the worker was chosen as the worker-that-counts for the employer's bonus, and the worker-that-counts got a bonus equal to their own payoff from the decision.

[12] https://elevenlabs.io/

original recordings using a pre-selected AI voice.[13] This allows us both anonymize worker recordings and also standardize voice characteristics, such as pitch and timbre, as much as possible. In doing so, we aim to control for confounding variables arising from unobservable voice characteristics across workers (e.g., vocal attractiveness and stereotypes (Zuckerman & Driver, 1989)). Importantly, the speech-to-speech transformation preserves the non-voice characteristics of the worker's speech such as pronunciation, intonation, and pace. Hence, the voice-transformed recording conveys the worker's accent. Second, we use the same software to generate **text-to-speech** audio clips based on the test transcript of workers' original recordings and using the same AI voice by gender as in the voice-transformed clips. These text-to-speech audio clips preserve the content of the worker's recording and use the same AI voice, but mask the worker's accent with a neutral North American English accent. In our experiment, the speech-to-speech transformed recordings were used in the *Accent Revealed* treatment arm while the text-to-speech clips were used on the *Accent Blind* treatment arm.

We measure a worker's accent in their voice-transformed recording using English Language Speech Assistant (ELSA), a software designed non-native speakers of English to improve their accents.[14] By uploading audio clips to the platform, users receive a detailed evaluation of their speech as well as an overall ELSA score ranging from 0 (completely beginner) to 100 percent (completely native). The overall score is based on the scores across various components of English speech: intonation, pronunciation, fluency, grammar, and vocabulary.[15,16] To measure a worker's *Accent*, we collected the intonation and pronunciation scores of the speech-to-speech recordings on ELSA.[17] We then classified workers as accented if their ELSA score on intonation or pronunciation was less than 90%, and non-accented otherwise. See an example of a recording and classification in Section A.2.

Sample

To create the Worker Pool from the 522 participants in the Worker Study, we first excluded workers who did not self-report their gender as male or female, were over 65 years of age, or did not belong to one of our target race/ethnicity groups: White, Asian, or Hispanic. We further eliminated workers whose recordings were either silent or low-quality.[18] Finally, for

[13] There are two separate voices, one for men and one for women.

[14] https://elsaspeak.com/en/.

[15] These are common components of measuring English proficiency. For example, see the rubric for the TOEFL-Speaking Section: https://www.ets.org/pdfs/toefl/toefl-ibt-speaking-rubrics.pdf

[16] ELSA also gives some approximated grades on various international tests.

[17] We excluded other categories because they are controlled factors in our experimental design: fluency (minimal variation in the length of the audio clips), while vocabulary and grammar are standardized due to workers reading a paragraph we provide them. Thus, the recordings do not reflect the workers' language ability in these categories.

[18] When possible, the workers' voice recordings were edited for quality (e.g., remove empty spaces).

each of the 12 worker categories (e.g., Female-White-Not Accented), we selected 6 workers for the Worker Pool. These workers were chosen to represent their category regarding their demographic characteristics such as education level and overall performance,[19] keeping in mind the quality of the recordings in terms of perceived accent and overall sound quality (e.g., no background noise).

We select 6 workers per worker category based on various trade-offs. First, a small number of workers per category allows us to collect as many employer evaluations per worker as possible. Since the employer study having four incentivized decision parts (see Figure 2.2), the lower limit on the number of workers per category is four. Second, workers might be perceived differently from each other for reasons that might be difficult to quantify. For example, while we make our best effort to homogenize the workers' recordings, voice-transformed recordings still have slight variations (e.g., due to the speaker's pace) across workers. By holding workers constant, for which having fewer workers is paramount, we would be able to reduce the effect of underlying characteristics. Finally, to have more variation in the workers presented to the employers, we preferred a number larger than 4 workers per category.

Table 2.1 presents the characteristics of the Worker Pool by race/ethnicity. Column (1) presents the characteristics of the complete sample, followed by the characteristics of Asians (column (2)), Hispanics (column (3)), and Whites (column (4)). On average, our sample of workers is 35 years old and more than half have a college degree or higher. Asians are 15 percentage points (hereafter, *pp*) more likely to have a college degree than the average, while the least likely to have college or higher educational attainment are Hispanics (-0.10 *pp*) followed by Whites (-0.06 *pp*).

The average scores of the Worker Pool workers are 7.1 out of 15 (or 47% of the total score) for mathematics, 9.9 (66%) for verbal, and 7.4 (49%) for the social norm test, respectively. Relative to the Worker Pool, Asians performed better the mathematics test, 32% higher than the average, while Hispanics and Whites performed worse (-19% and -13%). We observe a similar pattern in the verbal test, where Asians' scores are, on average, 6% higher than the average, while the scores of Hispanics and Whites are lower by -1% and -5%, respectively. Finally, the social norms test score is higher than the average for Asians and Hispanics (by 1.5% and 1%) and lower for Whites (-2%). Table A.1 in the appendix further breaks down these summary statistics with respect to accent.

[19]The sum of their scores on math, verbal and social skills tests.

Table 2.1: Worker Pool Descriptive Statistics

	(1) All	(2) Asian	(3) Hispanic	(4) White
Age	35.14	30.92	34.38	40.13
	(10.35)	(9.83)	(9.82)	(9.61)
College or More	0.60	0.75	0.50	0.54
	(0.49)	(0.44)	(0.51)	(0.51)
Math Score	7.07	9.33	5.75	6.13
	(3.53)	(3.27)	(3.33)	(2.94)
Verbal Score	9.90	10.50	9.79	9.42
	(1.84)	(2.00)	(1.79)	(1.61)
Social Score	7.35	7.46	7.42	7.17
	(2.32)	(2.64)	(1.93)	(2.43)
N	72	24	24	24

Note. This table presents the mean and standard deviation (in parentheses) of the sample of workers.

We further explore whether there is an accent score gap by estimating the relationship between accent and skill test score in Table 2.2. Column (1) presents the accent gap for the complete sample, followed by the results for race-ethnicity groups (columns (2)–(4)). Overall, there is qualitative evidence of an accent gap in performance for all skill tests. These gaps are also present across race/ethnicity categories, but not statistically significant, possibly due to the small number of workers per category.

Across skill tests and race/ethnicity, Asians have the smallest accent gap in the mathematics test, while Hispanics have the smallest accent gap in the verbal skills and social norms tests. On the other hand, Hispanics have the accent gap in Mathematics, while Whites have the largest for Verbal and Social skills.

Table 2.2: Relationship Between Accent and Test Scores

	(1) All	(2) Asian	(3) Hispanic	(4) White
DV: Math Score				
Accent	-0.917	-0.167	-1.833	-0.750
	(0.831)	(1.363)	(1.332)	(1.216)
DV: Verbal Score				
Accent	-0.806*	-0.833	-0.250	-1.333**
	(0.426)	(0.816)	(0.747)	(0.610)
DV: Social Score				
Accent	-0.917*	-0.417	-0.833	-1.500
	(0.540)	(1.097)	(0.787)	(0.961)
N	72	24	24	24

Note. This table estimates the effect of accent on scores in the workers' sample. * p<0.10, ** p<0.05, *** p<0.01.

2.4 Employers

In this section, we present descriptive statistics for our employer pool in Table 2.3. Column (1) presents summary statistics for the *AccentBlind* treatment arm, column (2) presents summary statistics for the *AccentRevealed* treatment arm. The average age of employers in the sample is 37 years old and about half female. More than 60% of the employers are White and have a college degree or higher education. Reassuringly, the summary statistics show that employers are comparable across the information treatment groups. Therefore, in terms of observables, we have balanced samples across the accent information treatment arms.

Table 2.3: Employers Pool Description

	(1) *AccentBlind*	(2) *AccentRevealed*
Age	36.96	36.95
	(11.99)	(11.63)
Female	0.47	0.48
	(0.50)	(0.50)
Asian	0.11	0.10
	(0.32)	(0.30)
Hispanic	0.09	0.08
	(0.29)	(0.28)
Black	0.18	0.21
	(0.38)	(0.41)
White	0.67	0.65
	(0.47)	(0.48)
College or More	0.61	0.57
	(0.49)	(0.50)
Reside in Large City	0.34	0.34
	(0.47)	(0.47)
Multilingual	0.22	0.15
	(0.42)	(0.36)
Employed Full-Time	0.57	0.53
	(0.50)	(0.50)
N	286	290

Note. This table presents the mean and standard deviation (in parenthesis) of the sample of employers. N represents the number of observations.

3 Empirical Strategy

The central goal of this paper is to explore whether employers' beliefs and decisions depend on a worker's accent. Observations are employer evaluations of workers (i.e., observations are at the employer+worker pair level). Moreover, as described in Section 2.2, we collected four types of information in the employer study. These four pieces of information allow us

answer the four research questions described below:

3.1 Does worker foreign accent affect employer assessment of worker performance?

The first question we address is whether employers display accent discrimination in assessing worker ability *prior* to receiving any information on the individual worker's performance. To answer this question, using the first 12 evaluations of each employer, we estimate:

$$y_{ie} = \beta_0 + \beta_1(Accent \times AccRev)_{ie} + \beta_2 Accent_i + \beta_3 AccRev_e + \mathbf{X}'_{ie}\Gamma + \epsilon_{ie}, \quad (1)$$

where the outcome variable y_{ie} refers to the employer e's belief on the of number of questions that worker i answered correctly out of 15. $\mathbf{X}'_{ie}$ is a vector of control variables that includes gender, age, education, and race and ethnicity: Asian and Hispanic (the reference group is White). $Accent_i$ is an indicator variable for whether the worker has an accent, and $AccRev_e$ is an indicator variable for whether the employer is in the Accent Revealed treatment group. Our treatment effect coefficient of interest is β_1, which captures the effect of revealing a worker's accent status to employers on the gap in evaluation between workers who have an accent and those who do not.

A separate question of interest is whether accent bias in evaluating worker ability is based on accurate or inaccurate perceptions. By controlling for actual test performance, we test whether differences in the perception of worker ability are attenuated. In practice, this means estimating an augmented version of Equation 1 that includes a control for the worker's actual test score, $Score_i$:

$$y_{ie} = \beta_0 + \beta_1(Accent \times AccRev)_{ie} + \beta_2 Accent_i + \beta_3 AccRev_e + \mathbf{X}'_{ie}\Gamma + \sigma Score_i + \epsilon_{ie}. \quad (2)$$

If the magnitude of β_1 is significantly attenuated in Equation 2 compared to in Equation 1, this would indicate that employers' beliefs in differences in performance between accented and non-accented workers are accurate, largely reflecting actual differences in performance between the two groups. On the other hand, if the magnitude of β_1 is similar or larger in Equation 2 compared to in Equation 1, this indicates employers' beliefs do not reflect actual differences in performance between the two groups.

3.2 Does the presence of a performance signal affect employer assessment of worker performance?

To disentangle the various factors contributing to accent gaps in assessments estimated in Section 3.1, we assess the extent to which employers update their beliefs about worker ability when given a signal about a worker's performance (the "Random-5" signal, see Section 2.2). Additionally, we investigate whether belief updating varies depending on whether an accent is present or not. To answer these question, we compare beliefs reported in the first 12 evaluations ($Signal_{ie} = 0$) to those reported in evaluations 13 - 18 ($Signal_{ie} = 1$), where $Signal_{ie}$ is an indicator variable equal to one if the employer was shown the performance signal ("Random-5").:

$$y_{ie} = \delta_0 + \delta_1(Accent \times AccRev \times Signal)_{ie} + \delta_2(Accent \times AccRev)_{ie} + \delta_3(Accent \times Signal)_{ie} + \\ + \delta_4(AccRev \times Signal)_{ie} + \delta_5 Accent_i + \delta_6 AccRev_e + \delta_7 Signal_{ie} + \mathbf{X}'_{ie}\Gamma + \epsilon_{ie}, \quad (3)$$

The outcome variable and other right-hand-side variables are as in Equation 1. The coefficient of interest in this specification, δ_1, captures the effect of providing a performance signal on employers' assessments of workers with accents relative to non-accented counterparts.

3.3 Do employer beliefs related to foreign accent result in wage differences?

From a policy perspective, a critical question is whether employer beliefs translate into differences in worker compensation. Using the employer decisions in the hiring task (evaluations 19-24, performance signal visible), we estimate:

$$WTP_{ie} = \phi_0 + \phi_1(Accent \times AccRev)_{ie} + \phi_2 Accent_i + \phi_3 AccRev_e + \mathbf{X}'_{ie}\Gamma + \epsilon_{ie}, \quad (4)$$

where the outcome variable, WTP_{ie}, is employer e's willingness to pay to hire worker i. The rest of the variables are the same as in Equation 1. Our coefficient of interest, ϕ_1, measures the effect a worker having a foreign accent on the employer's willingness to pay the worker relative a non-accented counterpart with similar characteristics.

3.4 Do employer beliefs about worker bargaining power depend on worker foreign accent?

Another factor that could affect wage offers is employers' beliefs about worker bargaining power. If beliefs differ based on worker foreign accent, employers without any taste-based preference for hiring may nevertheless offer different wages based on accent to workers they perceive to be of the same ability. This might be because they perceive individuals with accents to have a different willingness to accept in wage negotiations compared to those without a foreign accent. To test for this channel, we use data from the Willingness to Accept (WTA) guessing task, where employers report their WTA beliefs about workers' WTA in the Ultimatum Game (UG) (decisions 25-30 in the experiment). As before, employers observe performance signals for all of these decisions. We estimate:

$$WTA_{ie} = \kappa_0 + \kappa_1(Accent \times AccRev)_{ie} + \kappa_2 Accent_i + \kappa_3 AccRev_e + \mathbf{X}'_{ie}\Gamma + \epsilon_{ie}, \qquad (5)$$

where the right-hand side variables are as in Equation 1. The outcome variable, WTA_{ie}, is employer e's prediction of worker i's willingness to accept in the UG. The coefficient of interest, κ_1, measures the effect of a working having an accent on the employer's beliefs over worker's WTA.

4 Results

4.1 Does worker foreign accent affect employer assessment of worker performance?

Table 4.1 displays the results for our variable of interest from estimating Equation 1. Panel A details the estimation for Mathematics, while Panel B and Panel C focuses on the Verbal and Social skills, respectively. The columns are organized to sequentially present the results for the entire sample, followed by the results for Asian, Hispanic, and White workers.

The results show employers believe that workers with a foreign accent perform worse than non-accented workers on all the skills tests. Specifically, employers in the Accent Revealed treatment believe that workers with an accent scored 0.47 points less on the Math test, 0.74 points less on the Verbal test, and 0.53 points less on the Social norms test. These effects represent a 6.7% of the average score for Math and 7.4% and 7.2% for Verbal and Social, respectively.

Relative to non-accented workers of their own race/ethnicity, employers believe that ac-

cented Asians perform worse in Math and Verbal tests (-0.61 and -0.50 points, respectively). Accented Hispanic workers are perceived as performing worse in all tests. Finally, employers believe that accented White workers perform worse in Verbal and Social tests.

Table 4.1: Prior Beliefs over Workers' Performance

	(1) All	(2) Asian	(3) Hispanic	(4) White
Panel A. Mathematics Skills				
Accent × *AccRev*	-0.473***	-0.611**	-0.598**	-0.210
	(0.153)	(0.283)	(0.246)	(0.245)
N	2,316	772	772	772
Panel B. Verbal Skills				
Accent × *AccRev*	-0.736***	-0.500*	-0.982***	-0.708**
	(0.167)	(0.256)	(0.261)	(0.273)
N	2,292	764	764	764
Panel C. Social Knowledge				
Accent × *AccRev*	-0.531***	-0.347	-0.587**	-0.670***
	(0.147)	(0.263)	(0.282)	(0.255)
N	2,280	760	760	760

Note. The table displays the estimated coefficient of *Accent* × *AccRev* in equation (1). The following covariates are included in the estimations but are excluded from the table for brevity: *Accent*,*AccRev*, age, education, and race and ethnicity: Asian and Hispanic (the reference group is White). Columns: All include the sample of the employers' evaluations 1-12 for all workers; Asian include the subsample of the employers' evaluations 1-12 that were Asian workers; Hispanic include the subsample of the employers' evaluations 1-12 that were Hispanic workers; White include the sample of the employers' evaluations 1-12 that were White workers. Standard errors clustered at the employer level in parentheses. N represents the number of observations. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 4.2: Prior Beliefs over Workers' Performance: Conditional on Workers' True Test Score

	(1) All	(2) Asian	(3) Hispanic	(4) White
Panel A. Mathematics Skills				
$Accent \times AccRev$	-0.478***	-0.611**	-0.623**	-0.209
	(0.154)	(0.283)	(0.246)	(0.245)
$TestScore$	0.024	0.003	0.059**	-0.016
	(0.017)	(0.030)	(0.026)	(0.041)
N	2,316	772	772	772
Panel B. Verbal Skills				
$Accent \times AccRev$	-0.739***	-0.501*	-0.993***	-0.707**
	(0.167)	(0.256)	(0.261)	(0.274)
$TestScore$	0.030	0.012	0.102*	0.011
	(0.040)	(0.078)	(0.055)	(0.089)
N	2,292	764	764	764
Panel C. Social Knowledge				
$Accent \times AccRev$	-0.532***	-0.344	-0.582**	-0.672***
	(0.147)	(0.264)	(0.281)	(0.254)
$TestScore$	0.021	0.037	-0.041	0.042
	(0.029)	(0.043)	(0.050)	(0.047)
N	2280	760	760	760

Note. The table displays the estimated coefficients of $Accent \times AccRev$ and $TestScore$ in equation (2). The following covariates are included in the estimations but are excluded from the table for brevity: $Accent$,$AccRev$, age, education, and race and ethnicity: Asian and Hispanic (the reference group is White). Columns: All include the sample of the employers' evaluations 1-12 for all workers; Asian include the subsample of the employers' evaluations 1-12 that were Asian workers; Hispanic include the subsample of the employers' evaluations 1-12 that were Hispanic workers; White include the subsample of the employers' evaluations 1-12 that were White workers. Standard errors clustered at the employer level in parentheses. N represents the number of observations. * p<0.10, ** p<0.05, *** p<0.01.

To explore whether employers' beliefs are accurate, we estimate the prior beliefs conditional on test scores (Equation 2). If the accent gap is due to accurate beliefs, then controlling for the worker's test score should reduce the estimated accent gap. However, this is not what we see. As shown in Table 4.2, the worker's true test score is broadly uncorrelated with the employer's beliefs, and the accent gap in employers' beliefs over workers' performance remains largely unchanged.

Our results suggest that employers might inaccurately use the worker accent as a signal of the worker ability. Interestingly, this accent gap is present even when very little English knowledge is needed such as in our Mathematics test, but, it is worth noting that we find the largest effect on the Verbal skills test. The latter results indicate that employers overestimate the extent to which foreign accent can be used as a signal of English proficiency.[20]

4.2 Does the presence of a performance signal affect employer assessment of worker performance?

Table 4.3 shows the results of estimating Equation 3, which explores the role of a performance signal employer's beliefs over the workers' performance. We find, first, that showing a performance signal updates employers' overall beliefs over workers' performance. Precisely, when the "Random-5" signal is available during the worker's evaluation, employers lower the number of questions they believe the worker answered correctly by -1.58, -1.64, and -1.58 for Math, Verbal, and Social, respectively. Thus, the employers overestimated the workers' performance and adjusted their priors after receiving the performance information.[21] We see qualitatively similar effects across all race/ethnicity groups.

However, the signal does not affect the accent gap in beliefs over performance. Precisely, in the Accent Revealed treatment arm, we see that the signal does not impact the differences in beliefs over performances due to accent. Moreover, the accent gap that we estimated in Table 4.1 still persists after controlling for the presence of the signal. Therefore, the quality signal does not reduce the accent gap in employers' beliefs over worker performance.

[20] As described in Section 2, we gave the worker a paragraph to read, thus keeping everything but the worker's voice constant.

[21] Notably, there is no differential effect of the signal by treatment arm. Results are excluded from the table for brevity.

Table 4.3: Posterior Beliefs over Workers' Performance

	(1) All	(2) Asian	(3) Hispanic	(4) White
Panel A. Mathematics Skills				
Accent × *AccRev* × *Signal*	0.453	0.183	1.005	0.158
	(0.374)	(0.820)	(0.746)	(0.664)
Accent × *AccRev*	-0.474***	-0.603**	-0.598**	-0.211
	(0.153)	(0.283)	(0.246)	(0.245)
Signal	-1.582***	-1.371***	-1.708***	-1.651***
	(0.249)	(0.499)	(0.382)	(0.325)
N	3,474	1,131	1,158	1,185
Panel B. Verbal Skills				
Accent × *AccRev* × *Signal*	-0.074	-0.111	0.140	-0.280
	(0.410)	(0.820)	(0.744)	(0.621)
Accent × *AccRev*	-0.742***	-0.500*	-1.022***	-0.706**
	(0.167)	(0.258)	(0.260)	(0.273)
Signal	-1.645***	-1.463***	-1.217***	-2.286***
	(0.248)	(0.464)	(0.442)	(0.330)
N	3,438	1,130	1,161	1,147
Panel C. Social Knowledge				
Accent × *AccRev* × *Signal*	0.077	-1.678**	1.052	0.633
	(0.363)	(0.771)	(0.659)	(0.603)
Accent × *AccRev*	-0.529***	-0.362	-0.590**	-0.669***
	(0.147)	(0.265)	(0.281)	(0.256)
Signal	-1.582***	-2.129***	-1.269***	-1.370***
	(0.281)	(0.448)	(0.359)	(0.347)
N	3,420	1,106	1,157	1,157

Note. The table displays the estimated coefficients in equation (3). The following covariates are included in the estimations but are excluded from the table for brevity: *Accent*,*AccRev*, *AccRev* × *Signal*, *Accent* × *Signal*, age, education, and race and ethnicity: Asian and Hispanic (the reference group is White). Columns: All include the sample of the employers' evaluations 1-18 for all workers; Asian include the subsample of the employers' evaluations 1-18 that were Asian workers; Hispanic include the subsample of the employers' evaluations 1-18 that were Hispanic workers; White include the subsample of the employers' evaluations 1-18 that were White workers. Standard errors clustered at the employer level in parentheses. N represents the number of observations. * p<0.10, ** p<0.05, *** p<0.01.

4.3 Do employer beliefs related to foreign accent result in wage differences?

The results thus far show that employers believe that accented workers perform worse than non-accented counterparts across all the skills. We next explore whether this bias translates into a pay offers differential by worker accent.

Table 4.4 presents the key results derived from Equation 4. Overall, the analysis reveals no significant accent gap in employers' willingness to pay (WTP) for hiring workers by accent status. The only notable exception is a modest reduction in WTP—13 cents—for accented Hispanic workers in the Social task (Panel C), statistically significant at the 10 percent level.

Therefore, the differences in performance beliefs in Table 4.1–Table 4.3 do not appear to translate into meaningful differences in WTP. One potential explanation for the absence of a pronounced accent WTP gap lies in the structure of the experiment, as detailed in Section 2. Employer WTP not only reflects their valuation of a worker's performance but also directly affects workers since hired individuals receive the offered wage as a bonus. To avoid appearing biased, employers may opt for WTP amounts that do not overtly differentiate based on perceived performance. This pattern is consistent with findings from Coffman *et al.* (2021) and Gioia & Immordino (2024), which found that evaluators in a hiring lab experiment avoided gender-based discrimination when workers' gender was explicitly disclosed, but exhibited bias when workers were categorized by attributes not related to performance (such as birth months).

4.4 Do employer beliefs about Worker bargaining power depend on worker foreign accent?

Table 4.5 shows the results of estimating Equation 5. Our goal here is to explore whether any differences in employers' beliefs about how much workers demand to be compensated by accent status could also potentially affect employers' decisions about worker pay.

Overall, we find no differences on employers' beliefs about Workers' WTA in the UG by accent. The only effect we find, significant at the 10% level, is that employers in the Verbal task treatment believe accented Asians have a lower WTA by 9 cents. Overall, results indicate employers' beliefs about workers' bargaining power do not differ for workers based on whether they have a foreign accent or not.

The findings regarding willingness to accept (WTA) align with two key observations. First, as presented in Table A.2, there is no evidence that workers with accents differ in their WTA compared to non-accented workers. This suggests that employers' beliefs about

workers' WTA accurately reflect workers' actual WTA. Second, Table 4.4 reveals no evidence of a wage gap based on accent. Consequently, there is no indication of differing beliefs about workers' bargaining power. Together, these findings demonstrate consistency between the WTP and WTA results.

Table 4.4: Willingness to Pay (in cents)

	(1) All	(2) Asian	(3) Hispanic	(4) White
Panel A. Mathematics Skills				
Accent × *AccRev*	8.584	8.139	5.625	12.486
	(5.390)	(9.597)	(8.368)	(9.751)
N	1,158	401	427	330
Panel B. Verbal Skills				
Accent × *AccRev*	-8.213	-4.945	-5.942	-12.212
	(5.109)	(9.697)	(8.154)	(8.659)
N	1,146	377	435	334
Panel C. Social Knowledge				
Accent × *AccRev*	-6.322	-0.299	-13.238*	-4.533
	(5.100)	(9.558)	(7.542)	(8.973)
N	1,140	390	422	328

Note. The table displays the estimated coefficient of *Accent* × *AccRev* in equation (4). The following covariates are included in the estimations but are excluded from the table for brevity: *Accent*,*AccRev*, age, education, and race and ethnicity: Asian and Hispanic (the reference group is White). Columns: All include the sample of the employers' evaluations 19-24 for all workers; Asian include the subsample of the employers' evaluations 19-24 that were Asian workers; Hispanic include the subsample of the employers' evaluations 19-24 that were Hispanic workers; White include the subsample of the employers' evaluations 19-24 that were White workers. Standard errors clustered at the employer level in parentheses. N represents the number of observations. * p<0.10, ** p<0.05, *** p<0.01.

Table 4.5: Beliefs about Workers' WTA in the UG (in cents)

	(1)	(2)	(3)	(4)
	All	Asian	Hispanic	White
Panel A. Mathematics Skills				
Accent × *AccRev*	0.331	-1.418	-3.995	4.893
	(3.680)	(5.834)	(5.402)	(7.090)
N	1,158	414	388	356
Panel B. Verbal Skills				
Accent × *AccRev*	-1.162	-8.998*	4.235	0.559
	(3.643)	(4.972)	(6.298)	(7.519)
N	1,146	416	379	351
Panel C. Social Knowledge				
Accent × *AccRev*	0.255	1.051	-4.446	5.215
	(3.432)	(5.286)	(6.190)	(7.351)
N	1,140	431	382	327

Note. The table displays the estimated coefficient of *Accent* × *AccRev* in equation (5). The following covariates are included in the estimations but are excluded from the table for brevity: *Accent*,*AccRev*, age, education, and race and ethnicity: Asian and Hispanic (the reference group is White). Columns: All include the sample of the employers' evaluations 25-30 for all workers; Asian include the subsample of the employers' evaluations 25-30 that were Asian workers; Hispanic include the subsample of the employers' evaluations 25-30 that were Hispanic workers; White include the subsample of the employers' evaluations 25-30 that were White workers. Standard errors clustered at the employer level in parentheses. N represents the number of observations. * p<0.10, ** p<0.05, *** p<0.01.

5 Conclusion

This paper studies the effect of having a foreign accent on employer beliefs about worker productivity and hiring decisions. We use a lab hiring experiment, to disentangle the effect of accent on these decisions from other related characteristics, such as English proficiency

and race/ethnicity. We found that employers exhibit foreign accent bias regarding worker productivity. This bias is present across different skills and different race and ethnicity groups. A signal regarding worker performance does not reduce the accent gap. We find no accent-based differences in employer willingness to pay and beliefs over the worker willingness to accept in an ultimatum game.

This study helps presents novel evidence of discrimination due to a foreign accent in hiring decisions. In future reserach, we plan to build on these findings to further understand the reasons for accent bias and how this gap might arise. We will explore whether we see heterogeneity in the biases across different educational attainments and employer characteristics. Moreover, further analysis is needed to understand why the belief differences do not translate to differences in willingness to pay since revenue depends on workers' scores.

References

Adank, Patti, Evans, Bronwen G, Stuart-Smith, Jane, & Scott, Sophie K. 2009. Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human perception and performance*, **35**(2), 520.

Antonakis, John, d'Adda, Giovanna, Weber, Roberto A, & Zehnder, Christian. 2022. "Just words? Just speeches?" On the economic value of charismatic leadership. *Management Science*, **68**(9), 6355–6381.

Baese-Berk, Melissa M, Bradlow, Ann R, & Wright, Beverly A. 2013. Accent-independent adaptation to foreign accented speech. *The Journal of the acoustical society of America*, **133**(3), EL174–EL180.

Barrett, Rusty, Cramer, Jennifer, & McGowan, Kevin B. 2022. *English with an Accent: Language, Ideology, and Discrimination in the United States*. Routledge.

Becker, Gordon M., Degroot, Morris H., & Marschak, Jacob. 1964. Measuring utility by a single-response sequential method. *Behavioral Science*, **9**(3), 226–232.

Bellemare, Charles, Goussé, Marion, Lacroix, Guy, & Marchand, Steeve. 2023. Physical Disability and Labor Market Discrimination: Evidence from a Video Résumé Field Experiment. *American Economic Journal: Applied Economics*, **15**(4), 452–476.

Bertrand, Marianne, & Mullainathan, Sendhil. 2004. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, **94**(4).

Bohren, Aislinn, Haggag, Kareem, Imas, Alex, & Pope, Devin. 2023. Inaccurate Statistical Discrimination: An Identification Problem. *The Review of Economics and Statistics*, 1–45.

Bradlow, Ann R, & Bent, Tessa. 2008. Perceptual adaptation to non-native speech. *Cognition*, **106**(2), 707–729.

Bureau of Labor Statistics, U.S. Department of Labor. 2023. *Foreign-born workers were a record high 18.1 percent of the U.S. civilian labor force in 2022.*

Burns, Justine, & Keswell, Malcolm. 2015. Diversity and the provision of public goods: Experimental evidence from South Africa. *Journal of Economic Behavior & Organization*, **118**, 110–122.

Caballero, Jonathan A., & Pell, Marc D. 2020. Implicit effects of speaker accents and vocally-expressed confidence on decisions to trust. *Decision*, **7**(4), 314–331.

Chiswick, Barry R, & Miller, Paul W. 1995. The endogeneity between language and earnings: International analyses. *Journal of labor economics*, **13**(2), 246–288.

Coffman, Katherine B, Exley, Christine L, & Niederle, Muriel. 2021. The role of beliefs in driving gender discrimination. *Management Science*, **67**(6), 3551–3569.

Deprez-Sims, Anne-Sophie, & Morris, Scott B. 2010. Accents in the workplace: Their effects during a job interview. *International Journal of Psychology*, **45**(6), 417–426.

Derwing, Tracey M, & Munro, Murray J. 2009. Putting accent in its place: Rethinking obstacles to communication. *Language teaching*, **42**(4), 476–490.

Dietrich, Bryce J, Hayes, Matthew, & O'brien, Diana Z. 2019. Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review*, **113**(4), 941–962.

Eames, Taryn. 2024. Pronoun Disclosure and Hiring Discrimination: A Resume Audit Study. *Available at SSRN*.

Floccia, Caroline, Butler, Joseph, Goslin, Jeremy, & Ellis, Lucy. 2009. Regional and foreign accent processing in English: Can listeners adapt? *Journal of psycholinguistic research*, **38**, 379–412.

Fuertes, Jairo N, Gottdiener, William H, Martin, Helena, Gilbert, Tracey C, & Giles, Howard. 2012. A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, **42**(1), 120–133.

Ge, Qi, & Wu, Stephen. 2024. How do you say your name? Difficult-to-pronounce names and labor market outcomes. *American Economic Journal: Economic Policy*, **16**(4), 254–279.

Gioia, Francesca, & Immordino, Giovanni. 2024. Preference-based hiring decisions and incentives. *Journal of the Economic Science Association*, **10**(1), 1–14.

Gluszek, Agata, & Dovidio, John F. 2010. Speaking with a nonnative accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of language and social psychology*, **29**(2), 224–234.

Güth, Werner, Schmittberger, Rolf, & Schwarze, Bernd. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, **3**(4), 367–388.

Healy, Paul J. 2017. *Epistemic Game Theory Experiments: Utility Elicitation and Irrational Play*. Tech. rept. mimeo.

Hossain, Tanjim, & Okui, Ryo. 2013. The Binarized Scoring Rule. *The Review of Economic Studies*, **80**(3), 984–1001.

Huang, Laura, Frideger, Marcia, & Pearce, Jone L. 2013. Political skill: Explaining the effects of nonnative accent on managerial hiring and entrepreneurial investment decisions. *Journal of Applied Psychology*, **98**(6), 1005.

Isphording, Ingo E, & Sinning, Mathias. 2012. The returns to language skills in the US labor market. *Ruhr Economic Paper*.

Krupka, Erin L., & Weber, Roberto A. 2013. Identifying Social Norms using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association*, **11**(3), 495–524.

Li, Mingsheng, & Campbell, Jacqui. 2009. Accessing employment: Challenges faced by non-native English-speaking professional migrants. *Asian and Pacific Migration Journal*, **18**(3), 371–395.

Munro, Murray J, & Derwing, Tracey M. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, **45**(1), 73–97.

Munro, Murray J, Derwing, Tracey M, & Sato, Kyoko. 2006. Salient accents, covert attitudes: Consciousness-raising for pre-service second language teachers. *Prospect: an Australian journal of TESOL*, **21**(1), 65–77.

Oreopoulos, Philip. 2011. Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes. *American Economic Journal: Economic Policy*, **3**(4), 148–171.

Rubin, Donald L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher education*, **33**, 511–531.

Ruggles, Steven, Flood, Sarah, Sobek, Matthew, Backman, Daniel, Chen, Annie, Cooper, Grace, Richards, Stephanie, Rodgers, Renae, & Schouweiler, Megan. 2024. *IPUMS USA: Version 15.0, American Community Survey 2022*.

Schmaus, Miriam, & Kristen, Cornelia. 2022. Foreign accents in the early hiring process: A field experiment on accent-related ethnic discrimination in Germany. *International Migration Review*, **56**(2), 562–593.

Selten, Reinhard. 1967. Die strategiemethode zur erforschung des eingeschraenkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. *Beitraege zur experimentellen Wirtschaftsforschung*, **1**, 136–168.

Shuck, Gail. 2004. Conversational performance and the poetic construction of an ideology. *Language in society*, **33**(2), 195–222.

Shuck, Gail. 2006. Racializing the nonnative English speaker. *Journal of language, identity, and education*, **5**(4), 259–276.

Spence, Jessica L, Hornsey, Matthew J, & Imuta, Kana. 2021. Something about the way you speak: A meta-analysis on children's linguistic-based social preferences. *Child Development*, **92**(2), 517–535.

Spence, Jessica L., Hornsey, Matthew J., Stephenson, Eloise M., & Imuta, Kana. 2024. Is Your Accent Right for the Job? A Meta-Analysis on Accent Bias in Hiring Decisions. *Personality and Social Psychology Bulletin*, **50**(3), 371–386.

Tenzer, Helene, Pudelko, Markus, & Harzing, Anne-Wil. 2014. The impact of language barriers on trust formation in multinational teams. *Journal of International Business Studies*, **45**, 508–535.

Tombs, Alastair, & Rao Hill, Sally. 2014. The effect of service employees' accent on customer reactions. *European Journal of Marketing*, **48**(11/12), 2051–2070.

Tomlin, Bryan. 2024. Pronoun usage and gender identity's effects on market outcomes: Evidence from a preregistered field experiment. *Economics Letters*, **236**.

van Wijngaarden, Sander J, Steeneken, Herman JM, & Houtgast, Tammo. 2001. The effect of a non-native accent in Dutch on speech intelligibility. *The Journal of the Acoustical Society of America*, **109**(5_Supplement), 2473–2473.

Zuckerman, Miron, & Driver, Robert E. 1989. What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, **13**(2), 67–82.

Appendix

A.1 Figures

Figure A.1: Example of Test Questions: Mathematics Skills

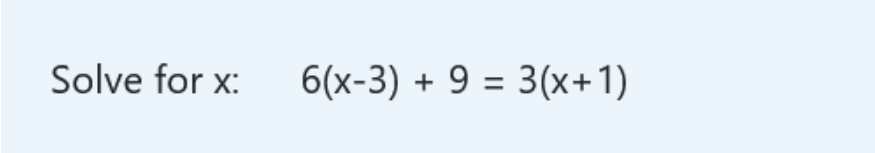


Figure A.2: Example of Test Questions: Verbal Skills

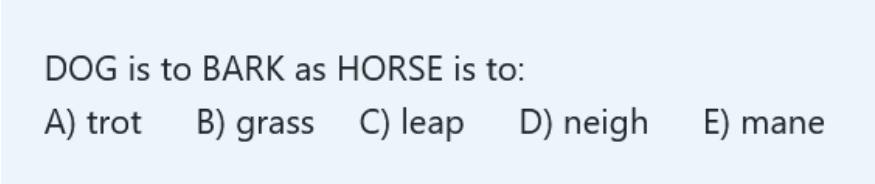


Figure A.3: Example of Test Questions: Knowledge of Social Norms

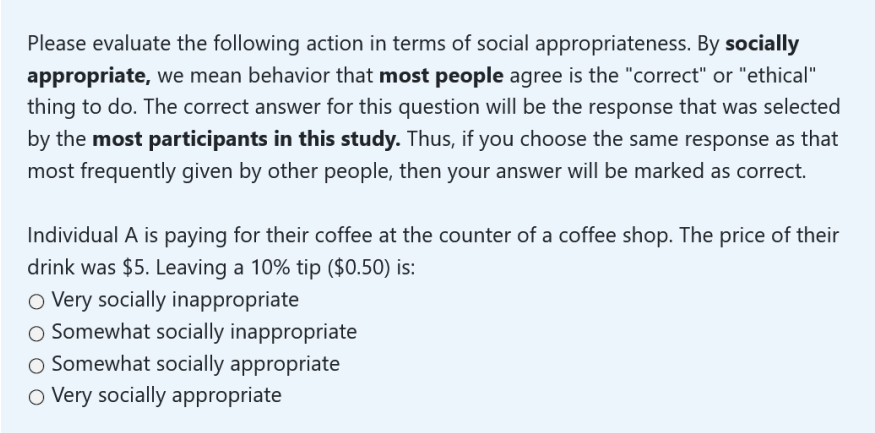


Figure A.4: Example of Signal



A.2 Example: Voice Recording and Accent



ELSA score for the voice-transformed (**speech-to-speech**) recording:

- **ELSA Pronunciation:** 80%
- **ELSA Intonation:** 77%

A.3 Tables

Table A.1: Sample Description by Accent Status

	(1) All	(2) Asian	(3) Hispanic	(4) White
		No Accent		
Age	33.19	28.83	31.33	39.42
	(9.21)	(6.78)	(7.19)	(10.26)
College or More	0.61	0.67	0.50	0.67
	(0.49)	(0.49)	(0.52)	(0.49)
Math Score	7.53	9.42	6.67	6.50
	(3.35)	(3.29)	(3.23)	(2.94)
Verbal Score	10.31	10.92	9.92	10.08
	(1.82)	(1.83)	(2.15)	(1.38)
Social Score	7.81	7.67	7.83	7.92
	(2.44)	(3.11)	(2.04)	(2.23)
N	36	12	12	12
		Accent		
Age	37.08	33.00	37.42	40.83
	(11.17)	(12.11)	(11.40)	(9.31)
College or More	0.58	0.83	0.50	0.42
	(0.50)	(0.39)	(0.52)	(0.51)
Math Score	6.61	9.25	4.83	5.75
	(3.69)	(3.39)	(3.30)	(3.02)
Verbal Score	9.50	10.08	9.67	8.75
	(1.80)	(2.15)	(1.44)	(1.60)
Social Score	6.89	7.25	7.00	6.42
	(2.14)	(2.18)	(1.81)	(2.47)
N	36	12	12	12

Note. This table presents the mean and standard deviation (in parenthesis) of the sample of workers by Accent Status. N represents the number of observations.

Table A.2: Relationship Between Accent and Worker WTA in the Ultimatum Game

	(1) All	(2) Asian	(3) Hispanic	(4) White
Accent	8.639	-0.167	8.583	17.500
	(8.936)	(15.085)	(15.618)	(16.586)
N	72	24	24	24

Note. This table estimates the effect of accent on the willingness to accept on the ultimatum game in the workers' sample. N represents the number of observations.