Smile Inequality: Oral Health and Labor Market Outcomes

Gursmeep (Rubina) Hundal, University of Chicago

31 August 2025

Abstract

This paper studies how a visible but remediable attribute, dental appearance, shapes labor market access for low-income youth in India. I combine three complementary randomized experiments: (1) an online **perception survey** ($N \approx 514$) eliciting trait attributions from AI-generated headshots; (2) a large-scale **correspondence study** ($N \approx 6{,}000$ job applications) varying oral health, gender, and caste signals; and (3) an **incentivized resume-rating** experiment ($N \approx 400$ HR professionals) that recovers not only hiring scores but also forecasts of customer, coworker, and industry reactions under a proper scoring rule. Across settings, poor oral appearance consistently shifts beliefs: survey respondents predict $\approx 9\%$ lower salaries for "bad teeth" profiles, and HR professionals penalize them by about 1.8 points on a 0–10 hireability scale. The IRR further shows that these penalties operate through customer-facing and coworker channels: forecasted client discomfort and peer aversion account for roughly one-third of the measured hiring penalty. In the field, these appearance effects are muted on average because entrenched caste hierarchies dominate callback decisions, but they resurface sharply in contexts where employers screen women and customer-facing applicants more closely. The findings advance our understanding of when treatable health signals matter in labor markets: visible oral health deficits systematically worsen first impressions and recruiter evaluations, yet their translation into realized exclusion depends on whether other identity-based signals or role requirements leave room for appearance-based screening.

1 Introduction

Visible oral health, the appearance of teeth in a job applicant's headshot, is a common, easily observed attribute that is also largely treatable. Yet for many low-income youth, unmet dental needs are persistent, and even a single visual cue can influence whether an applicant receives an interview.

This paper examines how such visible health-related signals shape early-stage hiring in urban India and how their effects depend on social hierarchies and hiring contexts.

We combine three complementary randomized designs. First, a within-subject **perception survey** ($N \approx 514$) elicits how AI-generated headshots that vary in dental appearance (and other visual traits) shift attributions of trust, confidence, diligence, and predicted earnings. Second, an **incentivized resume-rating experiment (IRR)** with nearly 400 HR professionals measures not only hiring scores but also incentivized forecasts of how customers, coworkers, and other employers would respond, allowing a direct decomposition of informational updating, taste-based preferences, and anticipated third-party reactions. Third, a large-scale **correspondence study** ($N \approx 6{,}000$ applications) submits fictitious resumes to real entry-level postings, randomizing oral-health signals, caste, and gender. Together, these three designs trace the pathway from first impressions to employer callbacks and illuminate the mechanisms by which appearance influences labor-market access.

This study makes three contributions. First, it introduces **oral health as a remediable signaling margin** in labor economics, bridging human-capital and appearance-based discrimination frameworks. While prior work highlights persistent returns to attractiveness, we show that a specific, treatable health attribute i.e., dental appearance, systematically alters evaluations. Second, we develop and empirically test a **screening model** (inspired by Arrow, 1973; Phelps, 1972) in which employer responses reflect a weighted mix of productivity beliefs (δ), personal distaste (τ), and anticipated customer/coworker reactions (κ). By combining a perception survey, a field experiment, and an incentivized HR survey, we provide unusually sharp identification of which channels matter most. Third, we offer new evidence on **intersectional discrimination**: oral-health penalties interact with gender and caste signals, revealing when treatable appearance deficits translate into exclusion and when entrenched hierarchies dominate.

The key findings are threefold. (i) In the perception survey, poor oral appearance substantially lowers ratings of trust, confidence, and diligence, and reduces predicted salaries by roughly 9%. (ii) In the IRR, HR professionals penalize candidates with visibly poor teeth by about 1.8 points on a 0–10 hireability scale, and forecast significantly worse customer and coworker reactions—channels that account for roughly one-third of the total penalty. These results provide direct evidence of both taste-based and third-party mechanisms. (iii) In the correspondence study, caste cues remain the strongest axis of exclusion, with high-caste surnames receiving callbacks nearly 5 percentage points above Dalit-sounding names. The oral-health penalty is muted on average, but reappears in specific contexts: among female applicants and in customer-facing jobs where appearance matters most.

Together these patterns highlight that visible oral health consistently worsens first impressions and professional evaluations, yet its translation into realized exclusion depends on whether other identity-based signals leave room for appearance-based screening.

The remainder of the paper proceeds as follows. Section 2 situates this work within literatures on discrimination, appearance, and health. Section 3 describes the experimental designs. Section 4 sets out the screening framework and Section 5 presents empirical findings followed by the empirical test of the screening model. Section 7 discusses policy implications, and Section 8 concludes.

2 Related literature

This paper sits at the intersection of three literatures: (i) statistical and taste-based discrimination in labor markets, (ii) the economics of health capital and appearance, and (iii) the empirical literature that uses audit and correspondence methods to measure hiring discrimination.

A large theoretical and empirical literature documents how observable characteristics shape employer decisions. Classic models highlight statistical discrimination, where imperfect signals lead employers to update beliefs about productivity (Phelps, 1972; Arrow, 1973), and taste-based models where employers act on preferences or prejudice. Experimental audit and correspondence studies have provided clean evidence that names, addresses, and photos affect callbacks and interviews (Bertrand and Mullainathan, 2004; Banerjee et al., 2009; Neumark, 2018). Recent work emphasizes intersectionality and dynamic accumulation of disadvantage across hiring stages (McLaughlin and Neumark, 2025; Bohren et al., 2022).

A parallel literature emphasizes that health and appearance affect labor-market returns. Grossman's framework treats health as productive capital (Grossman, 1972); empirical work links improved dental health to higher earnings in some settings (Glied and Neidell, 2010; Gallego et al., 2024; Singhal et al., 2016). Studies of attractiveness find persistent wage returns to physical appearance (Siermińska and Singhal, 2023; Rooth, 2009). Our paper contributes by focusing on oral health as a visible, treatable attribute that operates through both health-capital and signaling channels.

Psychology and behavioral economics document rapid and systematic trait inferences from faces (Willis and Todorov, 2006; Todorov and Oosterhof, 2011). These first impressions predict consequential outcomes ranging from elections to judicial decisions (Olivola et al., 2014; Antonakis and Dalgas, 2009). Recent economics work uses incentivized experiments to elicit hiring preferences in a deception-free way (Kessler et al., 2019); we build on that IRR innovation to measure how professional recruiters update beliefs and forecast others' responses in the presence of oral-health cues.

The three-experiment design (perception survey, correspondence study, IRR) allows us to observe (a) the perceptual penalties associated with poor oral appearance, (b) whether perceptual penalties translate into callback gaps in real hiring, and (c) how HR professionals, when incentivized to

reveal preferences and forecasts, process oral-health signals. This structure permits more precise identification of mechanisms than any single method alone and shows when a visible, treatable health signal has labor-market consequences and when stronger identity markers (notably caste and gender) dominate screening decisions.

3 Methodology

3.1 Perception Study

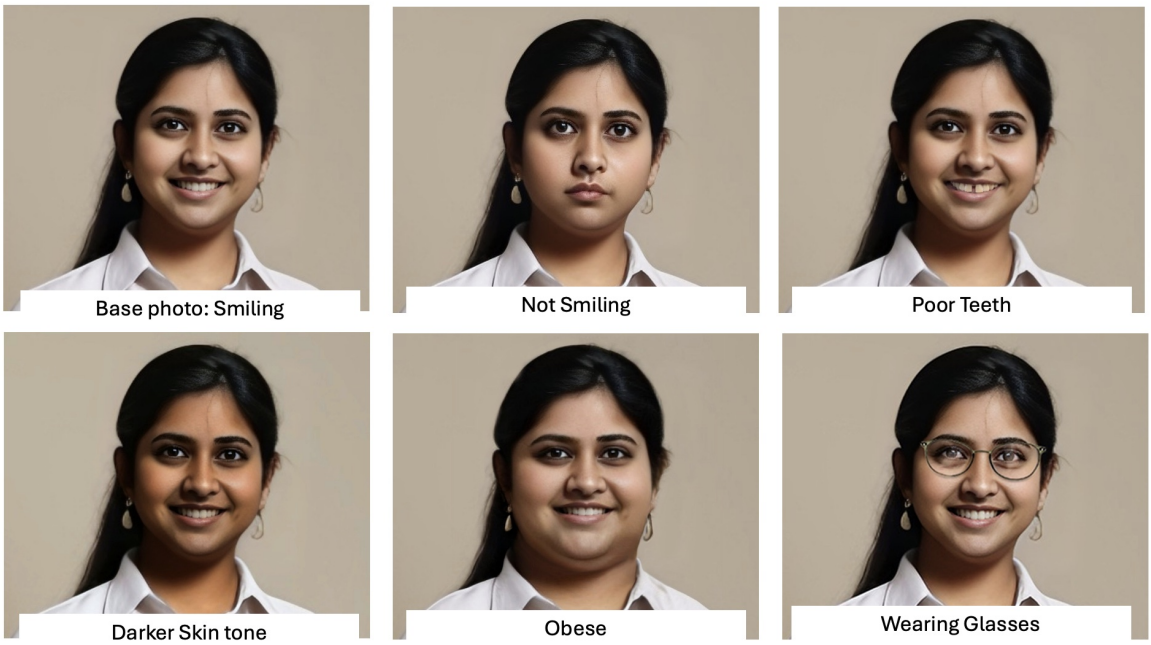
Overview and sample. The perception survey was administered on the Centiment/Qualtrics platform. After excluding respondents who failed embedded attention checks, the final analysis sample contains $N = 514$ respondents. Eligibility criteria required respondents to be adults (18+); no geographic or income quotas were imposed. Respondents completed the survey on desktop or mobile devices.

Stimuli and randomization. Each respondent viewed six AI-generated headshots (three male, three female). For each identity the respondent saw *exactly one* visual variant drawn *without replacement* from the set

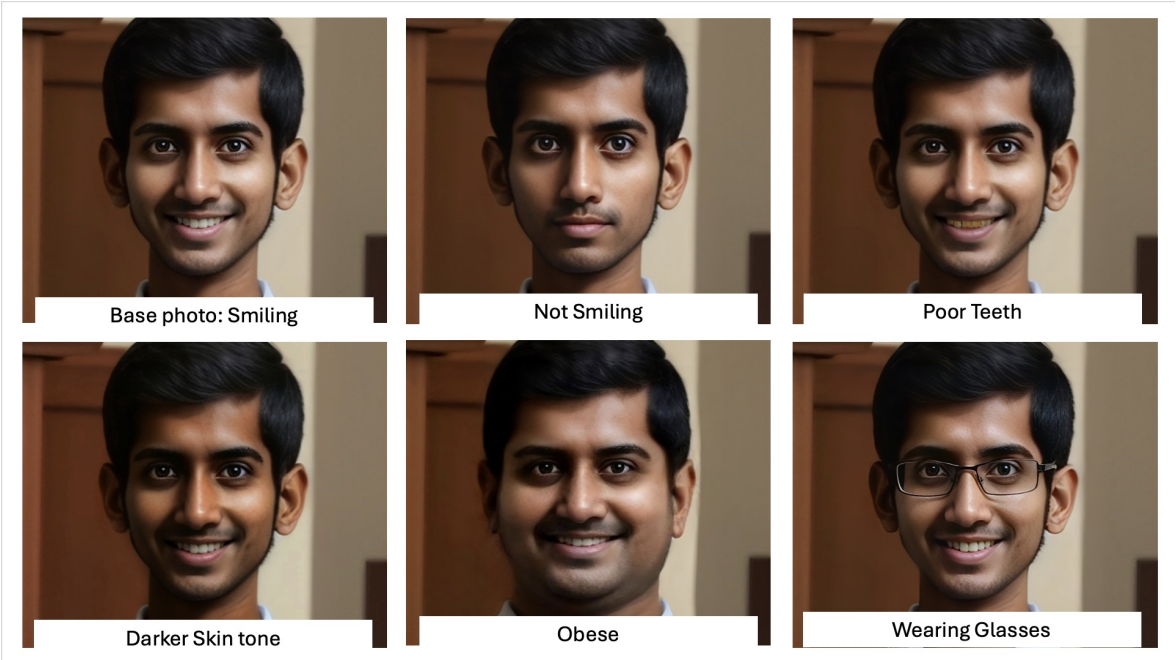
- base smiling
- neutral (no smile)
- smiling + poor teeth
- smiling + obese
- smiling + darker skin
- smiling + spectacles

Since variants are drawn without replacement, each respondent sees every treatment type exactly once across the six identities (i.e., no variant is shown more than once to the same respondent). The order of photo presentation was randomized independently for each respondent.

An additional short *qualification blurb* (e.g., "Completed a diploma in Office Management") was randomized at the respondent level: for half of respondents the qualification blurb was appended to exactly 3 of the 6 profiles (the 3 profiles selected uniformly at random), while the other half of respondents received no qualification blurbs. This between-respondent randomization of the blurb allows us to test whether minimal textual information attenuates visual penalties.



(a) Female identity: Base smiling; Not smiling; Poor teeth; Darker skin tone; Obese; Wearing glasses.



(b) Male identity: Base smiling; Not smiling; Poor teeth; Darker skin tone; Obese; Wearing glasses.

Figure 1: Example image variants shown to respondents in the perception survey (AI-generated; no real persons).

Stimuli validation. All headshots are AI-generated composite portraits; no photograph of a real person was used. The stimuli were produced to be age- and style-consistent, and a pilot confirmed that the base smiling and neutral variants were perceived as plausible and that the manipulations (poor dental appearance, obesity, darker skin, spectacles) were reliably perceived by independent raters.

Tasks and recorded outcomes. For each presented photo (identity + assigned visual variant + possibly a qualification blurb) respondents completed the following tasks:

1. **Trait ratings.** Eleven trait ratings on 0–10 sliders (integer responses enforced):
    - Trustworthiness
    - Attractiveness
    - Perceived wealth ("Rich")
    - Confidence
    - Likelihood of being highly educated
    - Healthy lifestyle
    - Hardworking / Diligent
    - Intelligence
    - Similarity-to-self (age/gender/looks)
    - Suitability for backend role (e.g., data entry)
    - Suitability for customer-facing role (e.g., receptionist, sales)

    Salary predictions: transformations and robustness. Salary forecasts are collected as open numeric monthly INR values. Our main specification uses log(monthly INR + 1) (log-level); robustness checks analyze levels, winsorized levels (1st/99th percentiles), and trimmed samples excluding extreme outliers. We will also report median regression estimates for skew-robust inference.

After rating all six profiles respondents completed a brief demographic module (gender, age, marital status, occupation, household income bracket, highest education completed, and state/union territory of residence).

Quality control. An attention-check question is embedded. Any respondent failing the attention check was excluded from the final sample (the $N = 514$ figure is after these exclusions). We also enforced monotone response checks and basic time thresholds to exclude implausible completions.

Primary and secondary perception outcomes (pre-analysis). Our pre-specified *primary perception outcome* is an *employability index* defined as the standardized mean of Trustworthiness,

Diligence, and Confidence (each standardized to mean zero, unit variance before averaging). Secondary outcomes include the eleven individual trait ratings listed above and the respondent's salary prediction.

Index construction and reliability. The pre-specified employability index equals the standardized mean of Trustworthiness, Diligence, and Confidence (each variable standardized to mean zero and unit variance before averaging).

Data handling for salary predictions. Respondents enter open numeric salary predictions, so for the main analyses we will use log(monthly INR) as the primary specification. As robustness checks we (i) analyze salaries in levels (INR), (ii) winsorize salary predictions at the 1st and 99th percentiles, and (iii) trim extreme outliers. We record both the level and the natural log so that alternative specifications are straightforward.

Randomization integrity and balancing checks. We report balance tables across respondent characteristics (age, gender, education, state, income bracket) for: (a) the assignment of each visual variant to identities (which, by design, should be orthogonal across respondents), and (b) the qualification-blurb assignment (3-of-6 vs. none). We also randomize photo presentation order and will control for presentation position in robustness checks.

Primary regression specification. Our main within-respondent specification is

$$y_{ir} = \alpha + \beta\,\mathrm{PoorOral}_{ir} + \sum_{t} \delta_t T_{ir}^{t} + \phi_i + \psi_r + \varepsilon_{ir},$$

where y_{ir} is the trait rating (or the standardized employability index) provided by respondent r for photo i; $\mathrm{PoorOral}_{ir}$ indicates the poor-teeth variant; T_{ir}^{t} are indicators for the other visual treatments; ϕ_i are photo (identity) fixed effects; and ψ_r are respondent fixed effects. Standard errors are clustered at the respondent level. As robustness checks we (i) estimate ordered probit/logit models for bounded 0–10 outcomes, (ii) run OLS without respondent fixed effects, and (iii) report heteroskedasticity-robust standard errors.

Notes on inference and clustering. Each respondent provides multiple (six) ratings, so inference in perception regressions will cluster standard errors at the respondent level. In the analysis section we detail the exact regression specifications and fixed effects; briefly, within-respondent regressions with photo and respondent fixed effects will be used to estimate the causal effect of visual treatments on trait ratings and salary predictions.

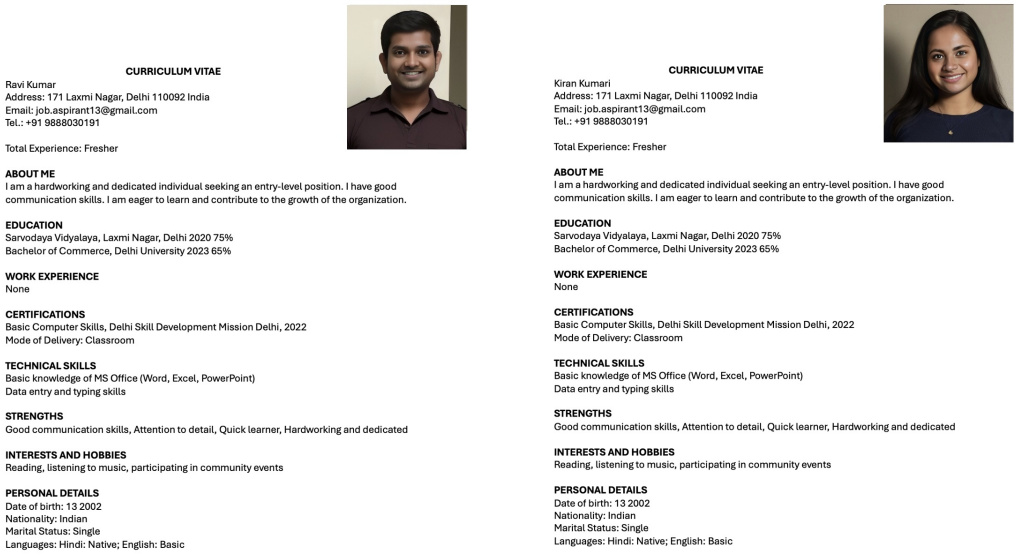
3.2 Correspondence Study

Design and realized sample. We run a large-scale correspondence (audit) experiment that submits one fictitious application per employer to genuine entry-level job postings. The realized sample comprises $N = 6{,}066$ submitted applications. Each employer received at most one application to reduce detection risk and to identify average treatment effects across a broad set of real vacancies.

Experimental margins. Each application varied along three randomized margins:

1. **Applicant gender**, signaled by first name (ten name templates: five male-typical, five female-typical; vernacular and religion-neutral).
2. **Caste signal**, signaled by surname format (high-caste, low-caste/Dalit-sounding, neutral such as "Kumar/Kumari," initial-only, or first-name only).
3. **Oral-health / photo condition**. For postings that accepted photos we randomized whether to include a headshot and, conditional on including one, which visual variant to attach (base smiling / good teeth; smiling with visibly poor teeth; and auxiliary variants used for secondary analyses). For postings that did not accept photos we submitted text-only résumés.

Sampling frame and submission protocol. Vacancies were scraped weekly from `Indeed.in` (entry-level / "fresher" categories) across Indian states during the pre-specified window (1 January–31 October 2025). We restricted to postings that plausibly matched our résumé templates and permitted online application. For each posting we recorded job title, posting date, location (state), advertised salary (when present), whether the posting accepted photos, and other relevant metadata. Resumes were substantively identical except for the randomized margins and were tailored at a surface level to the advertised role (e.g., Commerce-leaning phrasing for retail/reception jobs) to avoid obvious mismatch. Submissions were staggered in time and used realistic contact details.



CURRICULUM VITAE

Ravi Kumar
Address: 171 Laxmi Nagar, Delhi 110092 India
Email: job.aspirant13@gmail.com
Tel.: +91 9888030191

Total Experience: Fresher

ABOUT ME
I am a hardworking and dedicated individual seeking an entry-level position. I have good communication skills. I am eager to learn and contribute to the growth of the organization.

EDUCATION
Sarvodaya Vidyalaya, Laxmi Nagar, Delhi 2020 75%
Bachelor of Commerce, Delhi University 2023 65%

WORK EXPERIENCE
None

CERTIFICATIONS
Basic Computer Skills, Delhi Skill Development Mission Delhi, 2022
Mode of Delivery: Classroom

TECHNICAL SKILLS
Basic knowledge of MS Office (Word, Excel, PowerPoint)
Data entry and typing skills

STRENGTHS
Good communication skills, Attention to detail, Quick learner, Hardworking and dedicated

INTERESTS AND HOBBIES
Reading, listening to music, participating in community events

PERSONAL DETAILS
Date of birth: 13 2002
Nationality: Indian
Marital Status: Single
Languages: Hindi: Native; English: Basic

CURRICULUM VITAE

Kiran Kumari
Address: 171 Laxmi Nagar, Delhi 110092 India
Email: job.aspirant13@gmail.com
Tel.: +91 9888030191

Total Experience: Fresher

ABOUT ME
I am a hardworking and dedicated individual seeking an entry-level position. I have good communication skills. I am eager to learn and contribute to the growth of the organization.

EDUCATION
Sarvodaya Vidyalaya, Laxmi Nagar, Delhi 2020 75%
Bachelor of Commerce, Delhi University 2023 65%

WORK EXPERIENCE
None

CERTIFICATIONS
Basic Computer Skills, Delhi Skill Development Mission Delhi, 2022
Mode of Delivery: Classroom

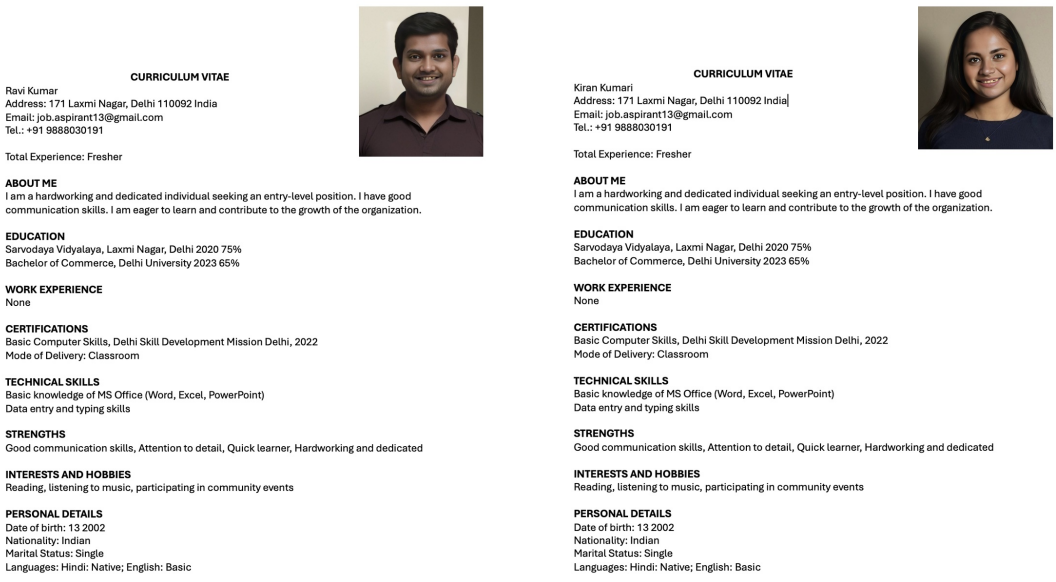
TECHNICAL SKILLS
Basic knowledge of MS Office (Word, Excel, PowerPoint)
Data entry and typing skills

STRENGTHS
Good communication skills, Attention to detail, Quick learner, Hardworking and dedicated

INTERESTS AND HOBBIES
Reading, listening to music, participating in community events

PERSONAL DETAILS
Date of birth: 13 2002
Nationality: Indian
Marital Status: Single
Languages: Hindi: Native; English: Basic

(a) Headshot with healthy teeth



CURRICULUM VITAE

Ravi Kumar
Address: 171 Laxmi Nagar, Delhi 110092 India
Email: job.aspirant13@gmail.com
Tel.: +91 9888030191

Total Experience: Fresher

ABOUT ME
I am a hardworking and dedicated individual seeking an entry-level position. I have good communication skills. I am eager to learn and contribute to the growth of the organization.

EDUCATION
Sarvodaya Vidyalaya, Laxmi Nagar, Delhi 2020 75%
Bachelor of Commerce, Delhi University 2023 65%

WORK EXPERIENCE
None

CERTIFICATIONS
Basic Computer Skills, Delhi Skill Development Mission Delhi, 2022
Mode of Delivery: Classroom

TECHNICAL SKILLS
Basic knowledge of MS Office (Word, Excel, PowerPoint)
Data entry and typing skills

STRENGTHS
Good communication skills, Attention to detail, Quick learner, Hardworking and dedicated

INTERESTS AND HOBBIES
Reading, listening to music, participating in community events

PERSONAL DETAILS
Date of birth: 13 2002
Nationality: Indian
Marital Status: Single
Languages: Hindi: Native; English: Basic

CURRICULUM VITAE

Kiran Kumari
Address: 171 Laxmi Nagar, Delhi 110092 India
Email: job.aspirant13@gmail.com
Tel.: +91 9888030191

Total Experience: Fresher

ABOUT ME
I am a hardworking and dedicated individual seeking an entry-level position. I have good communication skills. I am eager to learn and contribute to the growth of the organization.

EDUCATION
Sarvodaya Vidyalaya, Laxmi Nagar, Delhi 2020 75%
Bachelor of Commerce, Delhi University 2023 65%

WORK EXPERIENCE
None

CERTIFICATIONS
Basic Computer Skills, Delhi Skill Development Mission Delhi, 2022
Mode of Delivery: Classroom

TECHNICAL SKILLS
Basic knowledge of MS Office (Word, Excel, PowerPoint)
Data entry and typing skills

STRENGTHS
Good communication skills, Attention to detail, Quick learner, Hardworking and dedicated

INTERESTS AND HOBBIES
Reading, listening to music, participating in community events

PERSONAL DETAILS
Date of birth: 13 2002
Nationality: Indian
Marital Status: Single
Languages: Hindi: Native; English: Basic

(b) Headshot with visibly poor teeth

Figure 2: Example résumés used in the correspondence study. Only the headshot condition varies; text content is held constant apart from the randomized name/surname signals.

Outcomes and coding. The primary outcome is `Callback`, an indicator equal to 1 if a *positive* employer response (invitation to interview/test, request for materials that signals further consideration, or an offer) is received within **4–6 weeks** of submission, and 0 otherwise. Secondary outcomes include time-to-first-positive-response (days), response quality (interview invitation vs. informational request), any response (positive or negative), and requests for additional documents. Non-response or absence of a positive reply within the window is coded as `Callback`=0. All incoming messages are archived and coded by type and timestamp in the central registry.

Identification, threats, and mitigation. Identification of average treatment effects relies on the random assignment of name, surname format, and photo condition across eligible postings. Key threats are differential selection of job-types that accept photos and potential correlation between posting timing and observable job characteristics. We mitigate these threats by: (i) recording rich posting-level covariates (advertised salary, job title, state, acceptance of photos) and (ii) controlling for posting-date and state-by-role-type fixed effects in adjusted specifications. We also staggered submissions and limited to one application per employer to minimise detection and spillover.

Empirical specification and inference. Our baseline linear-probability specification is

$$\text{Callback}_j = \alpha + \beta\,\text{PoorOral}_j + \gamma\,\text{LowCaste}_j + \theta\,\text{Female}_j + \mathbf{Z}_j'\eta + \mu_{s(j),k(j)} + \varepsilon_j,$$

where PoorOral_j, LowCaste_j, and Female_j are treatment indicators; $\mathbf{Z}_j$ contains posting-level controls (salary midpoint and salary-missing indicator, posting-date dummies, simple job-title similarity controls, and photo-level mean trait ratings when available); and $\mu_{s,k}$ denotes state×role-type fixed effects. Standard errors are clustered at the posting (employer) level. Robustness checks include logit specifications, conditional logit on matched posting pairs (when feasible), re-weighting to address realized imbalances, and specifications that interact treatment indicators with indicators for whether a posting accepted photos.

Heterogeneity and minimum-cell rules. We pre-specify heterogeneity analyses by role type (customer-facing vs. back-office vs. remote), state (Delhi NCR, Maharashtra, other states), gender, and caste.

Randomization checks, balance, and power. We present balance tables for posting-level observables by treatment assignment and document any deviations from perfect balance. With the realized $N \approx 6{,}066$ and an observed baseline callback rate near 15% (see tracker), the study is powered to detect modest absolute differences between balanced arms.

Preregistered rules and exploratory analyses. Primary analyses follow the pre-specified outcomes and coding rules; multiple-testing adjustments are applied within outcome families. We retain the flexibility to conduct additional, clearly labeled exploratory analyses (e.g., alternative covariate sets, machine-learning heterogeneous-effect discovery) so long as they are reported as exploratory and not substituted for the pre-specified inferential claims.

3.3 Incentivized Resume–Rating Experiment

We run an incentivized resume–rating (IRR) experiment with hiring professionals to decompose how visible and textual signals (oral health, gender, caste, marital status, and credentials) affect evaluative hiring judgments. The pre-registered target sample is approximately $N = 250$ HR professionals. Each respondent evaluates **eight** candidate profiles (six canonical photo profiles plus two additional text-only profiles drawn at random). Image assets, and the IRR questionnaire are archived in the Appendix.

Presentation, sample, and randomization. Profiles are shown as short *profile cards* (resume snapshots). Within each respondent, **cards 1–6 are photo resumes** and **cards 7–12 are text-only resumes**. Every respondent sees all six photo cards (one per AI identity with the randomized visual variant) and **two text-only cards drawn uniformly at random without replacement from cards 7–12**. Gender and caste signals come from first name and surname format; remaining resume elements (marital status, certification, degree type, minor phrasing) are randomized across cards. No AI face is repeated within a respondent.[1]

Photo manipulations and profile dimensions. Photo resumes are assigned one of four visual variants with the following within-respondent allocation: (i) healthy oral appearance ("good teeth"), (ii) visibly poor oral appearance ("bad teeth"), (iii) obese appearance, and (iv) darker skin tone. The "good teeth" and "bad teeth" variants each appear twice per respondent; the "obese" and "darker skin" variants each appear once per respondent. Combined with the two text-only profiles, this yields 8 profiles per recruiter (6 photo + 2 text-only).Other profile margins (randomized orthogonally across profiles) include:

- **Marital status:** married / unmarried / omitted (equal probability).
- **Certification:** present / absent (50%).
- **Degree type:** BSc / BCom / BA (equal probability).

[1] In the pilot, many HR respondents declined to rate non-photo profiles, noting that they either do not receive such resumes, reject them, or ask the candidate to send a photo. We therefore revised the original plan of 6 photo + 6 non-photo cards to the implemented design of 6 photo + 2 non-photo cards.

Name	Gender	Caste signal	Assigned photo ID
Anmol Trivedi	Male	High	Pic1
Mehak Paswan	Female	Low	Pic2
Sameer Ambedkar	Male	Low	Pic3
Anjali Sharma	Female	High	Pic4
Varun K.	Male	Neutral	Pic5
Diya S.	Female	Neutral	Pic6
Two profiles from the following six are randomly shown (text-only):			
Ravi Joshi	Male	High	None
Sagar Valmiki	Male	Low	None
Rahul H.	Male	Neutral	None
Nisha Tripathi	Female	High	None
Tara Pasi	Female	Low	None
Muskan L.	Female	Neutral	None

Table 1: Canonical names and assigned base photos (six canonical + six supplemental text-only).

- **Fixed provenance:** institution names and addresses chosen to be plausible for entry-level roles from low-cost local colleges.

No candidate face is repeated within a respondent, preserving independence of facial identities in within-subject comparisons.

Tasks, incentives, and payments. For each card, respondents provide:

- **Hire score (primary outcome):** "How suitable is this candidate to move to the next stage of the hiring process?" on a 0–10 scale.
- **Belief forecasts (incentivized):** five forecasts designed to diagnose mechanisms:
  1. *Industry norms channel:* beliefs about how other employers would treat the same candidate in **three role families** — front-end customer-facing roles, back-end non-customer-facing roles, and remote roles.
  2. *Customer acceptance channel:* beliefs about how customers would feel about interacting with the candidate in customer-facing roles.
  3. *Coworker acceptance channel:* beliefs about how coworkers would feel about working with the candidate.

Two features promote truthful reporting. First, respondents were told that we may share high-scoring profiles with firms that hire for fresher roles, which gives a direct screening-relevance incentive for accurate Hire scores. Second, the *industry-norms* forecasts (three role families) and the *customer* forecast are **paid under a quadratic scoring rule** with external benchmarks (up to INR 50 per forecast dimension). The *coworker* item is recorded but **not** monetarily incentivized. The Hire score

itself is not paid directly.

Estimation and inference. The primary IRR regression is

$$\text{HireScore}_{ir} = \alpha + \beta\,\text{PoorOral}_{ir} + \mathbf{W}'_{ir}\lambda + \phi_r + \varepsilon_{ir},$$

where PoorOral_{ir} indicates the visibly poor-teeth photo for profile i as viewed by recruiter r, $\mathbf{W}_{ir}$ contains profile-level controls (certification, degree type, marital-status indicators, photo/no-photo and visual-variant dummies), and ϕ_r are recruiter fixed effects. Because the HireScore is bounded on $[0, 10]$ we report OLS estimates (interpretable as point changes on the 0–10 scale) and present ordered-logit / fractional-response specifications as robustness checks; marginal effects from nonlinear models are shown alongside OLS. Standard errors are clustered at the recruiter level.

Belief-elicitation outcomes are analyzed using both raw forecast errors and incentivized-score metrics; mediation and reduced-form decompositions that exploit orthogonal profile variation are used to attribute portions of the total oral-health effect to informational, customer, and coworker channels (analysis details in Section **??**).

Pre-specified outcomes, multiple testing, and exploratory analyses. The pre-specified primary IRR outcome is the recruiter-level mean `HireScore` difference by oral-health treatment. Secondary outcomes include incentivized forecast scores and the coworker-comfort item. Within IRR outcome families we adjust for multiple testing using the Benjamini–Hochberg false-discovery-rate procedure (Benjamini and Hochberg, 1995). We retain the flexibility to conduct clearly-labeled exploratory analyses (e.g., alternative covariate sets, machine-learning heterogeneity searches); exploratory analyses will be reported as such and not substituted for the pre-specified inferential claims.

Photo condition (profile-level)	**Probability of appearance**
No photo (text-only)	25% of profiles seen per respondent[2]
Good teeth (photo)	25% of profiles per respondent
Bad teeth (photo)	25% of profiles per respondent
Obese (photo)	12.5% of profiles per respondent
Darker skin (photo)	12.5% of profiles per respondent

Table 2: Within-respondent distribution of profile photo conditions (illustrative).

Caste signal (via name/surname)	Probability of appearance
High caste	33.3%
Low caste	33.3%
Neutral	33.3%

Table 3: Distribution of caste signals across name templates (design target).

Marital status	Probability of appearance
Married	33.3%
Unmarried	33.3%
Not mentioned	33.3%

Table 4: Design probabilities for marital-status margin.

Certification	Probability of appearance
Present	50%
Absent	50%

Table 5: Design probabilities for certification margin.

Degree type	Probability of appearance
BSc	33.3%
BCom	33.3%
BA	33.3%

Table 6: Design probabilities for degree-type margin.

4 Theoretical framework

This section sets out a simple framework that links observable appearance to employer screening decisions and clarifies the channels we identify empirically: (i) informational/statistical updating about productivity, (ii) taste-based exclusion, and (iii) concern about customer or coworker preferences.

Setup. Consider an employer who must decide whether to invite an applicant to interview. The employer's payoff from interviewing applicant a equals the expected surplus from hiring conditional on interview, net of screening costs. Let the applicant's true productivity be θ (unobserved at the screening stage) and let $s \in \mathcal{S}$ denote an observable signal (headshot features: dental appearance, gender cue, caste-salient surname, etc.). The employer forms a posterior belief $\mathbb{E}[\theta \mid s]$. Hiring is more likely when $\mathbb{E}[\theta \mid s]$ and/or other non-productivity concerns exceed thresholds.

A parsimonious linear decision rule for an employer who sorts on an index X is

$$\text{Invite}(s) = \mathbf{1}\{\delta \cdot \mathbb{E}[\theta \mid s] + \tau \cdot T(s) + \kappa \cdot C(s) \geq c\},$$

where

- $\mathbb{E}[\theta \mid s]$ is the employer's productivity expectation conditional on the signal;
- $T(s)$ captures taste-based disutility from interacting with applicants who exhibit attribute s (e.g., distaste for perceived poor hygiene);
- $C(s)$ is a customer/coworker externality channel (employer internalizes the effect on customers or coworkers' utility); and
- c is the screening threshold; δ, τ, κ are nonnegative weights.

Interpretation. Three empirically distinct channels follow naturally:

1. **Statistical discrimination:** If s predicts lower expected productivity, $\mathbb{E}[\theta \mid s]$ falls and the employer reduces hiring probability ($\delta > 0$). For oral health, this would require employers to believe that poor dental appearance correlates with lower skills or unreliability.
2. **Taste-based discrimination:** If employers intrinsically dislike s (e.g., find poor dental appearance aversive), $\tau < 0$ reduces invites irrespective of productivity signals.
3. **Third-party concerns:** If customers or coworkers react negatively, $\kappa \neq 0$ and employers anticipate revenue or workplace-friction losses (particularly in customer-facing roles), raising the effective screening threshold for candidates with s.

Mechanism identification in our design. The three experiments identify these channels as follows.

- The perception survey measures how s changes trait attributions and predicted earnings — a direct measure of $\mathbb{E}[\theta \mid s]$ and beliefs about third-party responses.
- The correspondence test reveals whether reduced beliefs or tastes translate into realized callback rates at the hiring gate (compares conditional invitation probabilities across randomized s).
- The IRR elicits professional evaluators' hire scores and forecasts (incentivized) about other actors; if HR professionals lower hire scores but forecast low customer sensitivity, this points to taste or internal-selection channels rather than customer-facing concerns.

Implications for interventions. The framework suggests distinct policy levers: if statistical discrimination dominates, information remedies (certifications, skill-verification, anonymized initial screening) may help; if taste or customer concerns dominate, interventions require norm change, regulation (photo-optional applications), or subsidized health interventions to remove the visible cue. The empirical sections below ask which channel(s) best explain the observed patterns in India.

5 Results

5.1 Perception survey

Tables 8–11 report respondent fixed-effects regressions (clustered standard errors) estimating how visual cues in the presented photograph affect a broad set of trait attributions and predicted log salary. The core treatments are (i) an image that signals poor oral health (`poororal`); (ii) a randomized short qualification prompt about the person shown (`qual`); and (iii) three secondary visual variants (darker skin tone, obesity, spectacles). All regressions control for the photo-gender indicator (`photomale`; images in positions 1, 3 and 5 are male) and absorb respondent fixed effects so identification comes from within-respondent comparisons across images.

Main effect of poor oral health. The poor-oral image substantially lowers perceived traits. For the primary employability-related items, the poor-oral treatment reduces trust by 0.357 (SE = 0.111, $p < 0.01$), confidence by 0.513 (SE = 0.118, $p < 0.01$), and diligence by 0.410 (SE = 0.113, $p < 0.01$). The composite employability index (average of standardized trust, confidence and diligence) declines by 0.201 (SE = 0.043, $p < 0.01$). These are substantively meaningful changes on the original item scales.

Qualification prompt. The randomized qualification prompt (`qual`) has small point estimates across outcomes and the `poororal×qual` interaction is imprecisely estimated. We therefore find no robust evidence that provision of the qualification statement meaningfully *closes* the trait gap induced by the poor-oral photograph.

Photo gender. Photo gender is an important and consistent predictor: images in male positions (1/3/5) receive significantly lower ratings on many traits. The photomale coefficients are large and negative for trust (0.371, SE = 0.059), confidence (0.417, SE = 0.056), diligence (0.310, SE = 0.054) and the employability index (0.171, SE = 0.021), all $p < 0.01$. These patterns suggest a second axis of systematic heterogeneity in attributions beyond the poor-oral manipulation.

Secondary visual features and gender heterogeneity. The secondary image features (darker, obese, glasses) affect different traits in different directions and sometimes interact with photo gender:

- *Darker skin tone.* The darker photographic variant is associated with higher ratings on several interpersonal traits. For example, trust increases by 0.367 (SE = 0.105, $p < 0.05$) and attractiveness by 0.462 (SE = 0.121, $p < 0.05$). These positive effects appear across both interpersonal and some capability outcomes in the reported models.
- *Obesity.* The obese variant shows mixed main effects but notable positive interactions with photo gender in some specifications: several obese×male coefficients are positive (and significant at conventional levels in a subset of traits), indicating that the obese variant sometimes elicits higher ratings when the photo is in a male position. This pattern highlights that obesity cues do not uniformly reduce perceived traits and that gender moderates responses.
- *Spectacles.* Images with spectacles tend to be judged more favorably on competence-related traits: for example, glasses is positively associated with wealth/confidence/intelligence in multiple models (coefficients often positive and some estimated precisely). In a number of models the glasses×male interaction is also positive.

Taken together the secondary-feature results show that different visual cues move different trait inferences, and that some cues exhibit significant gender heterogeneity. The direction and magnitude of these secondary effects are reported in the panels of Tables 8–11.

Predicted salary. Table 12 reports regressions of log(salary). The poor-oral photo is associated with a modest but statistically significant reduction in predicted log salary (coefficient −0.089, SE = 0.044, $p \approx 0.046$); this corresponds to an approximate $\exp(-0.089) - 1 \approx -8.5\%$ reduction in the predicted salary for the poor-oral image relative to other images. The qualification prompt has a small positive point estimate on log salary (0.029, SE = 0.018) but it is not statistically significant;

Table 7: Perception survey: Treatment effects on trait ratings by domain (respondent FE regressions)

Panel A: Interpersonal traits				
	Trust	Similarity	Attractiveness	
Poor oral (poororal)	-0.153	-0.335*	-0.613*	
	(0.122)	(0.129)	(0.138)	
Panel B: Competence traits				
	Confidence	Diligence	Intelligence	Education
Poor oral (poororal)	-0.289*	-0.293*	-0.348*	-0.510*
	(0.129)	(0.116)	(0.122)	(0.136)
Panel C: Physical/health traits				
	Healthy lifestyle	Perceived wealth		
Poor oral (poororal)	-0.595*	-0.421*		
	(0.140)	(0.141)		

Notes: Each column is a separate respondent fixed-effects regression. Standard errors in parentheses are clustered by respondent. $^*p < 0.05$. Covariates include: qualification blurb, darker skin tone, obese, glasses, photo male. Variables: poororal = visibly poor teeth; qual = qualification shown; photomale = male photo; darker = darker skin; obese = obese; glasses = spectacles.

again the interaction with `poororal` is near zero. Thus, while the poor-oral cue lowers predicted pay modestly, we do not find evidence that the qualification prompt meaningfully mitigates the salary penalty.

Robustness and caveats. All models absorb respondent fixed effects and report respondent-clustered standard errors, which controls for time-invariant respondent-level confounders and accounts for within-respondent correlation. The specification reported is the baseline: it is straightforward to augment these models with respondent covariates (age, education, income) or to report marginal means and predicted values for each treatment cell; such extensions are available on request. Finally, several secondary-feature $\times$ photo-gender interactions are significant and merit further theoretical interpretation; these heterogeneous patterns motivated the grouped presentation of trait tables below.

Table 8: Treatment effects on interpersonal traits (respondent FE; clustered SEs)

	Trust	Attractiveness	Similarity
Poor oral (poororal)	-0.153	-0.613*	-0.335*
	(0.122)	(0.138)	(0.129)
Qualification (qual)	0.056	-0.046	-0.007
	(0.058)	(0.066)	(0.068)
Poororal × qual	-0.035	0.075	0.119
	(0.158)	(0.191)	(0.181)
Darker (darker=1)	0.367*	0.462*	0.108
	(0.105)	(0.121)	(0.123)
Obese (obese=1)	0.099	-0.300*	-0.055
	(0.114)	(0.125)	(0.125)
Glasses (glasses=1)	0.251*	0.107	0.096
	(0.106)	(0.121)	(0.117)
Photo male (male=1)	-0.472*	-0.964*	-0.496*
	(0.083)	(0.099)	(0.099)
Constant	7.520*	7.862*	6.961*
	(0.069)	(0.079)	(0.072)
Observations	3,084	3,084	3,084
Respondent clusters	514	514	514

Notes: Each column is a separate respondent fixed-effects regression. Standard errors (in parentheses) are clustered by respondent id. $^*p < 0.05$.

Table 9: Treatment effects on capability traits (A) (respondent FE; clustered SEs)

	Wealth	Confidence	Education
Poor oral (poororal)	-0.421*	-0.289*	-0.510*
	(0.141)	(0.129)	(0.136)
Qualification (qual)	-0.022	0.022	0.007
	(0.063)	(0.059)	(0.063)
Poororal × qual	0.172	0.116	0.132
	(0.182)	(0.169)	(0.173)
Darker (darker=1)	0.111	0.380*	0.129
	(0.128)	(0.111)	(0.102)
Obese (obese=1)	0.135	0.059	-0.039
	(0.121)	(0.120)	(0.113)
Glasses (glasses=1)	0.324*	0.301*	0.363*
	(0.125)	(0.108)	(0.102)
Photo male (male=1)	-0.697*	-0.543*	-0.586*
	(0.087)	(0.094)	(0.091)
Constant	7.237*	7.868*	7.868*
	(0.072)	(0.067)	(0.066)
Observations	3,084	3,084	3,084
Respondent clusters	514	514	514

Notes: See Table 8 for notation. $^{*}p < 0.05$.

Table 10: Treatment effects on capability traits (B) (respondent FE; clustered SEs)

	Healthy	Diligence	Intelligence
Poor oral (poororal)	-0.595*	-0.293*	-0.348*
	(0.140)	(0.116)	(0.122)
Qualification (qual)	0.023	0.007	0.122*
	(0.060)	(0.058)	(0.057)
Poororal × qual	0.166	0.163	-0.003
	(0.179)	(0.163)	(0.166)
Darker (darker=1)	0.075	0.071	0.063
	(0.105)	(0.097)	(0.094)
Obese (obese=1)	-0.343*	-0.004	0.084
	(0.124)	(0.099)	(0.103)
Glasses (glasses=1)	0.097	0.144	0.307*
	(0.102)	(0.104)	(0.098)
Photo male (male=1)	-0.654*	-0.433*	-0.486*
	(0.089)	(0.080)	(0.085)
Constant	7.730*	7.834*	7.769*
	(0.067)	(0.058)	(0.059)
Observations	3,084	3,084	3,084
Respondent clusters	514	514	514

Notes: See Table 8 for notation. $^{*}p < 0.05$.

Table 11: Treatment effects on perceived role suitability (respondent FE; clustered SEs)

	Backend-facing	Customer-facing
Poor oral (poororal)	-0.262	-0.382*
	(0.138)	(0.148)
Qualification (qual)	-0.140*	-0.006
	(0.066)	(0.067)
Poororal × qual	0.088	0.204
	(0.184)	(0.199)
Darker (darker=1)	0.154	0.312*
	(0.126)	(0.123)
Obese (obese=1)	-0.044	-0.047
	(0.126)	(0.122)
Glasses (glasses=1)	0.162	0.226*
	(0.123)	(0.113)
Photo male (male=1)	-0.391*	-0.814*
	(0.097)	(0.104)
Constant	7.629*	7.815*
	(0.077)	(0.077)
Observations	3,084	3,084
Respondent clusters	514	514

Notes: See Table 8 for notation. $^{*}p < 0.05$.

Table 12: Log salary predicted by image treatment and qualification (respondent fixed effects; clustered SEs)

	ln(salary)
Poor oral (poororal)	−0.089*
	(0.044)
Qualification (qual)	0.029
	(0.018)
Poororal × qual	0.006
	(0.074)
Photo male (photomale)	−0.028
	(0.019)
Observations	3,078
Respondent clusters	513
R-squared	0.753

Notes: Dependent variable is ln(salary). Models include respondent fixed effects (absorbed). Standard errors clustered by respondent id (reported in parentheses). *$p < 0.05$.

5.2 Correspondence Study

The correspondence experiment provides direct evidence on whether the perceptual penalties documented above translate into actual exclusion at the gateway of the labor market. Table 13 condenses the main estimates.

Panel A focuses on neutral surnames (Kumar/Kumari) where resumes differ only in dental appearance. The estimated callback difference between applicants with visibly poor teeth and those with healthy teeth is close to zero and statistically indistinguishable from zero. The point estimates imply that poor oral appearance lowers callback rates by less than one percentage point, relative to a baseline of roughly 17 percent. This hints that employers notice poor teeth, but suggests that on average this cue may not independently trigger rejection at the resume stage.

Panel B shows that gender generates a much more pronounced callback gradient. Restricting to vacancies that did not explicitly specify a male or female preference, female applicants are 5.3 percentage points more likely to receive a callback than observationally identical male applicants. This effect is highly significant and large in magnitude: women's resumes are about one-third more likely to progress past the initial screen. Importantly, the oral-health penalty is concentrated within this advantaged group: among female applicants, callbacks fall by about 1.25 percentage

points when the photograph shows poor oral health, while among men callbacks are equally low irrespective of teeth condition.

Panel C documents that caste signals remain the dominant axis of exclusion. Applicants with high-caste surnames enjoy callback rates about five percentage points higher than otherwise identical resumes carrying Dalit-sounding surnames, corresponding to a 40–45 percent advantage. Neutral surnames such as Kumar/Kumari also perform substantially better than Dalit surnames, while "initial-only" names (e.g. Ravi K.) are treated nearly as unfavorably as Dalit names. These results underscore that in the Indian fresher labor market, social identity continues to operate as a powerful filter long before any interview stage.

Table 13 also shows that oral-health penalties are heterogeneous. The interaction estimates confirm that average callback penalties are small, but suggest important variation: the oral-health penalty is larger in customer-facing roles and is borne disproportionately by women. In backend or remote roles, visible oral deficits are essentially ignored, while in customer-facing positions they can lower female callbacks by one to two percentage points—nontrivial relative to the baseline. These patterns dovetail with the IRR evidence, where HR professionals explicitly penalized poor oral appearance in contexts where customer or coworker reactions were salient.

Taken together, the correspondence results show that visible oral health is not a universal bar to entry, but rather a contingent penalty. On average it pales next to caste discrimination, which remains large and robust, and it interacts sharply with gender: employers appear to "look harder" at female applicants and discount those whose photographs reveal dental deficits. These findings also clarify the mechanisms hinted at in the perception and IRR studies. First impressions of poor teeth translate into real penalties only when decision-makers have scope to weigh appearance heavily—most notably in customer-facing roles and when hiring women, who otherwise enjoy an advantage. Thus the broader lesson is that treatable health signals matter, but they matter most in contexts where social hierarchies leave room for appearance-based differentiation.

5.3 Incentivized resume rating (IRR)

The IRR complements the perception and correspondence designs by placing trained recruiters in a controlled evaluation task that looks very much like screening. Each respondent sees a set of short *resume snapshots* that we render as *profile cards*. The variable `card` indexes these templates and their display order. Cards 1–6 are photo resumes and cards 7–12 are text-only resumes. Within a respondent, cards differ by randomized visual treatment and small resume fields, but the template content that could mechanically drive ratings is soaked up by `card` fixed effects. Our preferred estimates therefore come from within-respondent comparisons that absorb respondent and card fixed effects and cluster standard errors by respondent. Each respondent rated six photo cards (cards 1–6)

Table 13: Correspondence Study: Callback Effects by Oral Health, Gender, and Caste

	LPM Coef.	Logit AME	Baseline Callback
Panel A. Neutral-surname (Kumar/Kumari)			
Poor oral health (vs good)	−0.004	−0.008	0.170
	(0.014)	(0.014)	
Panel B. Gender (no preference postings)			
Female (vs male)	0.053***	0.053***	0.160
	(0.010)	(0.010)	
Panel C. Caste (baseline = High)			
Low caste surname	−0.051**	−0.052**	0.170
	(0.016)	(0.016)	
Neutral surname (Kumar/Kumari)	−0.041***	−0.041***	0.170
	(0.011)	(0.011)	
Initial-only surname (e.g. K.)	−0.031*	−0.031*	0.170
	(0.016)	(0.016)	

Notes: Each row reports the estimated effect on the probability of callback. Columns 1 and 2 report coefficients from linear probability models (LPM) and average marginal effects (AME) from logit regressions, respectively; standard errors in parentheses. Baseline callback is the mean callback rate in the relevant comparison group. Robust SEs clustered by employer. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

and two text-only cards randomly drawn from cards 7–12, with respondent and card fixed effects absorbing template content and display position.

Main effect. In the photo sample, the bad-teeth image lowers the 0–10 hiring score by about 0.176 points relative to good teeth (SE = 0.045, $p < 0.001$). Darker-skin and obese variants are negative but statistically indistinguishable from zero. Figure 3a shows the reghdfe coefficients ordered from lowest to highest, and Figure 3b shows the parallel specification with the full `pic` factor. Table **??** reports coefficients and clustered standard errors.

Heterogeneity. Interactions indicate that the bad-teeth penalty does not statistically differ by caste signal in this IRR sample (bad teeth × high caste +0.187, SE = 0.119, $p = 0.118$; bad teeth × low caste +0.024, SE = 0.106, $p = 0.821$; Table 15). The female differential is also not statistically different from zero (Table 16). For text-only profiles there is no detectable pattern by gender or caste (Table 17), which is consistent with the idea that visible oral health matters when a photo is present.

Channels and incentives. To diagnose why recruiters penalize poor teeth, we elicit five forecasts after each card and pay using a proper quadratic scoring rule with external benchmarks. The five channels are:

1. **Industry norms channel:** recruiter beliefs about how *other employers* would treat the same candidate in three role families: front-end customer-facing jobs, back-end non-customer-facing

jobs, and remote jobs.

2. **Customer acceptance channel:** beliefs about how *customers* would feel about interacting with the candidate in customer-facing roles.
3. **Coworker acceptance channel:** beliefs about how *coworkers* would feel about working with the candidate.

These forecasts are incentivized so that truthful reporting is a best response, and they map directly into the three conceptual mechanisms in this paper: expected industry practice, customer reactions, and coworker reactions.

Mechanism results. With only the bad-teeth indicator in the model, poor teeth significantly reduce expected outcomes for front-end employer demand (-0.083, $p = 0.047$), customer acceptance (-0.104, $p = 0.016$), and coworker acceptance (-0.099, $p = 0.015$); back-end and remote forecasts are negative but imprecise (Table 18). Adding the darker-skin and obese controls leaves the strongest and most precise penalties on customers and coworkers (Table 19). Cross-channel equality tests do not reject equality relative to the customer channel, which is consistent with a pattern in which customer and coworker concerns move together and front-end employer beliefs move in the same direction but with somewhat less precision.

Attenuation when adding channels. When we add a standardized index that averages the three employer-role forecasts, together with standardized customer and coworker forecasts, the bad-teeth coefficient on the hiring score shrinks from -0.157 (SE = 0.036) to -0.102 (SE = 0.030); see Table 20. This $\approx 35\%$ attenuation suggests that a meaningful share of the hiring penalty operates through recruiters' own beliefs about industry practice and through anticipated reactions by customers and coworkers. The residual direct effect may reflect residual taste or unmeasured informational updating.

Interpretation. The IRR shows that when recruiters engage with resumes in a structured setting, visible oral health carries a small but robust penalty on hiring scores, and that penalty is strongly aligned with forecasts about customer and coworker acceptance and, to a lesser degree, with beliefs about what other employers would do in customer-facing roles. This helps reconcile the perception results, where visual cues shift first impressions, with the correspondence results, where realized callback penalties for oral health are modest on average: visible oral health matters most when it is both seen and salient for a role in which third parties are likely to react.

Table 14: IRR main effects among photo profiles

	(1) FE OLS: i.pic	(2) HDFE: bad, darker, obese
bad_teeth	-0.176***	
	(0.04)	
darker_skin	-0.026	
	(0.06)	
obese	-0.049	
	(0.05)	
Profile card (1-12)=2	0.232***	
	(0.07)	
Profile card (1-12)=3	0.001	
	(0.07)	
Profile card (1-12)=4	0.166***	
	(0.06)	
Profile card (1-12)=5	0.237***	
	(0.06)	
Profile card (1-12)=6	0.000	
	(0.06)	
Bad teeth (vs good)=1		-0.176***
		(0.04)
Darker skin (vs good)=1		-0.026
		(0.06)
Obese (vs good)=1		-0.049
		(0.05)
Constant	8.125***	8.231***
	(0.05)	(0.03)
Observations	2425	2422

Standard errors in parentheses

Respondent and card fixed effects. SE clustered by respondent.

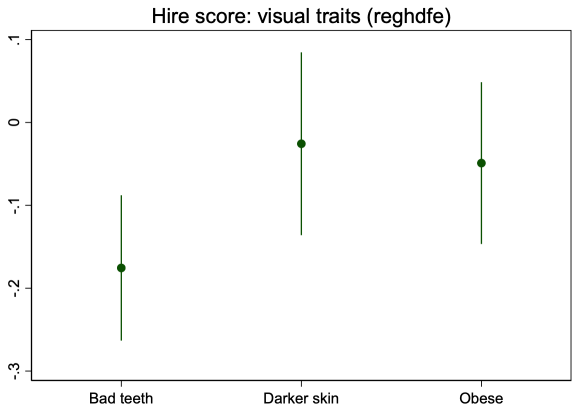
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Heterogeneity by caste: interactions with photo treatments

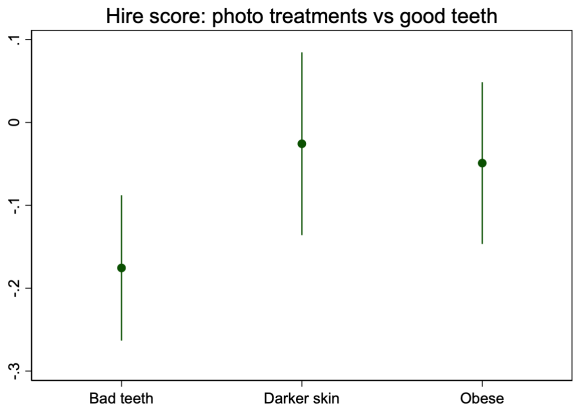
	(1) hire
Bad teeth (vs good)=1	-0.242***
	(0.08)
Bad teeth (vs good)=1 × High caste=1	0.187
	(0.12)
Bad teeth (vs good)=1 × Low caste=1	0.024
	(0.11)
Darker skin (vs good)=1	-0.046
	(0.10)
Darker skin (vs good)=1 × High caste=1	0.198
	(0.14)
Darker skin (vs good)=1 × Low caste=1	-0.128
	(0.16)
Obese (vs good)=1	-0.048
	(0.08)
Obese (vs good)=1 × High caste=1	0.151
	(0.13)
Obese (vs good)=1 × Low caste=1	-0.160
	(0.13)
Observations	2422

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$



(a) Hire score: visual traits (reghdfe)



(b) Hire score: photo treatments vs. good teeth

Figure 3: IRR coefficient plots (ordered).

Table 16: Heterogeneity by gender: interactions with photo treatments

	(1) hire
Bad teeth (vs good)=1	-0.115*
	(0.07)
Bad teeth (vs good)=1 × Female candidate=1	-0.125
	(0.09)
Darker skin (vs good)=1	0.036
	(0.09)
Darker skin (vs good)=1 × Female candidate=1	-0.117
	(0.12)
Obese (vs good)=1	0.026
	(0.07)
Obese (vs good)=1 × Female candidate=1	-0.152
	(0.10)
Observations	2422

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Non-photo resumes: gender and caste patterns (respondent FE)

	(1) hire
Female candidate=1	0.019
	(0.14)
High caste=1	0.034
	(0.19)
Low caste=1	-0.056
	(0.18)
Observations	802

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Mechanisms: effect of bad teeth on five forecasts (photo sample)

	(1) Emp front	(2) Emp back	(3) Emp remote	(4) Customers	(5) Coworkers
Bad teeth (vs good)=1	-0.083**	-0.019	-0.047	-0.104**	-0.099**
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Observations	2404	2396	2393	2402	2411

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 19: Mechanisms: effects of visual treatments on five forecasts (photo sample)

	(1) Emp front	(2) Emp back	(3) Emp remote	(4) Customers	(5) Coworkers
Bad teeth (vs good)=1	-0.092*	-0.015	-0.042	-0.133**	-0.148***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Darker skin (vs good)=1	-0.033	-0.034	-0.003	-0.053	-0.089
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Obese (vs good)=1	-0.006	0.046	0.023	-0.063	-0.106*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
Observations	2404	2396	2393	2402	2411

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 20: Exploratory attenuation: adding forecast channels (photo sample)

	(1) hire	(2) hire
Bad teeth (vs good)=1	-0.157***	-0.102***
	(0.04)	(0.03)
Standardized values of (emp_index)		0.463***
		(0.06)
Standardized values of (customers)		0.201***
		(0.04)
Standardized values of (coworker)		0.572***
		(0.06)
Constant	8.213***	8.191***
	(0.01)	(0.01)
Observations	2422	2393

Standard errors in parentheses

Resp and card FE; SE clustered by respondent.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 21: Cross-channel tests for bad-teeth effect (photo sample; baseline = Customers)

Channel	Effect b (SE)	p-value vs. customers
Customers (baseline)	−0.104 (0.043)	–
Front-end employees	−0.077 (0.041)	0.537
Back-end employees	−0.028 (0.047)	0.150
Remote employees	−0.038 (0.041)	0.206
Coworkers	−0.099 (0.040)	0.901

Notes: Entries in the first column report estimated effects of *Bad teeth* by channel from reghdfe models with respondent and card FE; the third column reports tests of equality relative to the customers channel (interaction = 0). Standard errors clustered by respondent.

6 Empirical Test of Screening Model

We next examine which screening mechanisms from Section 4 align with our experimental results. In the model, an employer invites a candidate if

$$\delta \cdot \mathbb{E}[\theta|s] + \tau \cdot T(s) + \kappa \cdot C(s) \geq c,$$

where δ captures statistical discrimination (expected productivity), τ reflects taste-based prejudice, and κ represents anticipated customer or coworker reactions.

Statistical discrimination (δ). If employers perceive poor oral health as a productivity signal, they should downgrade candidates' expected ability. Consistent with this, the perception survey shows that respondents forecast $\approx$ 8–9% lower salaries for candidates with poor teeth and rate them lower on confidence and diligence. However, the correspondence study finds no significant overall callback penalty, except among women, suggesting δ is modest in real employer behavior.

Taste-based discrimination (τ). A pure taste channel implies intrinsic dislike of the attribute. In the incentivized resume-rating experiment, recruiters penalized poor teeth by ≈ 1.8 points on a 0–10 scale, even when qualifications were held constant. This provides clear evidence of $\tau < 0$ operating as a taste-based penalty.

Customer/coworker concerns (κ). Employers may anticipate external reactions to visible oral health. In the IRR, recruiters expected negative customer and coworker responses to poor teeth, especially in customer-facing jobs. Controlling for these forecasts reduces the hiring penalty by $\approx$ 35%, indicating that much of the effect operates through κ. The correspondence study similarly shows that penalties for poor teeth are concentrated in customer-facing roles and among women, consistent with employers screening based on anticipated third-party preferences.

Summary. Together, the evidence suggests that taste-based penalties and anticipated customer/coworker concerns are the dominant channels. Statistical discrimination alone cannot explain the observed patterns, as perceptions of lower productivity do not translate into large callback gaps on average.

7 Policy relevance and discussion

Because oral health is both visible and remediable, the evidence in this paper carries direct implications for labor-market and health policy. The three experiments together highlight that (i) dental appearance shapes first impressions and professional evaluations, (ii) recruiters explicitly penalize poor oral health through taste-based and customer-facing channels, and (iii) these penalties are often masked in the field by stronger social hierarchies, especially caste. Policy should therefore target both the direct appearance-based mechanisms and the broader structures that condition their impact.

Photo-optional applications and blind screening. Appearance-based penalties emerge even when qualifications are held constant. Requiring or encouraging photo-blind screening at early stages could reduce the influence of non-productive visual cues. This is a low-cost procedural change that prevents bias from shaping callbacks while preserving later-stage evaluation of verified skills.

Targeted dental-care subsidies integrated with placement. Dental appearance is remediable. Our perception and IRR results show that better oral health improves employability ratings and forecasts of customer acceptance. Bundling subsidized dental care with job-placement programs for low-income youth, particularly women entering customer facing roles—could yield both health and labor-market returns. Cost-effectiveness depends on persistence of treatment effects, an important direction for future work.

Reinforcing verifiable skill signals. When employers use visual appearance as a noisy productivity proxy, certifications or skill badges can substitute as stronger, verifiable signals. Our IRR results suggest recruiters respond positively to credential-like information, implying that interventions that pair skills verification with health remediation could be especially effective.

Normative and regulatory interventions. The IRR evidence shows that a significant share of the penalty comes from anticipated customer and coworker reactions. Supply-side dental programs alone cannot resolve this. Policies need to combine anti-discrimination enforcement with demand-side interventions—such as stigma-reduction campaigns—that directly shift norms about appearance.

Equity across identity groups. Because caste-based exclusion remains large and robust, oral-health interventions alone will not equalize access. Policies that remove or mask caste identifiers in applications, or that actively target caste discrimination through affirmative hiring programs, are essential complements.

Operational lessons for practitioners. NGOs and placement agencies can act immediately: (i) prioritize dental outreach where clients apply for customer-facing roles, (ii) counsel applicants on photo use and quality when optional, and (iii) combine short-run cosmetic fixes (photography coaching, low-cost dental care) with longer-run clinical interventions.

8 Conclusion

This paper studies a neglected but policy-relevant margin of labor-market screening: visible oral health. Using three complementary experiments, we show that (i) poor dental appearance substantially lowers trust, confidence, and expected earnings in perception surveys; (ii) HR professionals impose economically meaningful hire-score penalties (about 1.8 points on a 0–10 scale) and forecast customer and coworker rejection of poor-oral candidates, with these channels accounting for one-third of the total effect; and (iii) in a large correspondence study, caste and gender hierarchies dominate callbacks, though oral-health penalties reemerge in customer-facing roles and among women.

Together these results highlight an important general point: treatable health deficits impede market access not only through productivity but through appearance-based channels, and their translation into realized exclusion depends on the strength of co-occurring identity signals and the hiring context. Effective policy requires a portfolio of tools—from photo-blind screening to subsidized dental programs to norm-shifting interventions—that address both the appearance-based mechanism and the entrenched hierarchies that condition it.

Future work should quantify the cost-effectiveness of specific interventions, such as bundled dental care and placement services, and assess longer-run labor-market impacts of remediating visible health cues. The experimental platforms developed here—AI-based perception tasks, large-scale correspondence scraping, and incentivized HR elicitation—provide a scalable roadmap for such policy-oriented evaluations.

Appendix: Quadratic Scoring Rule Implementation for Belief Elicitation

For each forecast question, respondents provide their prediction p_{ij} on a 0 to 10 scale for profile i and forecast item j. The true benchmark values t_j for each forecast item are obtained from externally collected customer and employer evaluations. To incentivize forecast accuracy, we implement a quadratic scoring rule following the standard form used in belief elicitation experiments (e.g., Schlag et al. 2015; DellaVigna et al. 2022; Enke and Zimmermann 2019).

For each forecast item, the respondent's bonus is calculated as:

$$S(p_{ij}, t_j) = \max\left\{0, B - k \cdot (p_{ij} - t_j)^2\right\}$$

where $B = 50$ INR is the maximum bonus for each forecast, and $k = \frac{B}{(10)^2} = 0.5$ ensures that a respondent earns the full bonus if their prediction is exactly correct, and zero if their prediction is maximally incorrect (i.e., 10 units away from the true value).

Since each HR respondent evaluates 8 profiles, we calculate their forecast bonus for each forecast dimension by first averaging their score across the 8 profiles:

$$\overline{S_j} = \frac{1}{8} \sum_{i=1}^{8} S(p_{ij}, t_j)$$

The respondent's total payment for the belief elicitation task is then:

$$\text{Total Bonus} = \sum_{j=1}^{4} \overline{S_j}$$

yielding a maximum possible bonus of INR 200 across the four forecast dimensions.

This scoring rule satisfies incentive compatibility under standard assumptions of expected utility maximization (see proper scoring rule literature: Brier 1950; Gneiting and Raftery 2007; Schlag et al. 2015).

Appendix: Survey Instruments

A. Perception Survey (Online, Centiment)

Consent Form

Study Number: IRB23-0921

Researchers: University of Chicago, Harris School of Public Policy

Description: You will be shown a few pictures and asked to provide your honest opinions on them. The survey will take approximately 8–10 minutes. No personal or sensitive information will be collected. Hosted by Centiment. Participation is voluntary. Responses will be confidential.

Risks and Benefits: No risks beyond daily life. No direct benefits, but contributes to understanding perceptions of faces.

Confidentiality: Responses are anonymized. De-identified data may be used in future research.

Contacts: research.study.query@gmail.com; SBS-IRB (773) 702-2915; sbs-irb@uchicago.edu.

Consent: By clicking "I agree," you confirm that you are at least 18 years old and agree to participate.

Survey Items

Q1.1 What are the first three things that come to your mind when you see this picture? (Open text)

Q1.2 On a scale of 0–10, please rate this person on:

Trustworthiness; Attractiveness; Wealth; Confidence; Education; Healthy lifestyle; Diligence; Intelligence; Similarity to self;

Suitability for backend role; (0 = Not at all, 10 = Very)

Suitability for customer-facing role. (0 = Not at all, 10 = Very)

Q1.3 Monthly salary: "This is [Name], 21 years old, working in [role]. What do you think is their monthly salary in INR?" (Numeric entry)

— These three questions repeat for 6 images (Picture 1 through Picture 6). —

Demographics

Gender (Male; Female; Other; Prefer not to say).

Marital status (Married; Widow/Widower; Divorced; Separated; Engaged; Single–looking; Single–not looking; Other).

Age (numeric).

Occupation (Student; Homemaker; Unemployed–looking; Unemployed–not looking; Govt.; Private; Business-owner; Freelancer; Small firm; Teacher/Professor; NGO; Professional; Retired; Other).

Household income (Less than INR10k; INR10–25k; INR25–50k; INR50k–1L; INR1–2L; More than INR2L).

Education (No schooling; Primary; Middle; Secondary; Higher Secondary; Diploma/ITI; Bachelor's; Master's; PhD).

State/UT of residence (full list provided).

B. HR Resume-Rating Survey (In-person, SME Sample)

Enumerator Recruitment Script

Hello, my name is [Enumerator Name], and I am working with Model Resource and the University of Chicago on a study about how hiring professionals evaluate entry-level candidates. The survey takes 15–20 minutes. You may skip any question. Compensation up to INR 300 based on forecasting accuracy. Participation is voluntary.

Consent Form

Study Number: IRB23-0921

Researcher: Rubina Hundal, University of Chicago.

Purpose: To understand how HR professionals at SMEs evaluate entry-level candidates.

Tasks: Answer firm background; evaluate candidate profiles; complete demographics.

Duration: 15–20 minutes.

Compensation: INR 100 base + up to INR 200 forecasting bonus (via GPay/UPI).

Confidentiality: Responses de-identified; identifiers deleted after payment.

Consent: By stating "I agree," you confirm you are 18+, understand the study, and agree to participate.

Survey Items

Firm and Role Background

Q1. Firm name.

Q2. Number of employees (1–10; . . . ; ¿1,000).

Q3. HR role involvement (resume review; shortlisting; interviewing; final decisions; other).

Q4. Are you directly involved in hiring entry-level employees? (Yes/No; if No $\rightarrow$ exit).

Q5. Types of roles usually hired (Customer-facing; Backend; Remote; Other).

Q6. Describe your role in hiring. (Open text).

Q7. Average time spent per resume (¡1 min; 1–3; 3–5; ¿5).

Candidate Evaluations (8 profiles)

Each respondent saw 6 photo resumes (good teeth / bad teeth / obese / darker skin) and 2 text-only. Profiles randomized across gender, caste, marital status, certification, degree. For each candidate:

– Hire-likelihood (0–10).

– Coworker comfort (0–10).

– Incentivized forecasts: employer likelihood (customer-facing; backend; remote), customer preference, coworker acceptance. Each 0–10. Forecast accuracy paid up to INR 50 each.

– Salary prediction (numeric INR).

– Trait ratings (0–10): Trustworthiness, Attractiveness, Wealth, Confidence, Education, Healthy lifestyle, Diligence, Intelligence, Similarity to self.

Compensation Module

Q140. Based on answers, you will earn INR 100 + INR [bonus].

Q141. Please enter your GPay/UPI ID.

Hiring Factors Module

Q130. Rank hiring factors (skills, qualifications, experience, communication, reliability, appearance, etc.).

Q131. Agreement scale (1=Strongly Disagree to 4=Strongly Agree): – Would pay higher wage for better fit. – Personal cleanliness is very important. – Keeping customers happy is main goal in customer-facing hires. – Technical skills are main goal in backend/remote hires. – Habits like smoking/breaks hurt performance. – Appearance affects hiring decisions. – Usually fill positions through referrals. – Providing more information would change ratings. – Believe other employers share same preferences.

Q132. What additional information would help you make better hiring decisions? (Open text).

Demographics

Q133. Name.

Q134. Age (numeric).

Q135. Gender (Male; Female; Other; Prefer not to say).

Q136. Caste (SC; ST; OBC; General; Other; Prefer not to say).

Q137. Years in HR (¡1; 1–3; 4–6; 7–10; ¿10).

Q138. Education (No schooling; Primary; Middle; Secondary; Higher secondary; Diploma/ITI; Bachelor's; Master's; PhD).

Q139. Marital status (Married; Widow/Widower; Divorced; Separated; Engaged; Single–looking; Single–not looking; Other).

Q142. Monthly income (¡10k; 10–25k; 25–50k; 50k–1L; 1–2L; ¿2L).

Closing

Q143. Phone number (optional).

Q144. Email address (optional).

Thank you for participating.

References

Antonakis, J. and Dalgas, O. (2009). Predicting elections: Child's play! *Science*, 323(5918):1183.

Arrow, K. J. (1973). The theory of discrimination. In Ashenfelter, O. C. and Rees, A., editors, *Discrimination in Labor Markets*, pages 3–33. Princeton University Press, Princeton, NJ.

Banerjee, A. V., Bertrand, M., Datta, S., and Mullainathan, S. (2009). Labor market discrimination in delhi: Evidence from a field experiment. *Journal of Comparative Economics*, 37(1):14–27.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289–300.

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013.

Bohren, J. A., Imas, A., and Hull, P. (2022). How discrimination accumulates across hiring stages. Technical report, Working paper. Working paper. Published versions / URLs (if any) can be substituted here; please replace with official citation/DOI when available.

Gallego, F. A., Larroulet Philippi, C., and Repetto, A. (2024). What's behind her smile? health, looks, and self-esteem. *American Economic Journal: Applied Economics*, 16(2):359–388.

Glied, S. L. and Neidell, M. (2010). The economic value of teeth. *Journal of Human Resources*, 45(2):468–496.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223–255.

Kessler, J. B., Low, C., and Sullivan, C. D. (2019). Incentivized resume rating: Eliciting employer preferences without deception. *American Economic Review*, 109(11):3713–3744.

McLaughlin, J. S. and Neumark, D. (2025). Is there intersectional labor market discrimination? Technical Report 33388, National Bureau of Economic Research.

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

Olivola, C. Y., Funk, F., and Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, 18(10):566–570.

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

Rooth, D.-O. (2009). Obesity, attractiveness, and differential treatment in hiring: A field experiment. *Journal of Human Resources*, 44(3):710–735.

Siermińska, E. and Singhal, K. (2023). Does it pay to be beautiful? *IZA World of Labor*, (161):161.

Singhal, S., Mamdani, M., Mitchell, A., Tenenbaum, H., Lebovic, G., and Quiňonez, C. (2016). Dental treatment and employment outcomes among social assistance recipients in ontario, canada. *Health Policy*, 120(10):1202–1208. Epub 2016 Sep 4.

Todorov, A. and Oosterhof, N. N. (2011). Modeling social perception of faces. *IEEE Signal Processing Magazine*, 28(2):117–122.

Willis, J. and Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7):592–598.