Inaccurate Statistical Discrimination: An Identification Problem

Aislinn Bohren, Kareem Haggag, Alex Imas & Devin G. Pope

Existence of Discrimination

Systemic discrimination has been well documented.

- Racial, ethnic, gender, sexual orientation.
- Documented in a variety of important markets: housing, credit, labor, online consumer markets.
- See Bertrand Duflo (2016) and Fang Moro (2011) for overview.

Literature defines two potential sources of discrimination

- **Taste- or preference-based discrimination:** due to animus against members of a group (Becker 1957).
- **Statistical discrimination:** uncertainty about productivity
  - Believe groups avg productivity is lower (Phelps 1972; Arrow 1973).
  - Believe group has higher variance (Aigner & Cain 1977).
- **Distinguishing** between these two types of discrimination is important for policy and welfare.
  - Methods: outcome-based tests, varying information set.
  - **Key identifying assumption:** beliefs are accurate.

Inaccurate Beliefs

- Evidence from psychology and economics that people have systematic biases in beliefs.
  - Inaccurate stereotypes based on representativeness heuristic (Bordalo et al. (2016)).
  - Enormous literature in psychology discussing how stereotypes can be inaccurate (see Fiske (2018) for review).
- Inaccurate beliefs may arise from a lack of information.
  - Inexperience (e.g. aliens arrive from Mars).
  - Accurate priors about population averages, but unsure how selection impacts population being evaluated (e.g. hiring & promotion criteria).
  - Knowledge of other peoples beliefs (e.g. interpreting a recommendation letter).

Why do Inaccurate Beliefs Matter?

- Evaluators use beliefs to learn about trait of candidate (e.g. productivity).
  - Belief about population distribution of trait.
  - Belief about how statistics (GPA, test scores) convey information about candidates trait.
  - Belief about prior evaluators beliefs (e.g. how to interpret reference).
- Inaccurate beliefs lead to inaccurate inference about the candidate.
  - May persist even in the face of a lot of information about candidate (learning literature on model misspecification).
  - May cause inefficient exit from market.
  - Bohren Imas Rosenberg (2019): dynamics of discrimination with inaccurate beliefs.

Important to distinguish between discrimination due to accurate vs. inaccurate beliefs

Methodological

- Relaxing assumption of accurate beliefs generates an identification problem.

Policy

- Effective interventions when beliefs are inaccurate very different than those for accurate statistical or taste-based sources.
- Welfare implications for inaccurate statistical discrimination differ from accurate statistical and taste-based discrimination.

Most papers on discrimination ignore the possibility of inaccurate beliefs

	# Papers	Percent
Total Papers	105	100.0%

Most papers on discrimination ignore the possibility of inaccurate beliefs

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%

Most papers on discrimination ignore the possibility of inaccurate beliefs

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%
Discuss taste-based versus statistical source	65	61.9%

Most papers on discrimination ignore the possibility of inaccurate beliefs

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%
Discuss taste-based versus statistical source	65	61.9%
Test for taste-based versus statistical source	49	46.7%

Most papers on discrimination ignore the possibility of inaccurate beliefs

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%
Discuss taste-based versus statistical source	65	61.9%
Test for taste-based versus statistical source	49	46.7%
Discuss accurate versus inaccurate beliefs	11	10.5%

Most papers on discrimination ignore the possibility of inaccurate beliefs

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%
Discuss taste-based versus statistical source	65	61.9%
Test for taste-based versus statistical source	49	46.7%
Discuss accurate versus inaccurate beliefs	11	10.5%
Test for inaccurate beliefs	7	6.7%
Measure beliefs	7	6.7%

Scope: papers published between 1990-2018 in AEJ: Applied, AEJ: Policy, AER, EMA, JEEA, JLE, JPE, ReStud, ReStat, QJE.

An Identification Problem

Worker:

- Group identity $g \in \{M, F\}$.
- Productivity $a \sim N(\mu_g, 1/\tau_g)$.
- Signal $s = a + \epsilon$, $\epsilon \sim N(0, 1/\eta_g)$.
- Focus on discrimination against group F.

Evaluator:

- Decides whether to hire worker, $v \in \{0, 1\}$.
- Holds subjective beliefs $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$.
  - Misspecified model when subjective parameters $\neq$ true parameters.
- Observes g and s; uses Bayes rule to updates belief about a.
- Hires worker if subjective posterior $\hat{E}_\theta[a|s, g]$ above group-specific hiring threshold $u_g \in \mathbb{R}$.
- $\theta = \{u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g\}_{g \in \{M,F\}}$ denotes evaluator's type.

Partiality

Categorize different forms of **preferences** and **beliefs.**

- Use partiality to refer to properties of model primitives.
- To distinguish from discrimination – property of behavior and consequence of primitives.

Definition (Preference Partiality)

An evaluator has *preference partiality* if $u_F \neq u_M$.

Definition (Belief Partiality)

An evaluator has *belief partiality* if $(\hat{\mu}_F, \hat{\tau}_F, \hat{\eta}_F) \neq (\hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M)$. *Accurate* if $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g) = (\mu_g, \tau_g, \eta_g)$ for $g \in \{M, F\}$ and otherwise *inaccurate*.

Discrimination

Let $v(s, g, \theta) \equiv \mathbb{1}\{\hat{E}_\theta[a|s, g] \geq u_g\}$ denote optimal hiring decision.

- Difference between hiring decisions

$$D(s, \theta) \equiv v(s, M, \theta) - v(s, F, \theta).$$

- **Discrimination** occurs against group F if $\exists s$ s.t. $D(s, \theta) > 0$.
- Interested in when different sets of beliefs and preferences give rise to same discriminatory behavior.

Definition (Equivalent Discrimination)

Two evaluators of types θ and θ' exhibit equivalent discrimination if $D(s, \theta) = D(s, \theta')$ for all $s \in \mathbb{R}$.

Equivalent Discrimination

Posterior mean ability monotonic wrt s

- Represent optimal hiring decision as cut-off rule wrt signal: hire if $s \geq \bar{s}(\theta, g)$.
- Two types have same signal cut-offs $\Rightarrow$ equivalent discrimination.

Proposition

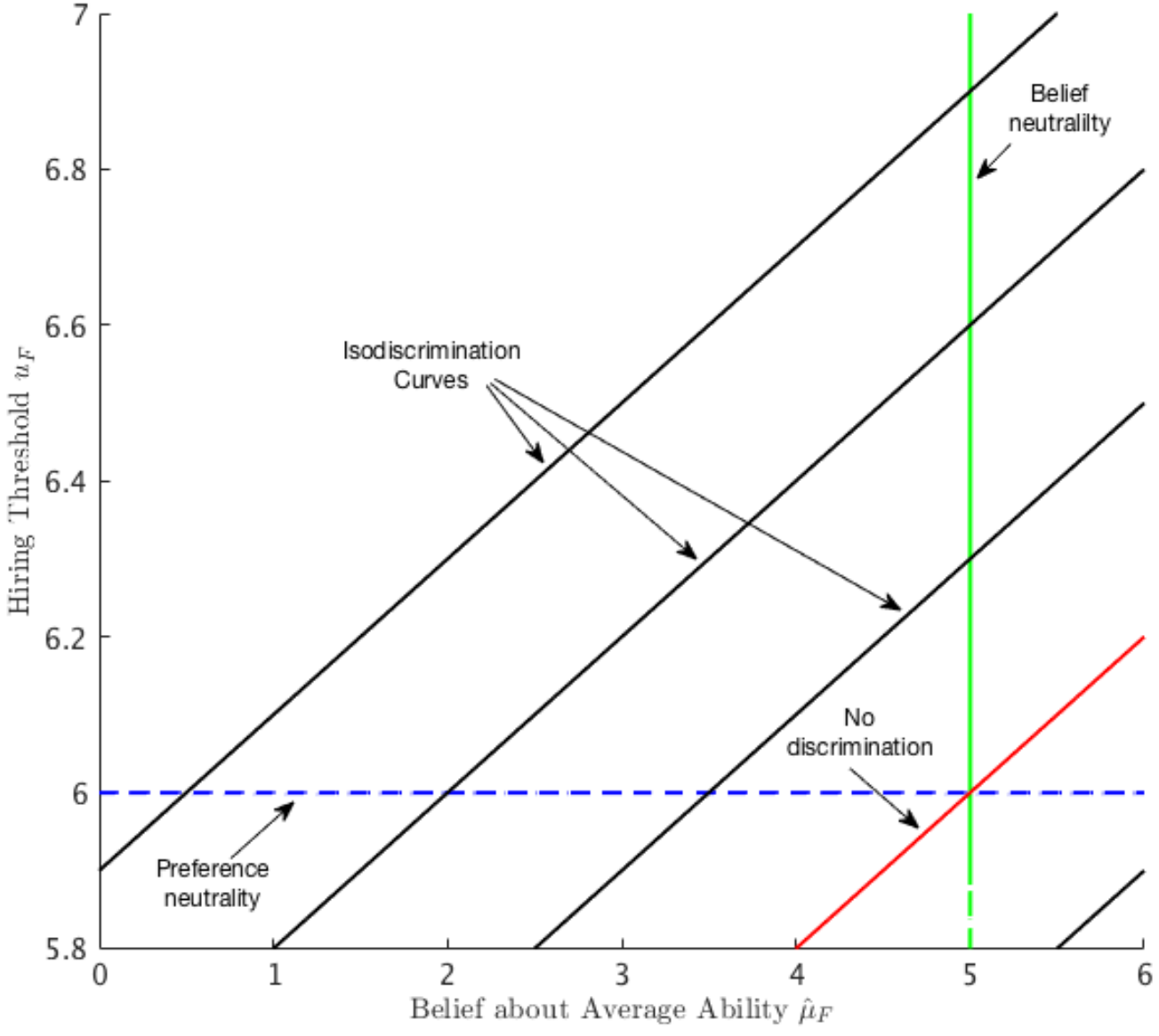
For any constants $(s_M, s_F) \in \mathbb{R}^2$ *with* $s_M \neq s_F$, *equivalent discrimination occurs for the set of types* $\{(u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)\}_{g \in \{M,F\}}$ *s.t.*

$$\frac{\hat{\tau}_M + \hat{\eta}_M}{\hat{\eta}_M} u_M - \frac{\hat{\tau}_M}{\hat{\eta}_M} \hat{\mu}_M = s_M \tag{1}$$

$$\frac{\hat{\tau}_F + \hat{\eta}_F}{\hat{\eta}_F} u_F - \frac{\hat{\tau}_F}{\hat{\eta}_F} \hat{\mu}_F = s_F \tag{2}$$

No discrimination occurs for set of types that satisfy (1) *and* (2) *for all* $s_M = s_F$.

Isodiscrimination Curves



Identifying Discrimination

Can property of interest can be backed out from available data?

- Assume, at the minimum, researcher observes group g and hiring decision v for each worker.

Existence of Discrimination

Show there exists an s such that $D(s, \theta) > 0$.

- If researcher also observes s, this is straightforward.
- Otherwise, can use correspondence or audit study: create fictitious workers with randomly assigned g and s.

Observations:

- Observing g, v and s for a set of workers from each group, with at least one signal $s \in [\overline{s}(\theta, M), \overline{s}(\theta, F))$, identifies the existence of discrimination.
- Observing g, v and s for a continuum of group M and F workers with signals in a neighborhood of $\overline{s}(\theta, M)$ and $\overline{s}(\theta, F)$, respectively, identifies isodiscrimination curve.

Source of Discrimination

Identify form of partiality that generates observed discriminatory behavior.

- Correspondence studies cannot distinguish between preference partiality and accurate belief partiality.
- Same insight extends to inaccurate belief partiality.
- Moreover, not possible to even rule out any potential sources.
  - Construct evaluator types with single category of preference or belief partiality that exhibit equivalent discrimination.
  - Establishes each form of partiality in isolation can generate given pattern of discrimination.

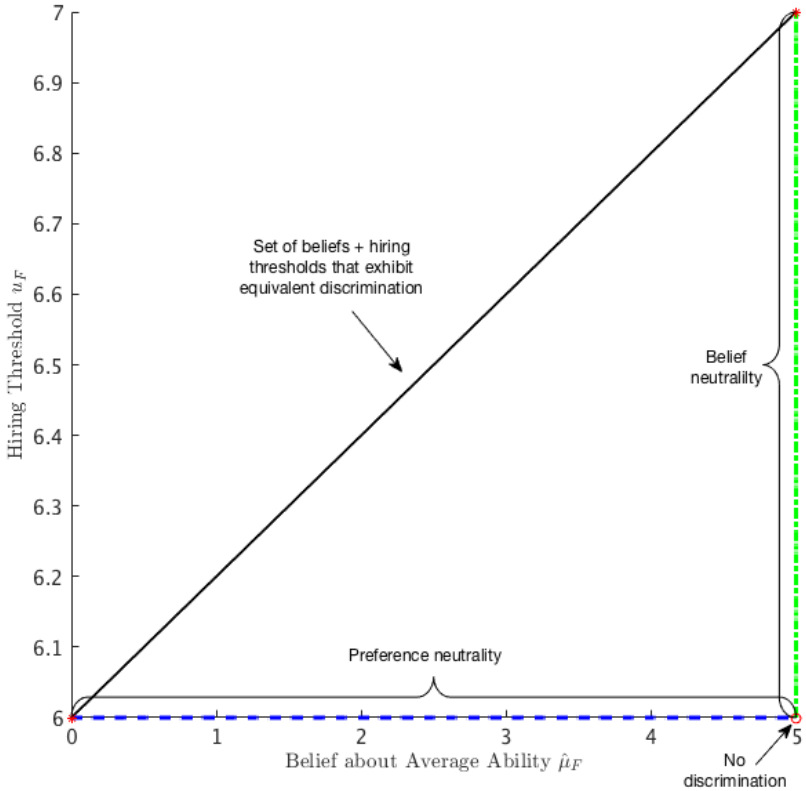
Source of Discrimination

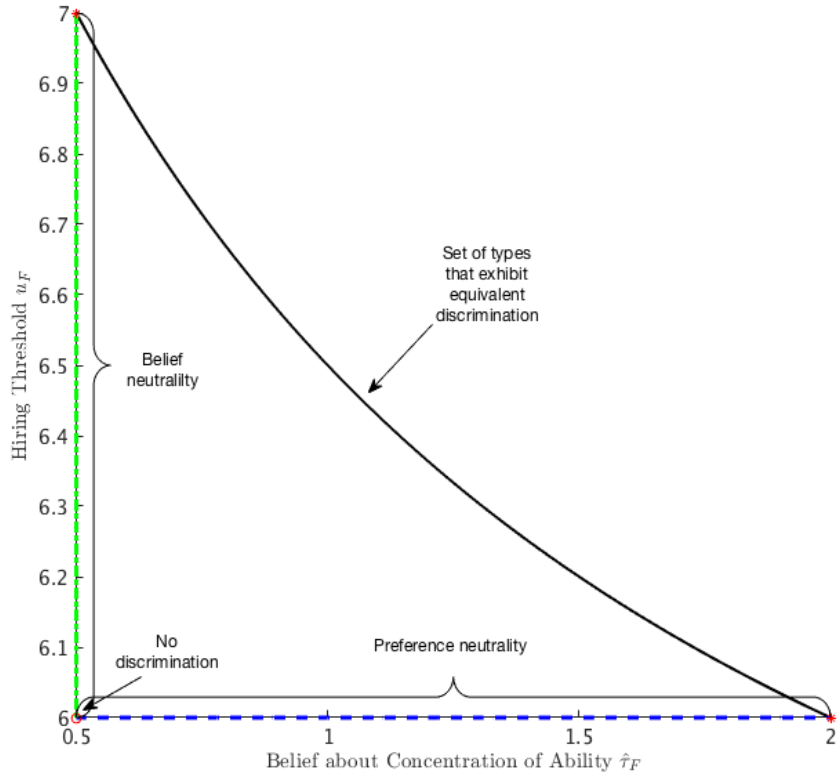
Proposition (Equivalent Sources)

For any $(s_M, s_F) \in \mathbb{R}^2$ with $s_F > s_M$, the corresponding isodiscrimination curve includes:

1. *A type with preference partiality and belief neutrality, $u_F > u_M$ and $(\hat{\mu}_F, \hat{\tau}_F, \hat{\eta}_F) = (\hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M)$.*
2. *A type with preference neutrality and belief partiality generated by lower expected productivity, $\hat{\mu}_F < \hat{\mu}_M$, and $(u_F, \hat{\tau}_F, \hat{\eta}_F) = (u_M, \hat{\tau}_M, \hat{\eta}_M)$.*
3. *A type with preference neutrality and belief partiality generated by higher (lower) concentration of productivity, $\hat{\tau}_F > \hat{\tau}_M$ ($\hat{\tau}_F < \hat{\tau}_M$), and $(u_F, \hat{\mu}_F, \hat{\eta}_F) = (u_M, \hat{\mu}_M, \hat{\eta}_M)$ for $\hat{\mu}_g < u_g$ ($\hat{\mu}_g > u_g$).*
4. *A type with preference neutrality and belief partiality generated by lower (higher) signal precision, $\hat{\eta}_F < \hat{\eta}_M$ ($\hat{\eta}_F > \hat{\eta}_M$), and $(u_F, \hat{\mu}_F, \hat{\tau}_F) = (u_M, \hat{\mu}_M, \hat{\tau}_M)$ for $\hat{\mu}_g < u_g$ ($\hat{\mu}_g > u_g$).*

Equivalent Sources





Identifying Source

Corollary

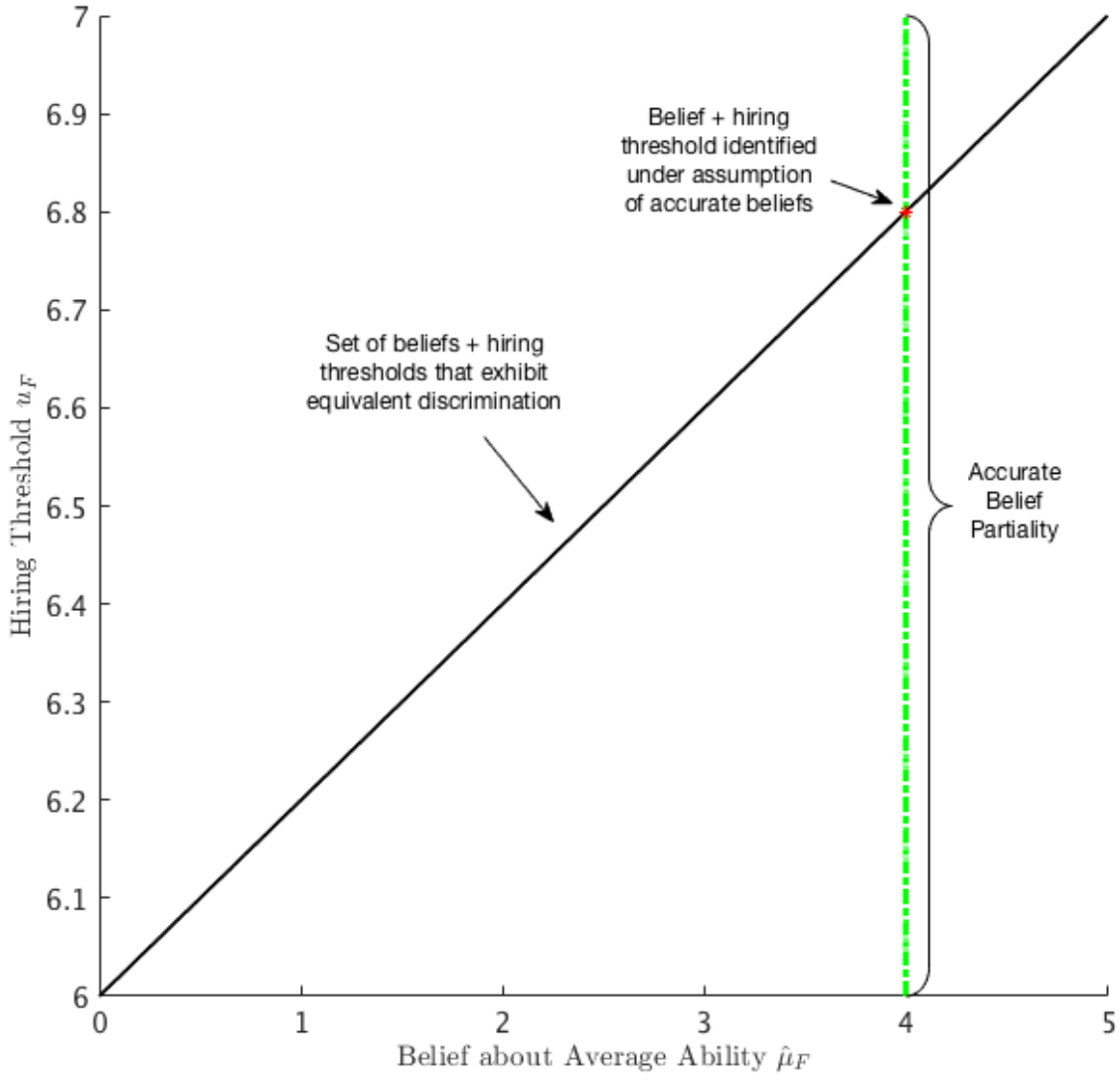
Identifying the isodiscrimination curve does not rule out any category of partiality.

Outcome-based Test.

Common method to identify source under assumption of accurate beliefs not valid when beliefs may be inaccurate.

- Suppose it is also possible to collect data on the productivity and signal distributions for each group.
- Under assumption of accurate beliefs, this identifies source.

Outcome-based Test: Accurate Beliefs



Outcome-based Test: Inaccurate Beliefs

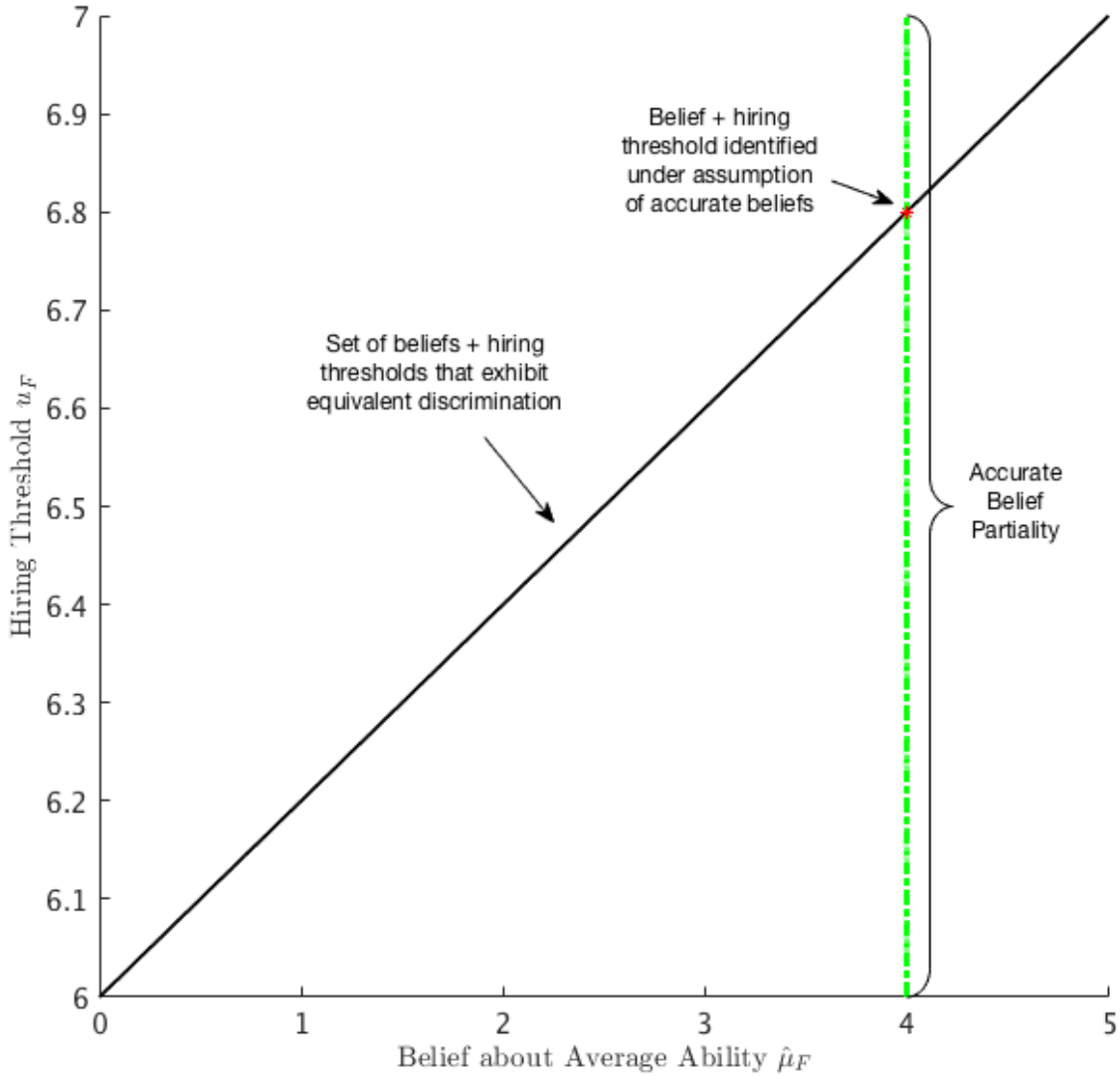
Identification crucially depends on accurate belief assumption.

- If true productivity and signal distributions identical, differential treatment may stem from preference partiality, inaccurate belief partiality, or combination of the two.
- If true productivity or signal distributions differ, differential treatment may stem from accurate beliefs (potentially coupled with preference partiality), preference partiality or inaccurate beliefs.

Observation (An Identification Failure)

Suppose a researcher observes the true productivity and signal distributions for each group and can identify the isodiscrimination curve. This does not identify the evaluator's type.

Outcome-based Test: Inaccurate Beliefs



Rejecting Accurate Statistical Discrimination

Sometimes possible to rule out accurate belief partiality as sole source of discrimination.

- See if given pattern of discrimination consistent with accurate belief partiality + preference neutrality.
- Of particular interest, since ruling out establishes discrimination stems from animus towards group or inaccurate beliefs about them.

Observation (Rejecting Accurate Statistical Discrimination)

Suppose a researcher observes the true productivity and signal distributions for each group and can identify the isodiscrimination curve with thresholds (s_M, s_F)*. If*

$$\frac{\tau_M \mu_M + \eta_M s_M}{\tau_M + \eta_M} \neq \frac{\tau_F \mu_F + \eta_F s_F}{\tau_F + \eta_F}, \tag{3}$$

the evaluator is not an accurate statistical discriminator.

Eliciting Beliefs.

If possible to collect data on evaluator's subjective beliefs, then can identify source.

Observation (Identifying Preferences from Beliefs)

Suppose a researcher can identify the isodiscrimination curve. Observing $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$ *identifies* u_g*, and therefore, the evaluator's type.*

Does not identify whether beliefs are accurate – to do so, must also observe true distributions.

Experiment

Illustrative example

- Demonstrate importance of inaccurate beliefs in how we study discrimination.

Experiment

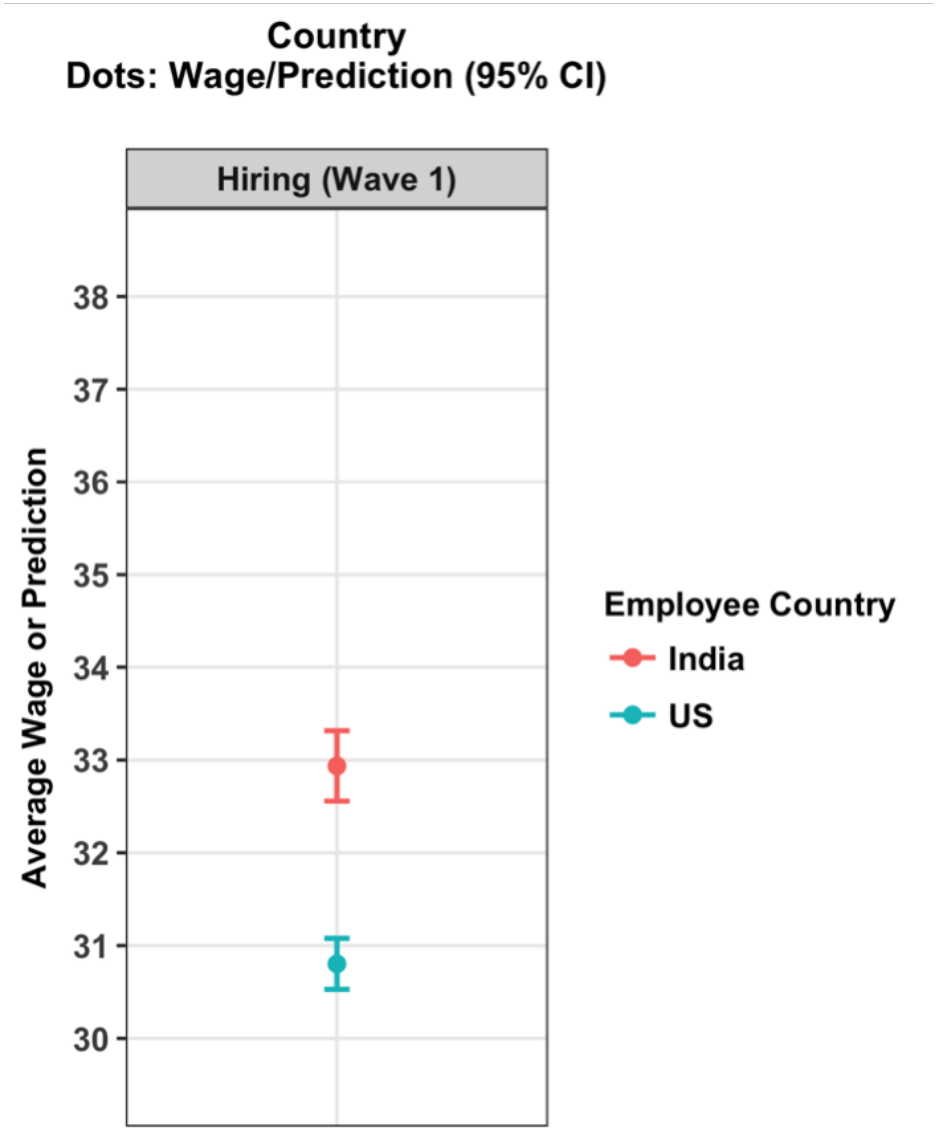
- **Part 1:** 600 Mturkers (400 from US; 200 from India) recruited to be "employees."
  - Completed 50 question math test.
  - Answered 8 questions to develop personal profile.
- **Part 2:** 600 different Mturkers (400 from US; 200 from India) recruited to be "employers."
  - Observed profiles of 20 different employees.
  - Decided maximum wage (wtp) to hire each (0-50 cents).
- **Payment**
  - Drawn random number x between 0 and 50 for each potential employee; if wtp $\geq x$, employer hires employee and pays wage x.
  - Employers earn 1 cent for each question that hired employee answered correctly.
  - Employers told avg # correct answers was 37.
  - Comprehension questions to check understanding.

Step 1: check for differential treatment

Three sets of group identities in experiment:

1. **Indian vs. American**
2. Male vs. Female
3. Old vs. Young

Indians employees receive higher wages than Americans



Summary

Part 1: Document Differential Treatment

- Indians favored relative to Americans.

Given the conventional pattern for studying discrimination, how would we test for statistical vs. taste-based discrimination?

Summary

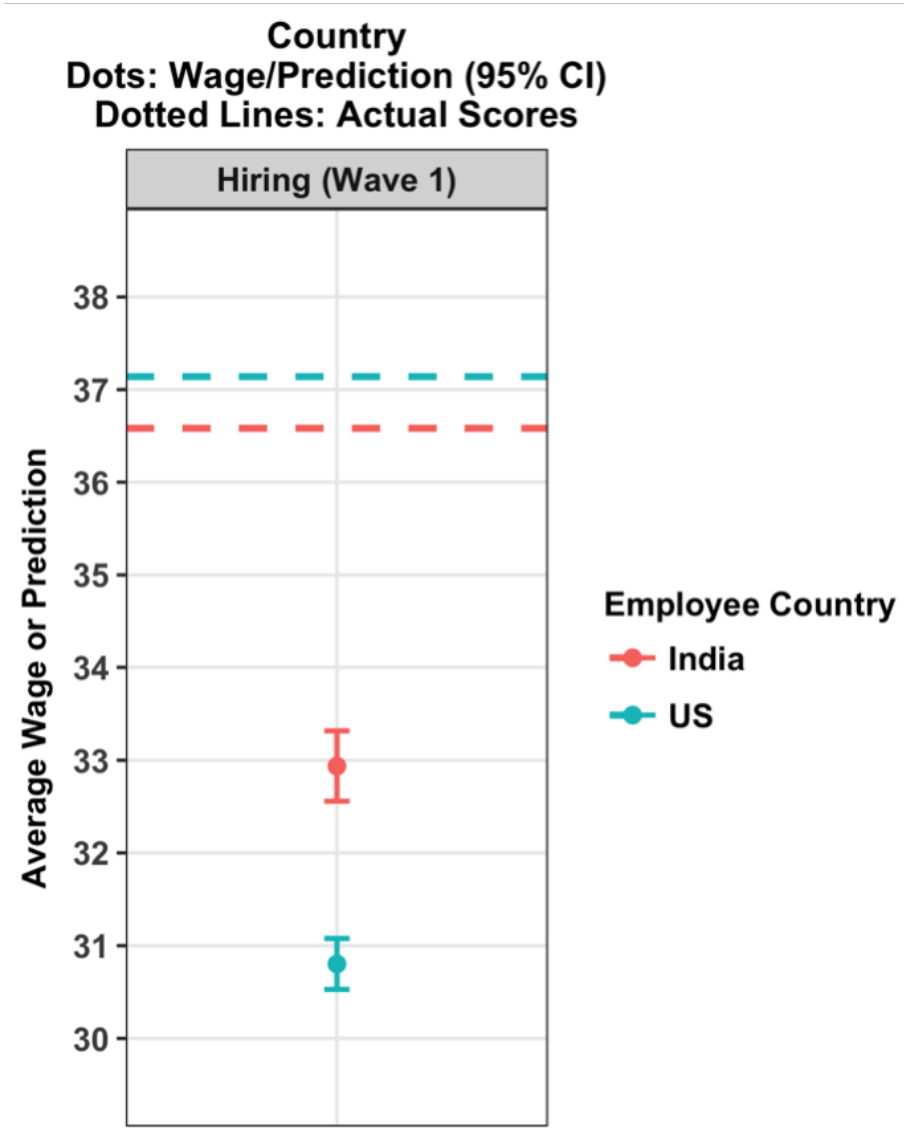
Part 1: Document Differential Treatment

- Indians favored relative to Americans.

Given the conventional pattern for studying discrimination, how would we test for statistical vs. taste-based discrimination?

- Compare to true performance distributions.

Taste-based discrimination against Americans.



Summary

Part 1: Document Differential Treatment

- Indians favored relative to Americans.

Part 2: Compare to True Distributions (e.g. standard technique)

- Taste-based discrimination against Americans.

But what if beliefs are wrong?

Part 3: Elicit Average Beliefs

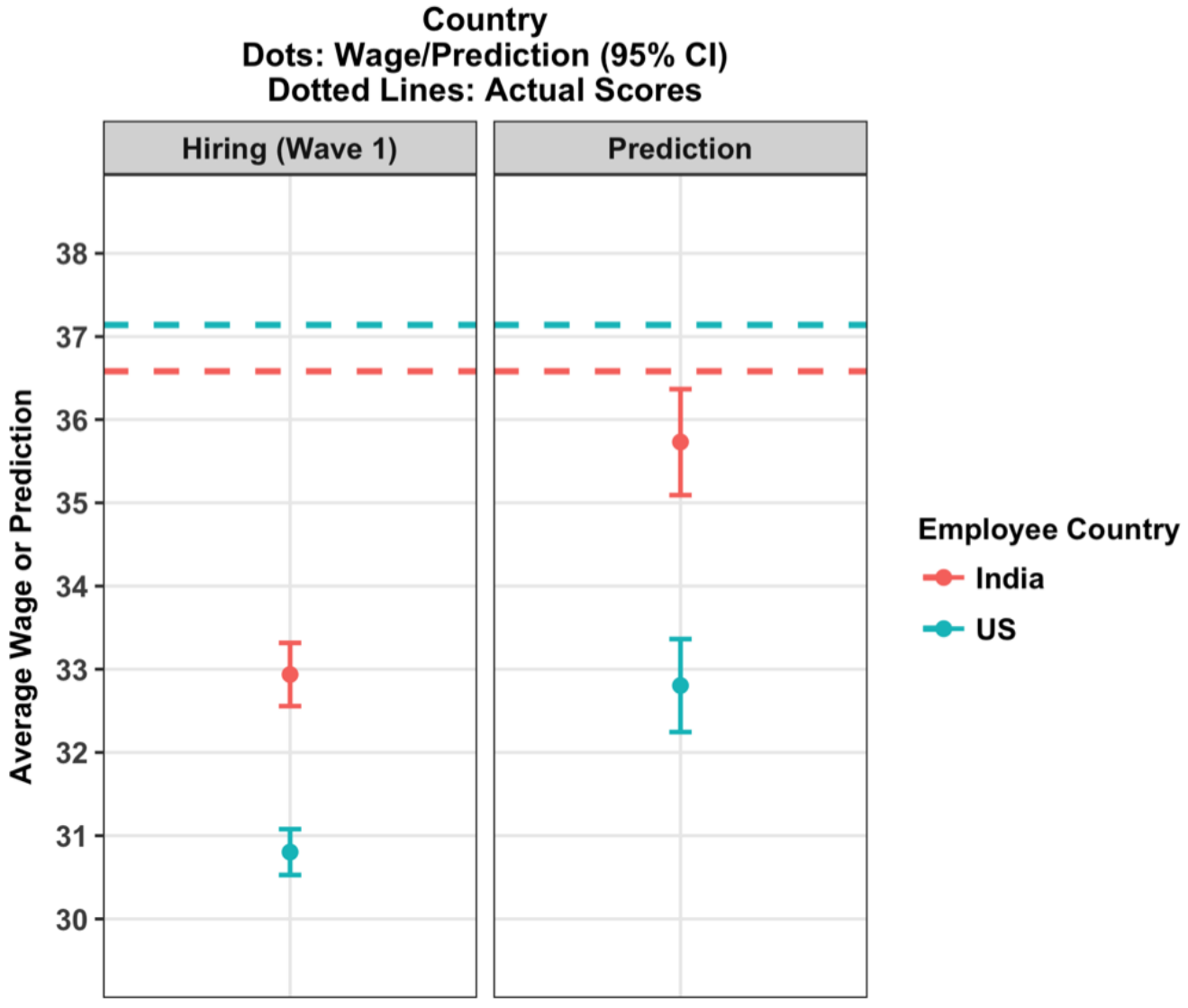
Employers asked following question:

"On average, how many questions do you think g answered correctly?"

where $g \in \{Americans, Indians\}$.

- Provided incentive compatible bonuses for performance on questions.
- Randomized between small and large incentives – no differences between two groups.

Inaccurate statistical discrimination against Americans.



Summary

Part 1: Document Differential Treatment

- Indians favored relative to Americans.

Part 2: Compare to True Distributions (e.g. standard technique)

- Taste-based discrimination against Americans.

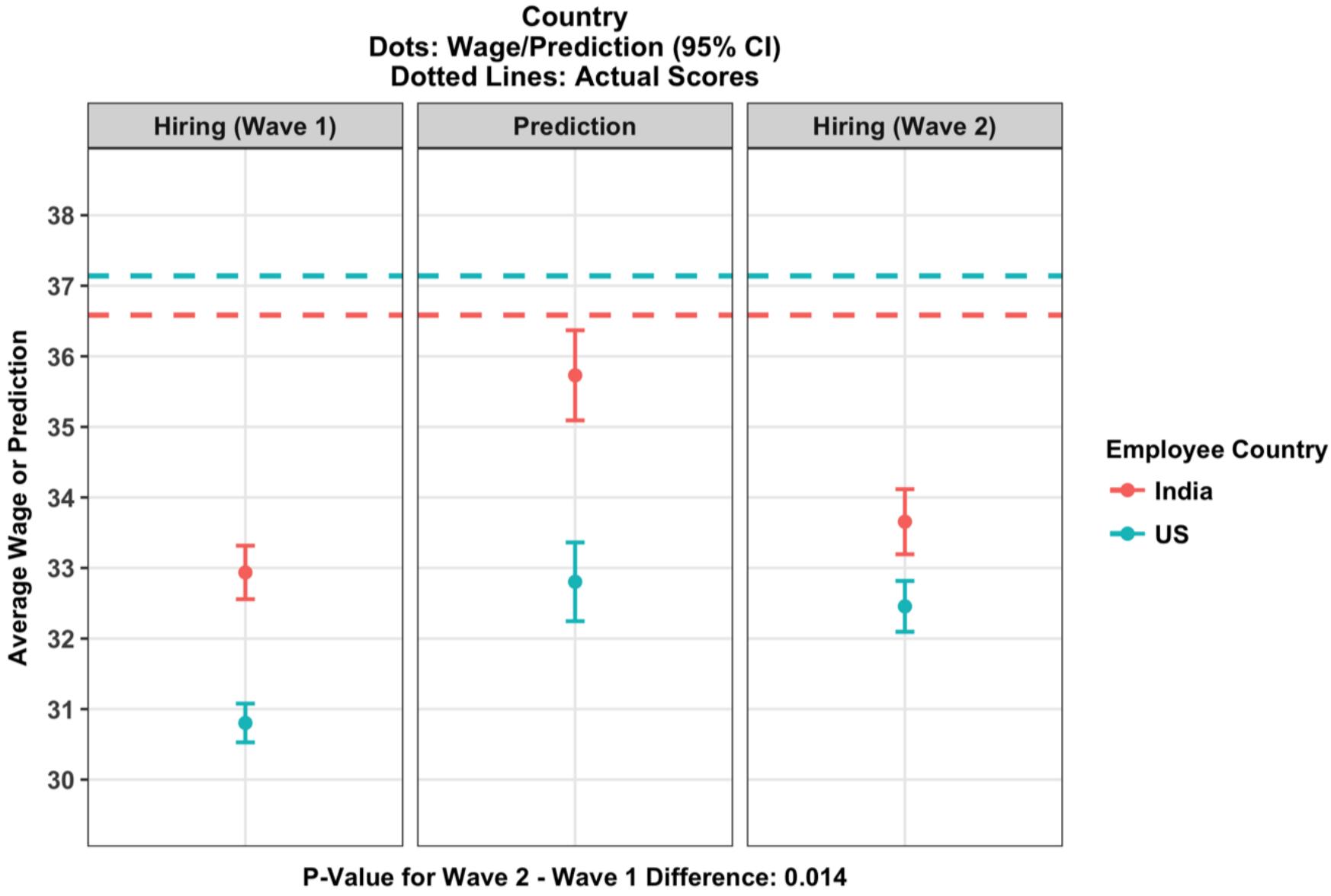
Part 3: Compare True and Perceived Distributions

- Inaccurate statistical discrimination against Americans.

Intervention: is it possible to correct beliefs?

- Employers were told *"Here are the correct answers for the belief questions that you answered."*
  - Americans answered 37.14 questions correctly on average.
  - Indians answered 36.58 questions correctly on average.
- Asked to make wage offers based on 10 additional employee profiles (same hiring/payment scheme).

Either taste-based or still some inaccurate statistical discrimination against Americans.



Conclusions

Inaccurate beliefs may be prevalent in many domains currently being studied by discrimination researchers.

- Most tests to separate statistical and taste-based sources rely on strong assumption: people hold accurate beliefs.
- Incorrectly imposing this assumption can lead to mistaken attribution of source, and therefore, misguided policy/welfare analysis.
- Relaxing this assumption generates an identification problem.
- Illustrative example using results from an online experiment demonstrates the importance of inaccurate beliefs in how we study discrimination.

Conclusions: Two Key Forms of Inaccurate Beliefs

Inaccurate priors / beliefs about population.

- Persistent discrimination due to selection effects / exit
- May not "wash out" in the long-run.
- Discrimination reversals can be sign of selection (Bohren Imas Rosenberg (2019)).

Inaccurate beliefs about signals.

- Persistent discrimination due to incorrect updating.
- Could be about exogenous signals (e.g. how predictive is a test) or endogenous signals (e.g. interpreting reference letter).