

Overconfidence & Prejudice

The Review of Economic Studies (2026)

Paul Heidhues¹, Botond Köszegi¹, Philipp Strack²

Department of Economics, Universität Bonn¹

Yale University, Department of Economics²

Renato Pirisi

October 1st, 2026

Discrimination in standard economics

- Many economic decisions involve judging other people: whom to hire, whom to promote, or whom to trust.
- In standard microeconomics, we explain and model these situations through **utility maximization**, given preferences and beliefs.
- **Becker (1957): a taste for discrimination.** An employer who dislikes employing members of a particular group may be willing to sacrifice profits to avoid hiring them.
- This approach takes prejudice as given. Two further questions are: where do prejudiced views of other groups come from? and, most importantly, **can they be entirely reconciled with tastes?**

Biased beliefs, overconfidence, and prejudice

- Research in behavioral economics documents **systematic biases in beliefs and judgments on others** (Bordalo et al., 2016; Heidhues et al., 2018).
- One example is **overconfidence**: people may overestimate their own ability or performance (Moore and Healy, 2008) relative to others'.
- This creates a tension when outcomes fall short of expectations: *"If I am so capable, why am I not more successful?"* How we answer may shape our views of discrimination and other people.
- **This paper** studies an agent who maintains an inflated belief about his own ability but otherwise learns from evidence using Bayes' rule, asking
 - Can overconfidence generate favoritism toward one's own group and prejudice against people that are different?
 - When does better information reduce these biases, and when it does not?
 - What are the effects of two groups being on opposite sides of social issues (competing)?

Setup: individuals, groups, and discrimination

- Society consists of I individuals in G disjoint groups, subject to K types of **discrimination** (i.e. economic or social forces that affect groups). Individual j has fixed ability $a_j \in \mathbb{R}$ and belongs to group g_j .
- Agent i repeatedly observes each individual's **recognition** q_j and signals η_k about discrimination. Both the true and subjective models satisfy:

$$q_j = a_j + \sum_{k=1}^K \phi_{g_j k} \theta_k + \epsilon_j^q, \quad j = 1, \dots, I,$$
$$\eta_k = \theta_k + \epsilon_k^\eta, \quad k = 1, \dots, K.$$

- $\theta_k \in \mathbb{R}$ is the fixed extent of type- k discrimination, while $\phi_{gk} \in \mathbb{R}$ is its fixed incidence on group g . Higher θ_k benefits groups with $\phi_{gk} > 0$ and harms those with $\phi_{gk} < 0$.
- Errors are mean-zero normal and i.i.d. over time. their variances are denoted by v_j^q and v_k^η .

Setup: individuals, groups and discrimination (cont.)

- Crucially, discrimination is **redistributive**: writing m_g for group g 's population share,

$$\sum_{g=1}^G m_g \phi_{gk} = 0 \quad \text{for every } k.$$

- The true abilities and levels of discrimination are $a = A$ and $\theta = \Theta$. Agent i correctly knows group memberships g_j and the incidences of discrimination ϕ_{gk} .
- He is uncertain about others' abilities a_{-i} , discrimination θ , and the covariance matrix Σ of the errors. His prior over (a_{-i}, θ) has full support on $\mathbb{R}^{I-1} \times \mathbb{R}^K$: he initially rules out no configuration of these parameters.

- **Overconfidence:** he is certain that his own ability exceeds its true value:

$$\mathbb{P}_i(a_i = \tilde{a}_i) = 1, \quad \tilde{a}_i > A_i.$$

- He updates his beliefs using **Bayes' rule**. We study their **long-run limit**, while his belief about his own ability remains fixed.
- Our analysis of long-run beliefs is appropriate for a person who has had sufficient scope to learn about society but has (like most individuals observed in the empirical work) remained overconfident so far.

Theorem 1: Long-run biases

Long-run biases

- Agent i 's beliefs concentrate (almost surely) on a single vector $(\tilde{\theta}^i, \tilde{a}^i)$: the posterior probability of any open neighborhood of this vector converges to one.
- His long-run bias about **type- k discrimination** is

$$\tilde{\theta}_k^i - \Theta_k = \frac{-\phi_{g_i k} v_k^\eta}{v_i^q + \sum_{k'} \phi_{g_i k'}^2 v_{k'}^\eta} (\tilde{a}_i - A_i). \quad (1)$$

- His long-run bias about **individual $j \neq i$'s ability** is

$$\tilde{a}_j^i - A_j = \frac{\sum_k \phi_{g_i k} \phi_{g_j k} v_k^\eta}{v_i^q + \sum_{k'} \phi_{g_i k'}^2 v_{k'}^\eta} (\tilde{a}_i - A_i). \quad (2)$$

Its sign depends on how similarly discrimination affects his group and j 's group.

Interpretation and discussion

- **Self-serving beliefs about discrimination.** The agent overestimates discrimination that harms him ($\phi_{g;k} < 0$) and underestimates discrimination that benefits him ($\phi_{g;k} > 0$). These beliefs explain why his recognition falls short of what he expects given his inflated view of his own ability.
- **Justifying group advantages.** Underestimating favorable treatment allows members of advantaged groups to attribute more of their success to ability, consistent with the idea of a “legitimizing myth.”
- **Disagreement across groups.** Members perceive more discrimination against their group than outsiders who are unaffected by, or benefit from, that discrimination. An agent is biased about a mechanism only if it affects him.
- **Beliefs about discrimination shape beliefs about ability.** For a given observed outcome, perceiving greater obstacles leads the agent to infer greater ability.

Proposition 1: In-group bias

Let A_g be group g 's average ability and $\tilde{a}_h^g$ its average assessment of other members of group h . Assume $v_i^g = v_g^g$ for all i in group g .

- **(I) In-group overestimation.** Each group overestimates its own ability, but correctly estimates average ability in the population:

$$\tilde{a}_g^g > A_g, \quad \sum_h m_h \tilde{a}_h^g = \sum_h m_h A_h.$$

- **(II) Absolute in-group bias.** If all groups have the same average ability, each group considers its own members better than average:

$$\tilde{a}_g^g > \sum_h m_h \tilde{a}_h^g.$$

- **(III) Relative in-group bias.** On average, people rate fellow group members more highly than outsiders:

$$\sum_{g,h} m_g m_h (\tilde{a}_g^g - \tilde{a}_h^g) > 0.$$

Understanding in-group bias

- **Overconfidence spills over to fellow group members.** They face the same discrimination as the agent. By overstating their obstacles and understating their advantages, he attributes too much of their recognition to ability.
- **Errors cancel across groups.** Because discrimination is redistributive, these mistaken beliefs do not distort his estimate of average ability in the population:

$$\sum_h m_h (\tilde{a}_h^g - A_h) = 0.$$

Overestimating his own group therefore implies underestimating other groups on average.

- **Equal group abilities: absolute in-group bias.** Each group sees itself as above the common ability level and outsiders as below it, on average.
- **More generally: relative in-group bias.** Averaging across the population, people rate fellow group members more highly than outsiders, even when true group abilities differ.

The effects of competition

How does the emergence of **opposing interests** affect a group's beliefs about other groups?

Suppose groups g and g' are not initially affected by the same types of discrimination:

$$\phi_{gk}\phi_{g'k} = 0 \quad \text{for all } k.$$

A new type of discrimination then places them on opposite sides:

$$m_g\phi_{g,K+1} + m_{g'}\phi_{g',K+1} = 0, \quad \phi_{g,K+1} \neq 0.$$

Proposition 2

The new type of discrimination:

- ❶ **Competition effect:** Lowers group g 's assessment of group g' .
- ❷ **Excuse effect:** Raises group g 's assessment of itself.
- ❸ **Bias substitution:** Raises group g 's average assessment of groups other than g and g' .

Competition and the excuse effect

- **Competition worsens views of the rival group.** Members of group g overestimate discrimination favoring g' or underestimate discrimination against it. They therefore attribute less of g' 's recognition to ability.
- **Connection to group conflict theory.** People hold more negative views of groups they perceive as competitors, for example over jobs or social transfers (Stephan et al. (1999)). Experiments find that inducing perceived competition can worsen attitudes even toward an imaginary immigrant group.
- **A new excuse improves views of one's own group.** The new type of discrimination provides another explanation for the agent's low recognition. Since fellow group members face the same obstacles, he also attributes more of their recognition to ability.

Bias substitution

- **A new explanation weakens existing biases.** By attributing part of his low recognition to the new type of discrimination, the agent relies less on biased beliefs about other types of discrimination.
- **Views of other groups improve on average.** While his assessment of the rival group g' falls, he attributes more of the remaining groups' recognition to their abilities.
- **Historical illustration.** Fouka et al. (2022) document that Black migration to northern U.S. cities reduced previously substantial stereotyping of Irish and Italian immigrants.

Application: Political Econ

Bias substitution provides one rationale for a common political tactic: to focus attention on an outside competitor and unify a heterogeneous population, thus reducing negative views their constituent groups hold of one another.

Can more information reduce prejudice?

- **Correct specification versus overconfidence.** With enough information, a correctly specified agent learns the truth. An overconfident agent can instead become certain of biased beliefs.
- **More observations of the same signals.** Since our results concern long-run beliefs, simply observing more realizations does not eliminate the limiting biases.
- **More precise information about others' recognition.** Reducing v_j^q for $j \neq i$ does not change agent i 's long-run biases: this variance does not appear in the expression for his assessment of j 's ability.

Why does this information fail to help?

Knowing more about other groups' successes does not resolve the central tension driving prejudice: the gap between the agent's inflated self-assessment and his actual recognition.

The kind of information that actually matters

Suppose discrimination of type k affects agent i : $\phi_{g_i k} \neq 0$.

Proposition 3

An increase in precision $1/v_k^\eta$:

- ➊ **Direct effect:** Reduces his bias about discrimination of type k .
- ➋ **No-excuse effect:** Lowers his assessment of other members of his own group.
- ➌ **Bias substitution across discrimination types:** Increases his bias about every other type of discrimination that affects him.
- ➍ **Indirect benefit:** Raises his average assessment of other groups.
- ➎ **Bias substitution across groups:** Amplifies existing biases about groups unaffected by discrimination of type k ($\phi_{gk} = 0$).

Better information: benefits and drawbacks

- **One excuse becomes less plausible.** More precise information makes it harder to sustain a distorted belief about discrimination of type k . This also reduces the overestimation of fellow group members' abilities.
- **The agent seeks alternative explanations.** His inflated self-assessment remains unchanged. He therefore places more weight on other types of discrimination to explain his low recognition, increasing his biases along those dimensions.
- **Effects on other groups are mixed.** His average assessment of outsiders improves, but existing prejudices about groups unaffected by type- k discrimination become stronger.

Key implication

Correcting one mistaken belief can amplify other biases when the underlying overconfidence persists.

Contact and prejudice

- **The contact hypothesis.** Contact between groups can reduce prejudice. One informational channel is learning about the ability of an out-group member.
- **Potentially large spillovers.** Under correctly specified learning, information about one person may reveal little about a large, diverse group. Here, it can change beliefs about discrimination affecting the entire group.
- **How the mechanism works.** Suppose agent i learns individual j 's true ability, and j is affected only by discrimination of type k . Then:

$$q_j - A_j = \phi_{g,j,k} \theta_k + \epsilon_j^q.$$

Knowing A_j turns j 's recognition into an additional signal about θ_k .

- Thus, better information about discrimination of type k reduces agent i 's bias about j 's entire group.

Balanced information reduces biases

Proposition 4

A proportional increase in the precisions $1/v_k^\eta$ of signals about **all types of discrimination** reduces all nonzero biases about discrimination and other individuals' abilities.

- **Improving information across all dimensions.** Unlike better information about a single type of discrimination, a balanced improvement reduces biases without amplifying them elsewhere.



Distinguishing the direction of causality

Reverse-causality alternative: Stubborn false beliefs about other groups or discrimination generate overconfidence, rather than the other way around.

Both explanations link overconfidence to biased social beliefs, but their predictions differ:

- 1 **Responses to information.** Our model predicts that information about discrimination changes beliefs about groups. In the alternative, those beliefs are fixed or follow from stubborn beliefs about discrimination.
- 2 **Responses to competition.** Our model predicts that competition with a new group changes views of unrelated groups through bias substitution. The alternative does not predict this effect.
- 3 **Self-assessment versus in-group assessment.** Our model predicts greater overestimation of oneself than of fellow group members. In the alternative, the bias about oneself equals the bias about the average in-group member.

Prejudice without discrimination or group knowledge

Prejudiced beliefs can arise even when the agent considers **no systematic discrimination** and has **no pre-existing notion of groups**.

- **Recognition includes a common shock.** With $I \geq 3$, agent i repeatedly observes

$$q_j = a_j + \psi_j \epsilon_g + \epsilon_j,$$

where ϵ_g is a common shock, ϵ_j is individual noise, and ψ_j measures exposure to the common shock. Both shocks have mean zero.

- **Overconfidence remains the only misspecification.** The agent fixes his own ability at $\tilde{a}_i > A_i$ but learns about others' abilities using Bayes' rule. He knows v_g , but not exposures ψ_j or individual noise variances v_j .
- **Groups reflect shared exposure.** Normalize the true exposure of agent i to $\Psi_i > 0$. Individuals with $\Psi_j > 0$ are "in the same boat" as i ; those with $\Psi_j < 0$ form his out-group. He does not initially know who belongs to either group.

Prejudice without group knowledge

Proposition 6

The agent's beliefs converge to a single vector $(\tilde{a}^i, \tilde{\psi}^i)$. For each individual $j \neq i$, his long-run belief about ability is

$$\tilde{a}_j^i = A_j + \frac{\Psi_i \Psi_j v_g}{v_i^g + \Psi_i^2 v_g} (\tilde{a}_i - A_i),$$

and his belief about exposure to the common shock is $\tilde{\psi}_j^i = \kappa \Psi_j$, $\kappa > 1$

- **The agent correctly infers group membership.** He learns the sign of each Ψ_j : individuals with $\Psi_j > 0$ are in the same boat as him, while those with $\Psi_j < 0$ belong to the out-group.
- **But he exaggerates group differences.** Since $\kappa > 1$, he overestimates the importance of the common shock in determining recognition.
- **Implication for ability beliefs.** If $\tilde{a}_i > A_i$, he overestimates the abilities of individuals with $\Psi_j > 0$ and underestimates those with $\Psi_j < 0$.

Overconfidence through biased learning

Take our original model, but now the agent observes $s_i = a_i + B + \epsilon_i^a$, $B > 0$ and believes with certainty that $B = 0$.

Proposition 7

The long-run bias about the agent's own ability is

$$\tilde{a}_i - A_i = \frac{v_i^q + \sum_k \phi_{gik}^2 v_k^\eta}{v_i^a + v_i^q + \sum_k \phi_{gik}^2 v_k^\eta} B.$$

For $j \neq i$, the bias about individual j 's ability is

$$\tilde{a}_j - A_j = \frac{\sum_k \phi_{gik} \phi_{gjk} v_k^\eta}{v_i^q + \sum_k \phi_{gik}^2 v_k^\eta} (\tilde{a}_i - A_i),$$

while the bias about discrimination of type k is

$$\tilde{\theta}_k - \Theta_k = -\frac{\phi_{gik} v_k^\eta}{v_i^q + \sum_\ell \phi_{gik}^2 v_\ell^\eta} (\tilde{a}_i - A_i).$$

Discussion: endogenous overconfidence

The relationship between social beliefs and overconfidence is the same as in the basic model. Hence, the predictions about social beliefs and the comparative statics seen thus far remain unchanged.

Corollary 1

Making the agent's own recognition a more precise signal of calibre—that is, lowering v_i^q —reduces his overconfidence and increases all his other biases.

- **The basic-model intuition.** If an outsider makes the agent more realistic about himself, all of his social prejudices decrease.
- **Why the result reverses here.** The agent explains his unexpectedly low performance using both discrimination and bad luck. When his own signal becomes more precise, bad luck becomes a less plausible explanation.
- He therefore relies more heavily on discrimination to explain the same outcome, increasing his other biases. **Reducing prejudice requires addressing the underlying misspecification:** overconfidence in the basic model, or the agent's misinterpretation of signals about himself in this variant.

Short-run beliefs

Consider a single type of discrimination, $K = 1$. Let $\theta^i(t)$ and $a_j^i(t)$ denote the beliefs of a correctly specified agent, and let $\tilde{\theta}^i(t)$ and $\tilde{a}_j^i(t)$ denote the beliefs of the overconfident agent. Let the prior variances of discrimination and j 's ability by $\bar{v}^\theta, \bar{v}_j^a$.

Proposition 8

The differences between the two agents' short-run beliefs are

$$\tilde{\theta}^i(t) - \theta^i(t) = \beta_t(\tilde{\theta}^i - \Theta),$$

and, for $j \neq i$,

$$\tilde{a}_j^i(t) - a_j^i(t) = \beta_t \frac{t/v_j^q}{1/\bar{v}_j^a + t/v_j^q} (\tilde{a}_j^i - A_j).$$

Where

$$\beta_t = \frac{t/v^\eta}{1/\bar{v}^\theta + t/v^\eta + \sum_j \phi_{gj}^2 \frac{1}{\bar{v}_j^a} \frac{1/v_i^q}{1/\bar{v}_i^a + t/v_i^q}} \in (0, 1).$$

Short-run versus long-run prejudice

- **Same direction of bias.** The short-run bias about discrimination has the same sign as the long-run bias:

$$\text{sign}\left(\tilde{\theta}^i(t) - \theta^i(t)\right) = \text{sign}\left(\tilde{\theta}^i - \Theta\right).$$

- **Smaller magnitude.** Since $\beta_t < 1$ and

$$\frac{t/v_j^q}{1/\bar{v}_j^q + t/v_j^q} < 1,$$

short-run biases about others' abilities are also attenuated.

- **Learning gradually reveals the bias.** As t increases, the attenuation factors approach one. Short-run beliefs therefore converge to the long-run beliefs.
- **Main qualitative results survive.** Even before beliefs fully converge, the model preserves the direction of the biases and the comparative-statics results derived for the long run.

Concluding remarks

- **A simple mechanism can generate systematic prejudice.** An overconfident agent interprets unexpectedly poor outcomes as evidence of discrimination and unexpectedly good outcomes as evidence of ability.
- **Self-serving learning shapes social beliefs.** This generates in-group bias, negative views of perceived competitors, and systematic misperceptions about discrimination.
- **Information has nuanced effects.** Better information about one dimension can reduce a bias there while shifting bias toward other explanations or groups. Balanced information is unambiguously beneficial.
- **The source of the bias matters.** Reducing prejudice requires addressing the agent's self-misperception, rather than correcting beliefs about other groups alone.