Overconfidence and Prejudice

Paul Heidhues[1] Botond Kőszegi[2] Philipp Strack[3]

[1]Düsseldorf Institute for Competition Economics

[2]Central European University

[3]Yale University

Motivation

- Attribution bias: Overconfidence systematically biases beliefs about external factors.
- We develop this idea into a **theory of prejudice and discrimination**:
  - Agent is observing society, learning from what he sees.
  - Single non-standard assumption: he's overconfident.
- Main mechanism:
  Agent thinks that his outcomes are not good enough
  $\Rightarrow$ he overestimates discrimination against his group(s);
  $\Rightarrow$ he overestimates in-group members.
- Such beliefs will manifest themselves in discrimination and conflict
  — but we intentionally consider on a model without actions, to focus on beliefs.

Motivation - Why

- This theory explains patterns in beliefs that are not predicted by *statistical* or *taste based* discrimination:
  1. Prejudiced (more negative) views of other groups.
  2. Disagreement about degrees of discrimination.
- Makes a variety of unique and subtle testable predictions:

  In-group bias, Bias-substitution, .
- Deepens our understanding of prejudice and discrimination and complements existing theories by clarifying the role ego-centric biases can play.
- Illustrates that misspecification can play an important role in understanding discrimination.

Insights

1. **Positively:** testable novel theory of how biases about other people and groups can be caused by overconfidence.

2. **Conceptually:** framework to think about how a person's biases depend on his position in society.

3. **Methodologically:** derive long-run beliefs in a high-dimensional model of misspecified learning.

Related Literature

1. Misspecified Learning
   - Berk (1966), Esponda and Pouzo (2016), Heidhues, Kőszegi and Strack (2018b), Fudenberg, Romanyuk and Strack (2017), Heidhues, Koszegi and Strack (2018a), Bohren (2016), Bohren and Hauser (2019a), Esponda and Pouzo (2019), He (2019), Frick, Iijima and Ishii (2019b,a), Esponda, Pouzo and Yamamoto (2019)
   - Here: no actions, but a high dimensional learning environment

Related Literature

1. Misspecified Learning
2. Prejudice and group biases
   - Frick, Iijima and Ishii (2022), Bohren and Hauser (2019b), Bohren, Haggag, Imas and Pope (2023)
   - Large literature mostly in psychology and sociology on prejudice and stereotypes
3. Statistical discrimination
   - Arrow (1973) and Phelps (1972)
   - Discrimination arises from updating based on group characteristics
   - Here: biases arise from updating based on misspecified model
4. Taste-based discrimination
   - Becker (1957)
5. Implications of overconfidence
   - Novel observation that overconfidence can cause prejudice

Model

Model of Social Inferences

- G **groups** and I **agents**, each with
  1. fixed "ability" $a_i \in \mathbb{R}$.
  2. group membership $g_i \in \{1, \ldots, G\}$

- **Signals**: Each agent i observes iid Normal signals of
  1. Agent j's "success"

$$q_j = a_j + \sum_{k=1}^{K} \phi_{g_j k}\, \theta_k + \epsilon_j^q,$$

  2. The intensity θ_k of type k discrimination

$$\eta_k = \theta_k + \epsilon_k^\eta,$$

- **Discrimination**:
  - $\phi_{gk} \in \mathbb{R}$ extend to which group g benefits from discrimination of type k.
  - Is redistributive $\sum_g m_g \phi_{gk} = 0$, where m_g is the share of group g.

Information and Learning

- **Overconfidence**: i's ability is a_i, but he believes it's $\tilde{a}_i > a_i$.

- **Full Support**:
  - i doesn't know other a_j, θ_k, or the variance & correlation of ϵ.
  - updates according to Bayes' Rule.
  - has full-support prior.
- We solve for long-run (limiting) beliefs.

A Few Comments

- Interpret main concepts generally.
  - a_j = ability/effort, deservingness of society's recognition (for past behavior/work/honesty).
  - q_j = achievement, social respect (transfers, perks, etc.).
- Key assumption: **persistent overconfidence**:
  - Supported by plenty of evidence from psychology, experimental economics, field settings.
  - Capture as point beliefs, a technically convenient reduced form.
  - Microfoundations: biased learning (in the paper), selective memory (Fudenberg et al, 2023).
- **We don't assume there's no discrimination.**
  - Results are about views relative to truth and others' views.
- We allow for signals about discrimination to discipline beliefs.
  - Could be arbitrarily uninformative (variance of ϵ^η large).
- Our model of discrimination can be derived from a model of competition (details).

Characterization of Long-Run Beliefs

Theorem (Long-Run Biases)

Agent i's beliefs concentrate on a single $(\tilde{\theta}^i, \tilde{a}^i, \tilde{\Sigma})$ almost surely. His long-run bias about discrimination of type k is

$$\tilde{\theta}^i_k - \theta_k = \frac{-\phi_{g_i k} v^\eta_k}{v^q_i + \sum_{k'} \phi^2_{g_i k'} v^\eta_{k'}} \cdot (\tilde{a}_i - A_i), \tag{1}$$

and his long-run bias about individual j's caliber is

$$\tilde{a}^i_j - A_j = \frac{\sum_k \phi_{g_i k} \phi_{g_j k} v^\eta_k}{v^q_i + \sum_{k'} \phi^2_{g_i k'} v^\eta_{k'}} \cdot (\tilde{a}_i - A_i). \tag{2}$$

- $v^\eta_k = var(\epsilon^\eta_k), v^q_i = var(\epsilon^q_i)$.
- Agent i's long-run beliefs are $\tilde{\theta}^i_k, \tilde{a}^i_j$.

Economic Implications

Implication 1: Self-Centered Views about Discrimination

$$\tilde{\theta}^i_k = \theta_k - \frac{\phi_{g_i k} v^\eta_k}{v^q_i + \sum_{k'} \phi^2_{g_i k'} v^\eta_{k'}} \cdot (\tilde{a}_i - A_i)$$

- Agent i
  1. underestimated discrimination if he benefits from it $\phi_{g_i k} > 0$;
  2. overestimates discrimination if he suffers from it $\phi_{g_i k} < 0$;
  3. underestimates discrimination against competitor groups, e.g. groups g' with $\phi_{g_i k}\phi_{g' k} < 0$.

- **Intuition:**
  - i is prone to feeling that his recognition is too low.
  - Discrimination explains this perceived injustice.

- **Predictions:**
  - Agent views discrimination against own group as worse than outsiders do.
  - Supported by plenty of evidence from opinion surveys.

Examples of Survey Evidence

❶ Racial Discrimination:

- 88% of blacks say that "the country needs to continue making changes to give blacks equal rights with whites," while only 54% of whites and 69% of Hispanics do — and the gap used to be even higher (Pew Research Center, 2017, Chapter 4).
- 70% of blacks, but only 37% of whites, say that blacks are treated less fairly by police than whites, with similar gaps regarding the treatment of blacks in courts, stores, public schools, health care, and on the job (Anderson, 2014).
- The majority of whites thinks that whites are discriminated against in America today, although most holding this opinion have not personally experienced specific discrimination (National Public Radio et al., 2017).

❷ Gender Discrimination:

- Similar gaps in opinions exist regarding gender discrimination.
- In a survey of STEM employees, 83% of men think that in their workplace women are usually treated fairly in the recruitment and hiring process, but only 67% of women think so; and 77% of men say that women are treated fairly in opportunities for promotion and advancement, but only 43% of women agree (Funk and Parker, 2018).

Aside: Relation to Other Models of Discrimination

- Classical theories:
  1. Statistical discrimination.
  2. Taste-based discrimination.
- Neither explains biased beliefs.
  1. Statistical discrimination: beliefs should be on average correct.
  2. Taste-based discrimination: beliefs play no role.
  3. Our theory predicts beliefs are biased in very specific directions.
- We hope that misspecification can help explain patterns not explained by standard theories in the discrimination literature and thereby complement them.

Implication 2: In-Group Bias

- Assume that information about group members is homogenous $g_i = g_j \Rightarrow v_i^q = v_j^q$.
- A_g be the average ability of group g,
- $\tilde{a}_{g'}^{g}$ the average opinion of group g about group g'

Proposition

1. ***In-Group Overestimation:*** *Each group overestimates itself relative to the truth ($\tilde{a}_g^g > A_g$), but on average estimates groups correctly ($\sum_{g'} m_{g'} \tilde{a}_{g'}^{g} = \sum_{g'} m_{g'} A_{g'}$).*

2. ***Absolute In-Group Bias:*** *If groups' average ability (A_g) are equal, then each group thinks others in their group are better than the average ($\tilde{a}_g^g > \sum_{g'} m_{g'} \tilde{a}_{g'}^{g}$).*

3. ***Relative In-Group Bias:*** *On average, a group's view of its fellow members relative to another group's members is positive: $\sum_{g,g'} m_g m_{g'} (\tilde{a}_g^g - \tilde{a}_{g'}^{g}) > 0$.*

Intuition

- Part I says that on average, an agent overestimates other members of his group.
- Intuitively, as they are subject to the same discrimination, he overestimates discrimination hurting fellow group members.
- Hence, attributes too much of their observed outcomes to ability.
- Agent understands that discrimination is redistributive.
- As a result, he estimates total caliber in the population correctly.
- Part II: If the average ability of groups are equal, then a person estimates his group to be above this level, and other groups to be below it on average.
- More generally, the average person estimates the average other member of his group to be better than average (Part III).

Evidence

- **Absolute in-group bias:**
  - Stylized fact in literature on prejudice.
  - Goes back to Sumner (1906), Allport (1954), Tajfel (1982), and is documented in many studies.
  - Also consistent with evidence on discriminatory behavior.

- **Relative in-group bias:**
  - When we don't see absolute in-group bias, there is usually relative in-group bias (see e.g. Zussman (2013)).

Implication 3: Biases Derive from Overconfidence

Monotonicity in overconfidence:

- i's biases are increase in his overconfidence $\tilde{a}_i - a_i$,
- or his perception that he gets less than he deserves.
- "impostor syndrome"/underconfidence leads to opposite biases.
- Consistent with evidence that narcissistic individuals are more prejudiced (Cichocka et al., 2017).

Implication 4: The role of Competition

- Suppose that groups g and g' are initially not affected by the same types of discrimination, i.e., $\phi_{gk}\phi_{g'k} = 0$ for all k.
- A new type of discrimination $K+1$ arises that pits groups g and g' on opposite sides: $m_g\phi_{gK+1} + m_{g'}\phi_{g'K+1} = 0$, with $\phi_{gK+1} \neq 0$.

Proposition

The new type of discrimination:

1. ***Competition Effect:*** *Lowers the view of group g about group g'.*
2. ***Excuse Effect:*** *Raises the view of group g about itself.*
3. ***Bias Substitution:*** *Raises the average view of group g about groups other than g and g'.*

Intuition & Discussion

- A member of group g overestimates discrimination in favor of group g'.
- This lowers group g's opinion of group g' (Part I).
- Thinking of others as competitors lowers agent's opinion of them.
  - This is a basic tenet of group conflict theory.
  - Example: racial mixing and immigration increase animus (Branton and Jones, 2005, Tabellini, 2019).
  - Esses et al. (1998) find that manipulating the sense of competition with an immigrant group leads subjects to see the group in a more negative light.
- New competition provides the person a new explanation for his and other group members low recognition (Part II).
- Part (III): less bias about other groups as $K + 1$ can in part explain "too low" outcomes.

Implication 6: Bias Substitution

- **Changes in environment can lead to reallocating biases.**
- There is a new social group that's common competitor of existing groups (e.g., new immigrants).
  - This naturally introduces a new type of discrimination, namely discrimination against the new group.
  - *i* comes to view new group negatively but everyone else positively.
- Example: inflow of blacks into northern U.S. cities reduced stereotyping of Irish and Italians (Fouka et al., 2019).

▸ Skip to Personal Contact

Implication 5: Useless/Harmful Types of Information

The effect of information:

- Better info about others' outcomes ($\downarrow v_j^q$) **has no effect** on i's biases.
- Better info about own outcomes ($\downarrow v_i^q$) **increases all of i's biases.**
  - "Bad luck" becomes worse explanation for low recognition, increasing need for other explanations.
- More info about group k discrimination ($\downarrow v_k^\eta$) with $\phi_{g,k} \neq 0$
  1. decreases bias about group k discrimination
  2. raises his bias regarding any other type of discrimination that affects him.
  3. Raises his bias about the average ability of any group g not affected by discrimination of type k.
  4. "bias substiution": Male white university professor learns that there is no-discrimination against men and now believes more in discrimination against whites.

Extensions

A Model with Characteristics

- Individual j has characteristics $c_j = (c_{j1}, \ldots, c_{jK}) \in \{0,1\}^K$.
- $c_{jk} = 1$ means she has characteristic k (e.g., is female or black).
- A group consists of individuals who share all characteristics.
- Discrimination of type k redistributes recognition between individuals who have characteristic k and those who do not

$$\phi_{ck} = \begin{cases} +1 & \text{if } c_k = 1 \\ -1 & \text{if } c_k = 0 \end{cases}.$$

- **Definition:** Agent i is more similar to individual j than to individual j' if whenever i and j' share a characteristic, so does j.

Proposition (Similarity Bias)

Consider the model with characteristics. If agent i is (strictly) more similar to j than to j', then i's long-run bias regarding the caliber of j is (strictly) greater than his long-run bias regarding the caliber of j', i.e., $\tilde{a}^i_j - A_j \geq \tilde{a}^i_{j'} - A_{j'}$ ($\tilde{a}^i_j - A_j > \tilde{a}^i_{j'} - A_{j'}$).

- Similarity bias — that a person has a more positively biased opinion about more similar others.
-

Personal Contact

- Extension of model: *i* also observes signals of (some) individuals' calibers.
  - Can only solve special cases: two competing groups, observe equally informative signal about everyone else's ability.
- Then, biases vis a vis relevant group decrease.
  - Similarly, observing more individuals lowers biases.
- Consistent with evidence for Allport's (1954) "contact hypothesis."
- Now, better information about the recognition of an out-group can increase prejudice against it.
- General conclusion: **whether information is beneficial crucially depends on its nature.**

▸ Skip Correlated Errors

Multidimensional Misspecified Learning

Multidimensional Misspecified Learning

- An agent makes inferences about a fixed vector of **fundamentals**

$$f = (f_1, \ldots, f_L)^T \in \mathbb{R}^L,$$

- In each period t, he observes a *signal*

$$r_t = M f + \epsilon_t \,,$$

- $M \in \mathbb{R}^{D \times L}$ is a matrix with rank L
- $\epsilon_t \in \mathbb{R}^D$ normally distributed with mean zero and covariance matrix Σ
- Updates beliefs using Bayesian rule for a prior P_0 over (f, Σ)
- **Assumption:** Agent is misspecified and dogmatically believes $f_i = \tilde{f}_i$
- **Question:** What are the agent's long-run beliefs?

Theorem (Long-Run Beliefs)

The agent's beliefs concentrate on a single point $(\tilde{f}, \tilde{\Sigma})$*. Furthermore:*

1. *If the agent has fixed beliefs* $\tilde{\Sigma}$*, but is uncertain about* $(\tilde{f}_j)_{j \neq i}$

$$\tilde{f}_j - f_j = \frac{(M^T \tilde{\Sigma}^{-1} M)^{-1}_{ij}}{(M^T \tilde{\Sigma}^{-1} M)^{-1}_{ii}} (\tilde{f}_i - f_i). \tag{3}$$

2. *If the agent has fixed beliefs about* $(\tilde{f}_j)_{j \neq i}$ *but is uncertain about* $\tilde{\Sigma}$

$$\tilde{\Sigma} - \Sigma = (M(\tilde{f} - F))(M(\tilde{f} - F))^T. \tag{4}$$

3. *If the agent is uncertain about both* $(\tilde{f}_j)_{j \neq i}$ *and* $\tilde{\Sigma}$

$$\tilde{f}_j - f_j = \frac{(M^T \Sigma^{-1} M)^{-1}_{ij}}{(M^T \Sigma^{-1} M)^{-1}_{ii}} (\tilde{f}_i - f_i), \tag{5}$$

and his bias about the covariance matrix is given by (4)*.*

Conclusion

What did we do:

1. Theory of prejudice resulting from overconfidence.
2. Framework to think about individuals social biases.
3. Long-run beliefs in high-dimensional misspecified learning model.

Remarks:

- Testable specific predictions about social biases.
- In line with some evidence.
- Complements *statistical* and *taste-based discrimination*, which do not explain biased beliefs.

Question:

- Social interaction.
- Endogenous groups.
- Self-fulfilling prophecies.
- Conspiracy theories.

Thank You

References

Allport, Gordon W., *The Nature of Prejudice*, Addison-Wesley Publishing Company, Inc., 1954.

Anderson, Monica, "Vast Majority of Blacks View the Criminal Justice System as Unfair," Technical Report, Pew Research Center 2014.

Arrow, Kenneth J., "The Theory of Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*, Princeton University Press, 1973.

Becker, Gary S., *The Economics of Discrimination*, Chicago: University of Chicago Press, 1957.

Berk, Robert H, "Limiting Behavior of Posterior Distributions when the Model is Incorrect," *The Annals of Mathematical Statistics*, 1966, *37* (1), 51–58.

Bohren, J Aislinn, "Informational Herding with Model Misspecification," *Journal of Economic Theory*, 2016, *163*, 222–247.

Bohren, J. Aislinn and Daniel Hauser, "Misinterpreting Social Outcomes and Information Campaigns," 2019. Working Paper.

Bohren, J Aislinn and Daniel N Hauser, "Misinterpreting Social Outcomes and Information Campaigns," Technical Report, Mimeo 2019.

___ , Kareem Haggag, Alex Imas, and Devin G Pope, "Inaccurate statistical discrimination: An identification problem," *Review of Economics and Statistics*, 2023, pp. 1–45.

Branton, Regina P. and Bradford S. Jones, "Reexamining Racial Attitudes: The Conditional Relationship Between Diversity and Socioeconomic Environment," *American Journal of Political Science*, 2005, *49* (2), 359–372.

Cichocka, Aleksandra, Kristof Dhont, and Arti P. Makwana, "On Self-Love and Outgroup Hate: Opposite Effects of Narcissism on Prejudice via Social Dominance Orientation and Right-Wing Authoritarianism," *European Journal of Personality*, 2017, *31* (4), 366–384.

Esponda, Ignacio and Demian Pouzo, "Berk–Nash equilibrium: A Framework for Modeling Agents with Misspecified Models," *Econometrica*, 2016, *84* (3), 1093–1130.

___ and ___ , "Equilibrium in Misspecified Markov Decision Processes," *arXiv preprint arXiv:1502.06901*, 2019.

___ , ___ , and Yuichi Yamamoto, "Asymptotic Behavior of Bayesian Learners with Misspecified Models," 2019.

Fouka, Vasiliki, Soumyajit Mazumder, and Marco Tabellini, "From Immigrants to Americans: Race and Assimilation during the Great Migration," 2019. Working Paper.

References ii

Frick, Mira, Ryota Iijima, and Yuhta Ishii, "Misinterpreting others and the fragility of social learning," 2019.

___ , ___ , and ___ , "Stability and Robustness in Misspecified Learning Models," 2019. Working Paper.

___ , ___ , and ___ , "Dispersed behavior and perceptions in assortative societies," *American Economic Review*, 2022, *112* (9), 3063–3105.

Fudenberg, Drew, Gleb Romanyuk, and Philipp Strack, "Active Learning with a Misspecified Prior," *Theoretical Economics*, 2017, *12* (3), 1155–1189.

Funk, Cary and Kim Parker, "Women and Men in STEM Often at Odds Over Workplace Equity," Technical Report, Pew Research Center January 2018.

He, Kevin, "Mislearning from Censored Data: The Gambler's Fallacy in Optimal-Stopping Problems," 2019.

Heidhues, Paul, Botond Koszegi, and Philipp Strack, "Convergence in Misspecified Learning Models with Endogenous Actions," *Available at SSRN 3312968*, 2018.

___ , Botond Kőszegi, and Philipp Strack, "Unrealistic Expectations and Misguided Learning," *Econometrica*, 2018, *86* (4), 1159–1214.

National Public Radio, Robert Wood Johnson Foundation, and Harvard T.H. Chan School of Public Health, "Discrimination in America: Experiences and Views of White Americans," Technical Report November 2017.

Pew Research Center, "The Partisan Divide on Political Values Grows Even Wider," Technical Report October 2017.

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

Sumner, William G., *Folkways*, New York: Ginn, 1906.

Tabellini, Marco, "Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration," *Review of Economic Studies*, 2019, *87* (1), 454–486.

Tajfel, Henri, "Social Psychology of Intergroup Relations," *Annual Review of Psychology*, 1982, *33* (1), 1–39.

Zussman, Asaf, "Ethnic Discrimination: Lessons from the Israeli Online Market for Used Cars," *Economic Journal*, 2013, *123* (11), 433–468.

Correlated Errors

- Suppose no groups ($K = 0$), so i observes unbiased signals of individuals' calibers ($q_j = a_j + \epsilon_j$).
- But generalize correlation structure: ϵ_j have covariance matrix Σ^q.

Proposition
Agent i's long-run bias about individual j is

$$\tilde{a}^i_j - a_j = \frac{\Sigma^q_{ij}}{\Sigma^q_{ii}}(\tilde{a}_i - a_i),$$

while his bias about the covariance matrix is

$$\tilde{\Sigma}^q_{jj'} - \Sigma^q_{jj'} = (\tilde{a}^i_j - a_j)(\tilde{a}^i_{j'} - A_{j'})\,.$$

Micro-foundation for Discrimination Model

- We take ϕ_k as exogenous, but they can be derived from a model of competition.
- Left $f(g, g')$ measure how often an individual of group g and g' compete.
- Let G_k be the groups benefitting from discrimination of type k
- We can define

$$\phi_{gk} = \begin{cases} \sum_{g' \in G \setminus G_k} f(g, g') & \text{if } g \in G_k, \text{ and} \\ -\sum_{g' \in G_k} f(g, g') & \text{if } g \in G \setminus G_k. \end{cases} \tag{6}$$

- Intuitively, the impact of discrimination of type k on an individual is determined by how many people he tends to compete with on the other side of the issue.
- Consistent with fierce competition with other members of their own group (e.g., whites compete with each other for college spaces), i.e., $f(g, g)$ is high.

Correlated Errors

- Key implication: *i* **overestimates (underestimates)** *j* **if** *j***'s recognition is positively (negatively) correlated with his own.**
- Intuition:
  - Suppose *i* knows Σ^q, and q_i and q_j are positively correlated.
    - *i* thinks q_i is systematically too low.
    - He concludes q_j must be systematically too low as well.
    - So he overestimates *j*.
  - But *i* misinfers Σ^q: overestimates covariance between q_j and $q_{j'}$ iff he misestimates *j* and *j*′ in same direction.
    - Suppose he overestimates both individuals.
    - Then, in a prototypical observation both q_j and $q_{j'}$ seem to him to be too low and therefore positively correlated.