

Bartos, Wojtech

Working Paper

Breaking Bias: Pathways to Reducing Discrimination

CESifo Working Paper, No. 12045

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Bartos, Wojtech (2025) : Breaking Bias: Pathways to Reducing Discrimination, CESifo Working Paper, No. 12045, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/325036>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

CES ifo

**12045
2025**

Juli 2025

Working Papers

Breaking Bias: Pathways to Reducing Discrimination

Vojtech Bartos

CES ifo

Imprint:

CESifo Working Papers

ISSN 2364-1428 (digital)

Publisher and distributor: Munich Society for the Promotion
of Economic Research - CESifo GmbH

Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740

Email office@cesifo.de
<https://www.cesifo.org>

Editor: Clemens Fuest

An electronic version of the paper may be downloaded free of charge

- from the CESifo website: www.ifo.de/en/cesifo/publications/cesifo-working-papers
- from the SSRN website: www.ssrn.com/index.cfm/en/cesifo/
- from the RePEc website: <https://ideas.repec.org/s/ces/ceswps.html>

Breaking Bias: Pathways to Reducing Discrimination

Vojtěch Bartoš¹

¹Department of Economics, Management, and Quantitative Methods, University of Milan, Via Conservatorio 7, 20122 Milan, Italy. Email: vojtech.bartos@unimi.it. ORCID ID: 0000-0002-1638-5983

Abstract

Discrimination remains a persistent challenge with significant moral and economic consequences. While extensive economics research has documented its existence, identifying effective strategies for reducing discrimination remains an evolving field. This chapter synthesizes theoretical perspectives from economics and psychology, integrating them with recent empirical findings on discrimination reduction. It reviews interventions that mitigate bias by manipulating the cost of discrimination, shaping information processing, leveraging intergroup contact, changing attitudes, and manipulating identity perceptions. The discussion covers taste-based and statistical discrimination models, implicit bias frameworks, and psychological theories of discrimination. A particular emphasis is placed on well-identified and sufficiently powered studies, including field experiments with incentivized measures. By bridging insights across disciplines, this chapter provides a comprehensive understanding of effective discrimination reduction strategies, highlights remaining gaps, and outlines directions for future research to develop scalable solutions.

Keywords: Discrimination reduction; Bias mitigation; Statistical and taste-based discrimination; Implicit bias; Intergroup contact

The final version of this chapter is forthcoming in Alessandro Bucciol and Simone Quercia (Eds.)

Research Handbook on Unethical Behavior. Edward Elgar.

Chapter 14

Breaking Bias: Pathways to Reducing Discrimination

14.1 Introduction

Discrimination of minority or underrepresented groups remains prevalent in today's world, underscoring the need for effective paths toward combating it. Beyond the moral imperative, discrimination imposes significant economic costs, leading to inefficiencies through reduced effort, self-confidence, and productivity (Hjort, 2014; Glover et al., 2017; Carlana, 2019). This chapter examines theoretical predictions from economics and psychology on reducing intergroup discrimination, reviews recent empirical evidence on effective methods, and highlights areas where further understanding is needed. Bertrand and Duflo (2017) argue that, while the existence of discrimination is well documented through rigorous field experiments across diverse fields, we lag behind in identifying what effectively reduces discrimination—a domain "ripe for picking". The inspiration from psychology at the time has mostly focused on contact hypothesis (Allport, 1954), although many other strategies have been proposed and tested, typically using small samples in laboratory environments. Some researchers have risen to the challenge and addressed some of these gaps. Nevertheless, as will become evident, there remain some fruits to be picked.

I review recent contributions in the area of discrimination reduction, aiming to link the evidence to existing theories in economics (Section 14.2) and psychology (Section 14.3). However, observed behavior can often be challenging to map directly onto specific theories, making it difficult to generalize

findings from individual empirical or experimental studies. Moreover, the theories themselves have evolved, offering new nuances that researchers can leverage when studying interventions to reduce bias.

The experimental and empirical evidence reviewed examine discrimination targeting individuals based on race, ethnicity, gender, religion, nationality, migration background, sexual orientation and identity, and income. Those who engage in discriminatory behavior include laboratory or online experiment participants, members of the general population, customers, employers, classmates, teachers, police officers, army cadets, judges, sports players, and referees. I focus on decision-making by individuals who have already developed discriminatory preferences, thus omitting an alternative approach: preventing the formation of discriminatory preferences in children during socialization. Fehr et al. (2008) demonstrate that parochialism emerges at a very young age, indicating that effective interventions would need to target pre-school children. This area offers ripe fruit for further exploration.

The approaches I cover are traditional interventions grounded in standard economic models of discrimination, such as manipulating the cost of discrimination (Section 14.4.1) and the provision (Section 14.4.2) or withholding (Section 14.4.3) of information. Specifically, I focus on the formation of prior beliefs, determinants of information processing—or lack of thereof—and how and whether individuals update their beliefs in light of new information. In the context of withholding information, I distinguish between mandated withholding required by laws and regulations, and strategic withholding or identity manipulation by individuals. Building on insights from psychology, I examine recent field experiments on the role of intergroup contact (Section 14.4.4). While intergroup contact has demonstrated positive effects, these interventions typically involve prolonged exposure. I therefore turn to shorter-term interventions aimed at reducing bias by directly altering attitudes (Section 14.4.5), including perspective-taking, habit-breaking, evaluative conditioning, exposure to role models, and the use of narratives. Additionally, I discuss how reduction of bias counteracts the adverse psychological effects of exposure to discrimination. Finally, I connect the evidence thus far to the concept of identity (Section 14.4.6), exploring how identity is malleable to exogenous shocks, and considering how manipulating perceptions of self and others' identities can be leveraged for positive outcomes—or, regrettably, for extremely negative ones. I discuss ideas for further research throughout the text and offer broader perspectives in the conclusions (Section 14.5).

It is important to clarify what this chapter covers and what it does not. It offers a perspective distinct from other literature reviews on discrimination by focusing strictly on ways to reduce dis-

crimination, rather than merely documenting its existence. I encourage readers to consult the other reviews for a more comprehensive understanding. Charles and Guryan (2011), Lang and Lehmann (2012), and Neumark (2018) review literature on discrimination in labor markets, while Lang and Kahn-Lang Spitzer (2020) focus on top also on criminal justice system. Bertrand and Duflo (2017) offer a comprehensive and updated review of field experiments on the topic. Li (2020) reviews literature on group identity and intergroup bias. Finally, Onuchic (2024) discusses recent advances in economic modeling of discrimination, specifically focusing on the role of misspecified models and also on the role of algorithmic decision-making. My focus also differs from Paluck et al. (2021) who exclusively focus on reviewing discrimination reduction literature in psychology. I extend this by focusing on both economics and selected work in psychology and political science. I also take seriously a main critique of Paluck et al. (2021) that too many studies use relatively small samples sizes, non-representative samples, rely on self-reported measures as outcomes, or are not pre-registered. Therefore, I put emphasis on well identified and sufficiently powered studies and often using samples beyond college students and convenience online workers. By emphasizing rigorous methodologies and integrating perspectives from both economics and psychology, I aim to summarize and challenge the discourse on effective strategies for reducing discrimination.

This chapter is also not a meta analysis. For those, I refer the reader to studies such as Paluck et al. (2021) on experiments on bias reduction in psychology, Lippens et al. (2023) on correspondence experiments, or Lane (2016) on laboratory experiments. Forscher et al. (2019) use a network meta-analysis to study effectiveness of bias reduction interventions on implicit bias using the Implicit Association Test (IAT; Greenwald et al., 1998). Achieving a comprehensive meta-analysis of discrimination reduction approaches would require a more harmonized outcome variable, such as the IAT for implicit bias, callbacks for correspondence studies, or monetary transfers in standardized experimental games. Although recent research offers a variety of innovative measures, the lack of harmonization complicates comparison across—and sometimes even within—studies.

14.2 Economic theories of discrimination and mitigation strategies

The foundational economic theories explaining why people discriminate are taste-based (Becker, 1971) and statistical (Phelps, 1972; Arrow, 1973) models. More recent theories have also emerged, which

I summarize here along with their implications for discrimination reduction strategies. Bohren et al. (2023b) propose a method for decomposing discrimination into its direct and systemic components. Systemic components capture the effects of discrimination that targets have experienced in the past, are currently experiencing, or expect to encounter in the future—factors that may influence their investment decisions and behavior. In this chapter, I focus exclusively on direct effects, though successful mitigation of direct discrimination should have positive spillover effects in the systemic domain.

14.2.1 Taste-based discrimination

Taste-based model assumes that distaste towards a minority enters directly into decision makers' (DM)¹ utility function in a form of a utility penalty. In other words, when interacting with an individual from a majority group $g = maj$, DMs utility is determined by the quality q of a relationship with an individual² and the utility function is $U_{maj} = q_{maj}$. In contrast, interacting with an individual from a minority group $g = min$ carries a penalty d_g ³ and utility is $U_{min} = q_{min} - d_{min}$. There is thus an interval where $q_{min} > q_{maj} \geq q_{min} - d_{min}$ where a majority individual of relatively lower quality is preferred to a higher quality minority individual. This also implies that DMs are willing to accept a loss—monetary or otherwise—to avoid an interaction with a member of a disliked group. Full information is assumed, hence $q_{maj} = E[q_{maj}]$.

A policy aimed at reducing taste-based discrimination at an individual level needs to affect the penalty d ⁴ or to manipulate the cost of the discriminatory act. Affecting the penalty may operate either via its direct reduction, for example by inducing empathy toward a discriminated group.⁵ Alternatively, it can be affected by manipulating identity of the target. This could operate via increasing salience of some other identity—selected from a range of social identities an individual may be associated with (Akerlof and Kranton, 2000, p. 731)—that can be used to describe the same individual. Finally, target's identity can be hidden altogether. In contrast, manipulating costs could operate by direct and enforceable anti-discrimination policies that exist to some extent in most countries, by resorting to

¹For the sake of simplicity, I assume throughout the text that DMs come from a majority group. The model can easily be generalized.

²For example, in labor markets the quality would be determined by applicants' productivity or returns to the worker, in education this would be student's quality, in sports this would be leniency of referees, in marriage or partnership markets this would be an expected quality of a match on whatever dimension deemed relevant to the DM, etc.

³For a majority, we can assume without loss of generality that $d_{maj} = 0$.

⁴While traditional economic theory relies on preference stability (Stigler and Becker, 1977), recent advances in empirical literature have been questioning the assumption (Chuang and Schechter, 2015; Schildberg-Hörisch, 2018).

⁵I use the term empathy to refer to an individual's ability to understand and respond to the experiences and feelings of another person or group. While empathy does not necessarily lead to action, it can facilitate helping behaviors toward that person or group.

social sanctioning, or by directly increasing the monetary cost of discrimination. In a market setting, discriminating firm may be driven out from the market by non-discriminating competitors (Becker, 1971).

14.2.2 Statistical discrimination

Statistical models, in contrast, assume that DMs do not have full information about individuals and instead infer characteristics based on the average quality and its distribution within a group., $E[q_g] = \alpha_g$, with $q_g \sim N(\alpha_g, \sigma_{q_g}^2)$. In the absence of individual signals, a rational DM compares $E[q_g]$'s for evaluated individuals and picks the one with higher $E[q_g]$. Effect of information provision depends on signal informativeness. Namely, DMs only observe a noisy quality indicator y_i from which they try to extract information about underlying true individual quality q_i , with $y_i = q_i + u_g$, $u \sim N(0, \sigma_{u_g}^2)$. The signal is processed as a weighted average of the individual signal y_i and the group quality average α_g . The weight on group average corresponds to the informativeness of individual signal determined by $\frac{var(q)}{var(q)+var(u)}$, where higher $var(u)$ implies that DMs would put more weight on the group average quality α_g . In line with Cornell and Welch (1996), I let the variance of informativeness differ by group. It can well be that having less information about minority individuals to begin with makes signals about a minority individual less informative, relative to a majority applicant. I also do not make assumptions as to whether the beliefs about the averages and distributions are corrects or not (Bohren et al., 2023a) and how such beliefs are shaped (e.g., Eyting, 2022)⁶. When beliefs are correct, the observed discriminative behavior under statistical discrimination can be rationalized. However, already Phelps (1972, p. 661) argues against such simplistic interpretation and complacency: "Discrimination is no less damaging to its victims for being statistical". Eventually, even accurate statistical discrimination produces more discrimination down the line, even across domains as evidence of inferior quality of a specific group mounts, creating a vicious cycle. This can further be reinforced by self-confirming expectations on the side of members of a statistically discriminated group (Lundberg and Startz, 1983; Coate and Loury, 1993): if I know that my quality signal is less informative and is rewarded less, I will not invest in the quality in the first place. In practice, such individuals may not value education or training, further harming their and their groups' prospects. Further, individuals may also perform worse when stereotyped against.

A policy aimed at reducing taste-based discrimination should provide more information about an

⁶Also, invalid beliefs may also be shaped by invalid inference drawn from non-representative samples (Lang and Kahn-Lang Spitzer, 2020).

individual. Alternatively, if underlying beliefs about a group are biased, correcting such beliefs about the distribution would be useful. Lastly, if DMs find it easier to judge individuals from own group relative to individuals from other groups, perhaps due to language or cultural barriers, learning about other groups may reduce the uncertainty over interpretation of signals from outgroup members.

14.2.3 Attention discrimination

More recently, Bartoš et al. (2016) proposed a theory of attention discrimination. This extends the model of statistical discrimination by assuming that information processing is costly, building on models of rational inattention (for an excellent review see Maćkowiak et al., 2023). DMs decide whether to process individual quality signals only when the ex-ante expected value of the signal exceeds the cost of information processing. We test the model predictions in Internet field experiments on rental housing and labor markets and find supportive evidence. In “lemon dropping” markets where DMs screen out the worst quality individuals, more attention would be paid to signals that would be expected to change a prior belief to the domain of failure. In contrast, in “cherry picking” markets where only top quality individuals are selected, signals ex-ante assumed to be changing the prior in favor of the positive outcome would be paid more attention to. Consequently, even highly qualified individuals may not be paid attention to or bad individuals are not screened out, typically magnifying discrimination beyond what statistical models would assume. This asymmetry in attention further demonstrates why attention (and more broadly statistical) discrimination may be beneficial to private individuals but is harmful to the society. High quality individuals from a group discriminated against may, for example, remain unemployed for too long.

A policy aimed at reducing attention discrimination may take two forms. First, by hiding the signal about group attribute of an individual. For example, by removing any information about a group characteristic from a resume or from own profile, or blind auditions, a common practice in major orchestras nowadays. Second, by increasing salience of quality signals for individuals from negatively-stereotyped groups, who would otherwise not be paid attention to by DMs. In labor market setting, this could be done by highlighting qualities of an individual very early on in a cover email, even if the exact same information is provided in a resume. This would increase the chance of recruiters to pay attention to that particular resume that would otherwise go unnoticed to the rejection pile. Notice that such practices should not have any effect if the underlying discrimination is driven by DM tastes. A group attribute would anyways be revealed down the line and the negative penalty would apply in

DM’s final decision.

14.2.4 Implicit discrimination

The models discussed above all assume that the discriminatory decision-making process is a conscious act.⁷ In contrast, an implicit discrimination model, drawing on extensive psychology literature, challenges this assumption. It suggests that individuals often engage in subconscious processing of stimuli. The basic theory is mute on what shapes the attitudes and how malleable these are or should be. No formal model has been proposed, but it may be similar to that of taste-based discrimination, however, the penalty d_g applies only when a decision is processed unconsciously. That may happen when a decision is split second, routine, or habitual. This has similarities with two-system models (Bernheim and Rangel, 2004; Fudenberg and Levine, 2006) that model behavior as a conflict between a conscious self, and an automated and impulsive self. Implicit discrimination is typically measured using the Implicit Association Test (IAT) (Greenwald et al., 1998). IAT measures the speed of unconscious mental associations. The basic assumption is that an association is faster when it is compatible with the underlying mental model of an individual (such as that between a photo of a White person and a word “good”), while it is slower when the association goes counter to the model.

A policy aimed at reducing implicit discrimination may be the same as those proposed to manipulate the penalty d_g directly, as proposed for taste-based discrimination. On top, slowing-down decision making processes as well as attracting attention to the decision itself may both attenuate implicit discrimination, by shifting the decision-making to the conscious self, rather than the impulsive one that is prone to implicit bias. This could also be done by highlighting the possibility of implicit biases in general, or on an individual level.

14.3 Psychological theories of discrimination reduction

Before examining the empirical evidence on discrimination reduction, this section discusses the main theories on the topic as considered in psychology. I aim to link these theories to economic models, although they typically fall outside the standard framework of economic modeling. This section draws heavily on the most recent comprehensive review of the literature on prejudice reduction by Paluck et al. (2021). The chapter explores the most prominent models to explain the malleability of discrimination,

⁷Psychology would call conscious discriminative decision-making as “explicit discriminatory attitudes”.

including intergroup contact (both in-person and imaginary), cognitive and emotional training, social categorization, peer influence, and the roles of value consistency and self-worth.

14.3.1 Contact hypothesis

Literature in psychology posits that either interpersonal contact or imaginary contact can reduce discrimination. This idea is grounded in the theory of Allport (1954), who proposed that prejudice may be reduced if intergroup contact satisfies four facilitating conditions: it leads to a common goal, exhibits intergroup cooperation, has support from authorities, and involves individuals of equal status.⁸ The effects of interpersonal contact can operate through all the economic models discussed above. It may directly influence preferences, correct potentially mistaken belief distributions, enhance the DM's ability to process information from outgroup members through intergroup learning, and shape how mental associations regarding outgroups are formed and processed. The facilitating conditions address all these channels, while their absence may lead to less intense or less informative contact. The intensity of these effects will depend on initial conditions, such as baseline knowledge of the outgroup's quality distribution or the initial outgroup penalty d_{\min} . Thus, the effects of interventions may vary from group to group, and from setting to setting.⁹ Moreover, non-cooperative—or outright adversarial—contact may increase hostility toward outgroups through heightened parochialism. This mechanism has been documented in the most severe forms of adversarial intergroup contact, such as war (Bauer et al., 2016), but it may also operate in less intense non-cooperative interactions, such as competitive sports (Lowe, 2021). Distinguishing between imaginary contact and actual interpersonal contact is crucial, as imaginary interactions do not provide new information, potentially operating through different channels.

14.3.2 Cognitive and emotional training

This stream of thinking assumes that mental associations are malleable. It can be broadly categorized into several subgroups, with different implications for the economic fundamentals. Counteracting negative associations with outgroup members may be achieved by directly *altering cognitive associations*. If a typical association with a member of an outgroup is negative, actively exposing a DM to any positive signals alongside visualisations of members of the outgroup may affect such associations. Since no new

⁸Allport warns that if these conditions are not met, contact may even increase the likelihood of intergroup conflict.

⁹Suggestively, a meta-analysis by Paluck et al. (2019) shows that the effects of contact interventions are strongest for disability and weakest for race.

information is learned about a particular group and the explicit penalty d_{min} is unlikely to be affected, such model likely operates exclusively via unconscious channels. In case DMs also learn about a social norm shunning discriminative behavior, an associated cost of discrimination may increase for that DM.

An alternative is *habit-breaking* (e.g., Devine et al., 2012). This theory assumes that mere increased awareness of own biased behavior may lead to reduced discrimination. In the language of implicit discrimination, habit breaking is supposed to activate decision making by the conscious, rather than the unconscious self. Similarly, since emotions—or *affect*—may trigger bias, regulating emotions through conscious training may counteract unconsciously triggered discriminatory acts. As long as the emotion regulation operates through channels not directly related to discrimination, this manifests itself rather through slowing down of the decision-making process, as well as through increased salience of own bias. If the emotion regulation is related to discrimination directly, a direct effect on preferences may be triggered, too. Similarly, a theory focused on shaping preferences towards outgroup members operates through *perspective taking* (e.g., Galinsky and Moskowitz, 2000). It asks DMs to put themselves in the shoes of imagined outgroup members. While perspective taking also slows down decision-making process and reduces implicit bias, it may also have a direct effect on social preferences or empathy towards outgroup members, hence directly affecting the penalty parameter d_{min} .

In contrast, *activation of counter-stereotypical associations* through exposure of DMs to counter-stereotypical individuals from a given outgroups may reduce discrimination also via a statistical motives. Imagine providing information about highly successful individual from a group typically associated with low success rates. DMs whose belief distribution was skewed towards the left may now update with a new signal. Holding such incorrect beliefs is possible. Either individuals may not be aware of existence of such types, or their beliefs may be produced by selection neglect (Enke, 2020) or salience putting more weight on low quality of a given group (Bordalo et al., 2016). Drawing DMs’ attention to a neglected part of the quality distribution or highlighting that their belief-generating process may have been incomplete may reduce discrimination.

14.3.3 Social categorization

Group identity is a social construct that is malleable and responds to economic incentives (e.g., Akerlof and Kranton, 2000; Atkin et al., 2021). Social categorization takes this seriously and assumes that it is possible to reclassify outgroups (Gaertner et al., 2000). An intervention building on this model manipulates the DM to either embrace a common identity (e.g., we all are human beings) or to think

of the other person in terms of a different identity against which the DMs is less biased. Since the theory assumes direct manipulation of identity without necessarily providing any new information, it is assumed to operate directly via manipulating the penalty. While before an intervention the penalty was d_1 , manipulating an identity lowers the penalty to d_2 , which is attributed to the group with which the other person is currently associated, including potentially being an ingroup where the penalty is assumed to be $d_{ingroup} = 0$. However, while manipulating how the other is perceived, new information may be provided. If, for example, one’s perceived identity shifts from that person’s race (e.g., “Black”) to their occupation (e.g., “teacher”), distribution of beliefs about a quality of a Black individual in general differs from those of a Black teacher. Under this scenario, statistical models would apply.

14.3.4 Peer influence

We are surrounded by other people, we learn from our peers, and we care about what they think of us. Peer effects may influence discrimination in two distinct ways.¹⁰ First, in line with statistical models, information from peers may be readily available, and its value may be higher than that from other information sources. Consequently, DMs may update their own beliefs by learning from ingroup peers about their attitudes toward and perceptions of outgroups. Second, group-specific social norms may be learned from or enforced by peers. Norm enforcement (or the expectation thereof), or simply understanding the appropriateness of discriminatory behavior, may shape how DMs behave toward outgroup members. This second mechanism is more closely related to taste-based discrimination. The composition of the penalty parameter d_g would need to be extended in line with theories of identity (Akerlof and Kranton, 2000), where individuals gain utility from not deviating from a social norm associated with their identity.¹¹ In both cases—under social learning and social norms—the influence of peers should increase with the social status of the messenger. Information or the threat of social sanctioning from a respected ingroup leader is more influential than that from a random ingroup member.

¹⁰In economics, peer effects have been mostly studied in the domains of education and labor markets. Boucher et al. (2024) propose a general theory of peer effects that can accommodate studying the effects of peers on discrimination.

¹¹Barr et al. (2018) propose a simple theoretical framework. In their model, they introduce an additional component $N(a_i|g, s)$ to the standard taste-based discrimination model. This component captures DMs’ willingness to moderate their own discriminatory behavior, depending on the level of perceived appropriateness a_i of their behavior, conditional on the outgroup’s identity g , and the salience of action appropriateness, s . The effects of the distaste penalty and the moderation parameter enter the utility function in a weighted manner, where the weight is individual-specific. In other words, some people care more about how they are viewed by others, while others care more about their underlying preferences.

14.3.5 Value consistency and self-worth

While we care about how others perceive us, we also place significant importance on our own moral self-image, or *self-worth*. For instance, Bénabou and Tirole (2011a) propose a formal self-signaling model that explores how beliefs about the self are shaped, particularly since individuals often have uncertainty about their own moral type. Interventions that increase the salience of the morality of non-discriminatory actions and one's own moral values (*affirmation* interventions) may encourage decision-makers (DMs) to avoid discriminatory practices. Such an intervention may function both by slowing down decision-making processes—thereby addressing implicit discrimination—and by integrating noisy beliefs about one's own moral type directly into the utility function. One approach to achieving this is by highlighting the DM's own outgroup bias. However, it is important to note that maintaining one's identity may also be achieved by differentiating oneself from the outgroup, potentially even by actively harmful acts.

14.4 Empirical evidence on discrimination reduction

We now turn to empirical studies documenting malleability of discriminatory behavior. I broadly try to classify the evidence to fall under the umbrella of the respective theories of discrimination. But often multiple theories may be at play simultaneously or are at least consistent with the empirical findings. We start from more standard approaches in economics such as manipulating costs of discrimination, and providing or withholding information. Then we turn to evidence from psychology or inspired by its theories, ranging from intergroup contact, changes in attitudes, and manipulation of group identity.

14.4.1 Increasing cost of discrimination

Taste-based discrimination model assumes discriminators face a penalty d_{min} from interacting with an individual from a minority group. That parameter is modelled as fixed (Stigler and Becker, 1977). Increasing the price of discrimination may thus reduce discriminatory behavior, or in other words, there is a "price of prejudice" (Becker, 1971). Formally, if $q_{min} > q_{maj} + d_{min}$ or alternatively if there is a direct cost for not hiring a minority satisfying the condition, such as affirmative action policies, a minority individual should be selected. Standard tests of discrimination such as correspondence studies reduce the cost of discrimination to zero by design by equalizing characteristics of both majority and minority applicants ($q_{min} = q_{maj}$). Thus, they do not allow for testing the implication.

Hedegaard and Tyran (2018) fill this gap in the literature and directly manipulate the price of discrimination by relaxing the characteristic equalization between groups. In a natural field experiment, they invite Danish secondary school students with Danish- and Muslim-sounding names to prepare letters for distribution. Students get paid a piece rate. Students work on the task in two consecutive weeks. In the first week, students work independently and build their productivity histories. In the second week, students are told that they will now work in teams of two. The wage for the second week is determined by the piece rate for the joint team production. The students then get to choose the workday. Together with the day, students are told the first name of their working partner on that day—either Danish- or Muslim-sounding—and the first week productivity of that student. The choice between a workday and a teammate from a specific group thus offers a plausible deniability.¹² The random draw of pairs of teammates to select from offers a distribution of productivity differences. This serves as a manipulation of price of choosing a less productive worker. Prevalence of discrimination is common and students are willing to forego a non-negligible amount to avoid working with Muslim-sounding named students. Consistent with the price-of-prejudice hypothesis, the estimated elasticity of discrimination with respect to price is very high, at -0.9. The authors try to relax a possible confound of (perceived) group-specific production function complementarities using a beliefs survey. However, actual productivity differences may still be better understood.¹³ More recently, in line with price of prejudice, Rao (2019) also shows sensitivity of rich Indian students’ discrimination against poor students to a monetary prize in relay race. His design is similar as it offers rich students a choice between a better performing poor student and a worse performing rich student. The exogenous source of variation is the monetary prize for a winning team. We discuss the paper in detail in Section 14.4.4.

Paths forward. Future research may take more seriously the concern about (beliefs about) complementarities in joint production functions in cross-group interactions. Hjort (2014) shows that productivity of ethnically heterogeneous teams may be lower relative to homogeneous ones. What remains unclear is also how people justify their differential reaction to discrimination under different prices. Do they change the narrative of how they perceive minorities to maintain their self-worth (Bénabou and Tirole, 2011a)? Can this affect their beliefs about outgroups long-term or is the effect rather one off?

¹²Preference for work day is ruled out in a NoName treatment in which only productivity of the student and not the name is revealed. Another treatment in which only name and no productivity is revealed allows for testing statistical motives.

¹³In Bartoš et al. (2024b), we document adverse causal effects of advisor race on advisee productivity.

14.4.2 Beliefs and information provision, processing, and inference

Statistical discrimination models typically assume that decision makers have correct priors about group-specific quality distribution. Decision making can be improved by provision of additional information about individuals. For example, provision of information about White job applicants' higher levels of education increases their chances of receiving an interview invitation (Bertrand and Mullainathan, 2004). Such positive returns to positive individual-level of information has been documented in a range of settings (e.g., List, 2004; Lippens et al., 2022), including in settings when information is not readily available and decision makers have to incur a cost, such as by engaging in online search for private information on social media (Acquisti and Fong, 2020). However, information may often either not be processed, not be processed correctly, or priors may be biased. We now discuss supporting evidence and paths forward on these fronts.

Endogenous information processing. Bartoš et al. (2016) show that information processing may be endogenous. If information processing is costly and if the expected value of an additional signal is unlikely to change the prior belief in discrete decisions such as job recruitment, a decision maker may decide not to process the available signal. Similar reasoning may explain the lack of returns to quality for Black applicants in Bertrand and Mullainathan (2004). Four parameters affect whether information is processed: cost of information processing, the prior belief, the expected signal value, and the expected signal variance. New technologies such as use of artificial intelligence in assisting in hiring decisions may, for example, reduce the cost of information processing. Cross-cultural learning and contact discussed in section 14.4.4 may, beyond its possible direct effects on preferences, also reduce the perceived or actual differences in learning from signals sent by outgroup members (Cornell and Welch, 1996). This may be of particular importance in settings where observing underlying individual quality is difficult.

Incorrect beliefs. A parameter affecting the decision to process additional information is the prior belief about the group quality. Such beliefs are typically assumed to be correct, relying on the theory of rational expectations. Yet a range of literature in psychology and economics documents that people's beliefs are often misperceived (Bursztyn and Yang, 2022), even in domains that can readily be fact-checked (Nyhan, 2020). Such misperceptions are typically larger about outgroup members, relative to ingroup members (Bursztyn and Yang, 2022), possibly due to greater uncertainty about a less known group, or due to motivated beliefs as we discuss below.

Bohren et al. (2023a) present a theoretical framework that allows for inaccurate beliefs and experimentally test the models predictions. An important result of the model is that unless individual beliefs are measured and contrasted to correct beliefs (or rather true distributions), a discriminatory outcome cannot be unambiguously attributed to either taste- or statical-based model. They test the model in a stylized online labor market experiment. Employers recruited workers from a pool of workers differing in characteristics such as nationality and gender to solve logic tasks. Employers were more likely to offer lower wages to Americans and slightly lower to females, as opposed to Indians and males. This is despite the fact that Americans outperform Indians in the task, and women underperform men more substantially. Without data on individual beliefs, the behavior would be attributed to taste-based discrimination against Americans and men. Yet, elicited beliefs reveal that Americans are expected to perform worse than Indians, while women are perceived to underperform less than they do in reality. Still, the process of belief formation may already be influenced by tastes. An exogenous provision of signals about true underlying productivity of a given group should lead to increased precision of underlying beliefs and consequent behavior. Bohren et al. test this prediction by offering actual performance data and show that wages are adjusted in line with belief updating. A taste-based model would not predict such updating.¹⁴ Inaccurate beliefs of individuals may further propagate by spreading to others through social learning or through computer algorithms learning from such beliefs (Kleinberg et al., 2018). Provision of correct statistical information thus seems a promising path forward (Feld et al., 2022; Grigorieff et al., 2020), even if corrected beliefs may not always translate in behavioral change (Haaland and Roth, 2023; Hopkins et al., 2019), especially in the long-run (Bursztyn and Yang, 2022).

Statistical inference and learning. Biased beliefs about a minority group may emerge from incorrect inference from statistical data or from endogenous data generation, even if the underlying information is unbiased and is paid attention to. On inference, consider an example of a police force that responds to a (perceived) higher crime rate in an area with a high share of minority group population that also happens to be more frequently policed. Increased policing results in increased number of crime reported. Neglecting such selection issues may lead to incorrect inference of a causal link between minority group membership and increased propensity to commit crime. This in turn affects police officers’ beliefs and possibly generates a vicious cycle. Hübert and Little (2023) provide evidence

¹⁴Experimenter demand may have played a role as the information was provided within the same study. Haaland et al. (2023) advise to use obfuscated surveys to reduce such confounds, even though De Quidt et al. (2018) suggest that experimenter demand effects are relatively small.

consistent with this. Sample selection neglect is prevalent in other settings (Enke, 2020), and may be exacerbated by environments and social bubbles in which decision makers operate (Levy and Razin, 2019). On a constructive side, Enke (2020) shows that nudging people to pay attention to the sample composition reduces the bias. Levy and Razin (2017) propose that integration policies in schooling systems may break the cycle of belief polarization and segregation.

Lepage (2024) instead focuses on how datasets are created in the first place and studies its effect on exacerbating discrimination. In a dynamic setting, Lepage shows that decision makers learn about group quality distribution from own experience with individual group members. In a hiring decisions, individual qualities of hires inform hiring manager priors about the average group quality. Negative experiences with group members generate negative bias against that group, and consequently affect hiring. This affects subsequent learning—even if perfectly Bayesian and unbiased—about this group and generates persistently negatively skewed belief distribution. These biases are more severe for groups about which hiring managers hold initially more uncertain beliefs, typically minorities, even if the two groups are equally productive on average. A stylized hiring experiment provides evidence consistent with the theory.¹⁵ Incentivizing hiring managers to hire more individuals from the uncertain group reduces the bias. Affirmative action policies may thus have a side effect in terms of improved learning about minority group underlying quality distribution. Similarly, relying on algorithms trained on large datasets may overcome the bias resulting from biased learning from own experience.

Motivated reasoning. Eyting (2022) goes further in understanding possible determinants of inaccurate beliefs. In a stylized hiring online experiment, he shows that beliefs may be driven by own motives and such motivated beliefs may in turn result in discriminatory behavior. In binary hiring decisions, employers seek positive information about potential hires from their preferred group. Specifically, when a previous piece of information was contradicting their motive, they would keep on seeking additional information. Once a piece of information is consistent with their motive, they are more likely to stop searching. Subsequently, employers act on the information received. Bohren et al. (2023a) would classify such behavior as taste-based discrimination, while in fact motivated information processing precedes the final outcome. This paper further blurs the lines between taste and belief based discrimination models by endogenizing belief formation based on own motives. Further, the paper suggests why information provision to recalibrate inaccurate beliefs may fall short of their intended effects. If the provided information is not in line with own motives, individuals may put less

¹⁵Benson and Lepage (2024) show further supportive evidence from administrative records of a US retailer.

weight on such signals¹⁶ and/or press on and search for information more in line with their motives. Such asymmetric learning occurs in situations in which at least some "wiggle room" in interpretation of such signals exists, with clear implications for limiting the space for ambiguous interpretation of information in information provision interventions.

Pluralistic ignorance. Another source of imprecision in beliefs may stem from the lack of individuals' understanding that an observed group behavior may not reflect the true underlying preferences of the plurality of others. Norm conformity, despite being in contrast to individual preferences and beliefs, may thus self-perpetuate, cementing possibly obsolete norms, including discriminatory ones. Katz et al. (1931) coined the term *pluralistic ignorance* to describe this phenomenon. Such misperceptions may lead individuals to acquiesce to norms they internally disagree with but do not stand up to it in public. Bursztyn et al. (2020) document this on a situation of support for female working outside of the household in Saudi Arabia. By collecting private individual data on support for such practice among young married males, they find overwhelming support. Simultaneously, they document substantial underestimation of such support by others. Experimentally providing credible information about the true support, they document subsequent long-lasting behavioral change that leads to increased rates of job applications among their respondents' wives. Similar unraveling has been documented in other settings (Miller, 2023).

Stereotypes. (Bordalo et al., 2016) show how beliefs may be misperceived in their model of stereotypes. The model builds on representativeness heuristic documented by Kahneman and Tversky (1972). The core idea is that frequency of relatively more salient traits of a given group, relative to another group, gets overweighted. Consequently, even relatively small differences in distributions of traits can become exacerbated. This exaggeration in mean differences between groups may be one source of incorrect beliefs about different groups discussed above, even if there is a "kernel of truth" in such belief formation. The paper provides both abstract experimental evidence, as well as supportive empirical evidence using voter attitudes data.

A representative heuristic may distort not only beliefs but also belief updating when processing new information. Esponda et al. (2023) derive a theory of "representative signal distortion". Their

¹⁶Motivated reasoning is also consistent with findings of Campos-Mercade and Mengel (2024) who show that non-Bayesians are more likely to discriminate against members of disadvantaged group than Bayesians. In their experiment it is rather that respondents do not incorporate the information accordingly. However, initial employer motives may be another reason for not updating beliefs in line with Bayesian updating, even when being exposed to information. Delavande and Zafar (2018) show that anti-American attitudes persist as Pakistani respondents do not update enough on new information, also possibly in line with their initial motives.

theory suggests that when evaluating signals from individuals belonging to a certain group, decision makers put more weight to such individual signals representative of that group—relative to a reference group—than they should. In turn, evaluators believe that such representative individual signals are more representative of the entire group than it should be. This may further magnify the effect of biased priors in Bordalo et al. (2016). Experimental evidence they provide is in line with such biased updating process. In the context of statistical discrimination, such biases both increase statistical discrimination by exacerbating differences in initial beliefs but such beliefs also diverge and become less accurate with provision of new evidence, since such information is processed with a positive bias for members of higher quality group, while the opposite is true for members of a lower quality group. Importantly, Esponda et al. provide a remedy. Since Kahneman and Tversky (1972) theory requires a reference group for representativeness heuristic to occur, not allowing for direct comparisons should reduce a bias. In a hiring decision, e.g., this implies evaluating applications sequentially, by group. Alternatively, withholding group information during information processing should eliminate the bias, too. We discuss information withholding in more detail in Section 14.4.3.

Paths forward. While positive information provision often helps, it may not be a panacea. To have a positive effect, information needs to be paid attention to. It also needs to be as unambiguous as possible to minimize the wiggle room allowing for motivated reasoning. Further, direct comparison should be reduced to minimize representativeness heuristic. With the increased knowledge in the domain of how information is processed, we should now turn our attention to how exactly information campaigns should work to durably change attitudes towards minorities and how to design hiring or application procedures give a chance for equal treatment of information provided by everyone. For the former, information provision about hardships may affect preferences directly, perhaps through increasing empathy. We provide some suggestive evidence for such channel in (Bartoš et al., 2024a) but more work is needed to establish the link precisely. Computer algorithms may provide an option for the latter. We now turn to a different approach, namely withholding group attribute, when possible. This reduces both attention discrimination as well as representative signal distortion.

14.4.3 Withholding information

Hiding group attributes may be the simplest way of reducing discrimination. Since group attributes on their own should not have a direct effect on innate quality q , there is a cause for hiding

information about group identity. Even though taste-based model would suggest discrimination persists once the signal is eventually revealed, the attention discrimination model (Bartoš et al., 2016) as well as models building on representativeness heuristic (Esponda et al., 2023) both provide an argument for such information withholding. In the former case, potentially relevant quality information about applicants may not be paid attention to. In the latter case, updating based on the information may be biased. However, information withholding policies need to be treated with caution as unintended consequences and indirect costs need to be taken into account.

Mandated information withholding. In their seminal paper, Goldin and Rouse (2000) show that introduction of blind auditioning in US symphony orchestras led to an increased fraction of women being hired, clearly documenting jurors male-bias in the absence of blind auditioning. The lack of progress in discrimination reduction has led to introduction of legally mandated anonymous application procedures in labor markets broadly. First, the intended effect is to increase the cost of discriminatory behavior (See Section 14.4.1). Second, passing of such rules is intended to send a signal to the society about the non-discriminatory norm (Benabou and Tirole, 2011b). This is referred to as "expressive role of law". Such policies have shown some progress in reducing gender discrimination in job recruitment in Swedish labor markets (Åslund and Skans Nordström, 2012), and blind reviews may be theoretically welfare-increasing (Taylor and Yildirim, 2011).¹⁷ In the era of Internet, however, information not disclosed in the application may often be found using alternative means, such as using social media, if deemed relevant (Acquisti and Fong, 2020).

However, mandated withholding of individual signals may have unintended consequences. When individual information is unavailable, decision makers resort to relying on existing prior beliefs. Agan and Starr (2018) present an example from a policy restricting employers from asking about criminal records of their applicants, labeled "Ban the Box". Using a large-scale correspondence study conducted before and after the law was introduced in New Jersey and in New York in 2015, the paper finds that while reporting criminal records on own application leads to substantially lower callbacks prior to the introduction of the law. However, withholding information about criminal records after the law introduction has a negative effect on callbacks for Blacks, coupled with an increase in callbacks for whites. This asymmetry stems from a simple heuristic that whites are unlikely to have a criminal record, whereas Blacks are. Employers deprived of individual information thus rely on race as an

¹⁷Name-blind resumes have been introduced for hiring in the public sector in countries like Belgium, the Netherlands, and Sweden. Several major global companies have voluntarily introduced the policy in their recruitment.

imperfect proxy for criminal convictions. Similar evidence is found using quasi-experimental evidence exploiting differences in timing of introduction of the law (Doleac and Hansen, 2020). This behavior is consistent with models of statistical discrimination. However, employer beliefs about conviction rates among Blacks are biased upwards. Such exaggeration is either in line with models of inaccurate statistical discrimination (Bohren et al., 2023a) or stereotypes (Bordalo et al., 2016). Raphael (2021) extensively reviews the literature on the unintended effects of the policy, as well as related policies such as bans on employer credit checks (Ballance et al., 2020; Bartik and Nelson, 2024) or drug testing (Wozniak, 2015).

Strategic information withholding and norm conformity. Individuals in some cases do not need to rely on legal restrictions on information provision but may instead strategically manipulate or hide their identity. In a version of a trust game (Introduced originally in Berg et al., 1995) with Georgian high school students, Kudashvili and Lergetporer (2022) show that, in a pre-play signaling stage, Armenian minority trustees are willing to conceal their ethnicity by misreporting their names to sound Georgian. Their behavior is rational, as they accurately anticipate discriminatory behavior from the majority Georgian students. This manipulation effectively eliminates discrimination in this context. The minority behavior is consistent with actual observed strategic name-changes among minorities (Arai and Thoursie, 2009). Relatedly, (Zussman, 2013) document identity hiding by Arabs sellers in an Israeli online market for used cars. They collect a large number of advertisements and show, using a follow-up phone survey, that Arabs are more likely than Jews to leave the name field blank in their advertisements. The response is rational as the same paper documents substantial discrimination against both Arab sellers and buyers using a correspondence study.

In some cases, instead of hiding a trait that is impossible to hide, people from stereotyped groups signal through their behavior conformity with the majority prescribed role to fit in. For example, some Black student engage in costly social signaling through investment in characteristics typically associated with white students, a behavior described as "acting white". This comes at a cost of having fewer same-race friends at school (Fryer and Torelli, 2010). Similarly, female MBA students shy away from signaling preferences for career-enhancing actions when observed by their male peers in order not to hamper their marriage market prospects (Bursztyn et al., 2017)—this is referred to as "acting wife".

Paths forward. Understanding such strategic withholding of information or identity manipulation is important also for interpretation of experimental results on discrimination. These studies typically clearly signal ethnicity and hence do not account for this endogenous response by minority groups. While psychological costs of hiding own identity have been documented (e.g., Camacho et al., 2020), its economic impact is not well understood.¹⁸

14.4.4 Intergroup contact

We now turn to evidence that takes contact hypothesis seriously. Beyond general effects, we discuss the role of facilitating conditions, whether the effects of contact generalize to all members of the outgroup, and whether the effects of contact last beyond the period of studied exposure. Pettigrew and Tropp (2006) and Paluck et al. (2019) offer useful reviews of the state of literature mainly in psychology, but not exclusively. A general summary is that intergroup contact seems to be effective in reducing discrimination. A methodological issue with most of the literature is that it focuses on self-reported outcomes, measured typically after a short period of time. Focusing only on pre-registered studies, Lowe (2024) adds a concern of publication bias. The effect of intergroup contact is still positive for pre-registered studies, but of a considerably smaller magnitude compared to non-registered studies.¹⁹

Few papers stand out for their clever identification and measurement. Intergroup contact can sometimes be studied in a natural setting, going beyond more traditional laboratory based studies. Assignment to college dorms and to army barracks during military training is often determined by a lottery and can serve as a natural experiment. Several studies used such settings to document reduced prejudice (Boisjoly et al., 2006; Carrell et al., 2019; Corno et al., 2022; Finseraas and Kotsadam, 2017; Marmaros and Sacerdote, 2006; Rao, 2019; Scacco and Warren, 2018). The naturalness of the setting is important also for measurement related issues, since it minimizes possible reporting issues. If, for example, the training intervention is directly aimed at improving intergroup relationships, surveyed individuals may feel obliged to report positive attitudes towards outgroups to please the surveyor (De Quidt et al., 2018). Beyond naturalness of the setting, eliciting true underlying revealed preference is of crucial importance, as self-reported measures may be subject to social desirability bias. Paluck et al. (2019) in their meta-study find that estimated effect sizes of intergroup contact are much smaller when comparing behavioral measures to self-reported attitudes. An important addition to the toolbox

¹⁸Even though the link between stress and poverty has been documented (Haushofer and Fehr, 2014).

¹⁹Presenting results of a pre-specified analysis, even if along other exploratory analysis, offers reassurance to the reader that the result was not obtained selectively Olken (2015).

of researchers are measures of implicit attitudes. Corno et al. (2022) use the implicit association test (IAT; Greenwald et al., 1998) to show that being randomly assigned a Black roommate at a South African university reduces white students' implicit attitudes towards Blacks. Barnhardt (2009) finds similar effects in her study examining inter-religious attitudes using a policy of randomly allocating public housing to Muslims and Hindus. She finds that Hindu children exposed to Muslims have more positive implicit attitudes as measured using the IAT. While IAT may be manipulated to some extent (Fiedler and Bluemke, 2005), it is unlikely that this would be the case in either of the two studies. In both cases, it was the first time participants were exposed to the IAT and in neither case it is likely that a possible experimenter demand would be treatment specific.

Several papers stand out for their use of novel behavioral measures. Rao (2019) studies how an Indian policy aimed at eliminating costs of accessing elite private schools for poor students improved rich students' behavior towards poor students in general. The paper exploits two sources of variation from the natural experiment: first, the program was implemented in a staggered way across schools and across classrooms (within the same school), and uses a difference in differences framework to estimate the effects. Further, it exploits a source of exogenous intensity of contact within classroom where students are assigned to peer study groups in an alphabetical order. The findings that rich students reduce their bias against poor children in general are documented on a range of domains. Allocations to anonymous poor students from a different school in a dictator games increase, suggesting a genuine increase in prosocial preferences towards poor students. But interestingly, prosocial behavior seems to be affected in general, also towards rich students. While a dictator game offers a comparable measure of individual preferences, unconfounded by a threat of norms enforcement or peer effects, some may consider it artificial and possibly low stakes. Rao thus turns to field measures. Having poor classmates increases willingness of rich students from treated schools to participate in a weekend fundraiser for disadvantaged kids traditionally organized by the schools themselves, creating a natural setting to study generosity, even if the observable nature of the measure may confound the outcome by possible increased peer pressure. Since economic models of discrimination discuss the importance of social interaction for segregation in labor (Beaman and Magruder, 2012) or housing (Schelling, 1971) markets, Rao measures willingness to interact by organizing a relay race with a cash prize. Prior to the relay race, students had a chance to observe performance of their potential teammates in individual sprints. Rao then introduces a trade-off (discussed in Section 14.4.1): rich students can choose a low-performing rich teammate or a high sprint performance poor teammate. Rich students from treated

classrooms are more likely to pick higher performing poor students.

Role of facilitating conditions. The contact hypothesis, however, depends on the fulfillment of facilitating conditions. When these conditions are not met, contact may be ineffective or even have negative effects.²⁰ Lowe (2021) tests this prediction in a field experiment with cricket players from different castes in India by manipulating both team composition and the composition of leagues in the teams compete. Playing on the same team induces positive, collaborative contact, while competing against another teams induces negative contact with opponents. Consistent with Allport’s original hypothesis, the study finds evidence supporting these predictions. The effect is driven by preference-based channels: First, cooperative contract increases willingness to interact with other caste players, adversarial contact reduces it. Second, the positive effect is experienced even by backup players who only meet their teammates for the duration of the match. Third, the positive effect of cooperative contract operates not merely by winning together with stronger other-caste teammates but rather by working together, suggesting that perceptions about cooperativeness of other caste players is moderated by the type of contact.²¹ To rule out statistical-based channels, both types of contact increase information about quality of players in general. It seems that, at least in this setting, group attribution error a la Allison and Messick (1985), i.e. group-specific attribution of success to luck and ability, may not be at play. The paper also uses behavioral measures instead of self-reports to support its findings and the effects are measured up to three weeks after the leagues ended.

Allport’s hypothesis suggests equal status is another facilitating condition required. Lowe (2021) finds that the effect of collaborative contact is present in both upper and lower caste members, speaking against equal status as a necessary condition for positive contact to occur.

Effects on peers versus general outgroup. Does positive contact reduce discrimination towards the targeted outgroup in general or only to a narrower group of individuals targeted by the interventions? Mousa (2020) answers this question in another experiment manipulating composition of sports team within a sports league, this time of majority Christian and Muslim football players in a Christian enclave in Iraq, following displacement by ISIS. Mousa uses a set of behavioral measures measuring both interactions with outgroup peers (such as training with Muslim players, voting for a Muslim to

²⁰Experiencing negative contact may lead to a higher increase in discrimination when compared to a reduction of discrimination related to positive contact (Barlow et al., 2012). Model of stereotypes offers a theoretical rationale behind this asymmetry (Bordalo et al., 2016).

²¹Economic settings shape how individuals perceive one another. As opposed to an experimentally induced competitive environment, under cooperative conditions, individuals are more likely to describe their counterparts as more similar to themselves (Hagenbach and Kranton, 2023).

receive a best player award, or signing up for a next season tournament into a mixed team), as well as with outgroup strangers (visiting a majority Muslim city, attending a mixed event, or donating to an NGO serving mixed communities). She finds that while playing in the same team with Muslims improves behavior towards Muslim peer football players, it does not affect any measure on the domain of Muslim strangers. A potential confound is that the peer/stranger measures are not directly comparable, so it cannot be concluded with certainty that the difference can be clearly attributable to the difference in effectiveness of contact with outgroup peers and outgroup strangers, even if the consistency of null-results on the three stranger measures is suggestive. Yet under some circumstances, contact manipulation may deliver positive effects even to outgroups strangers. For example, Ghosh et al. (2024) document that willingness to engage with strangers increases in their randomized experiment with Hindu and Muslim youth in India. Lowe (2024) documents in a meta-analysis that the effect of contact with strangers is almost twice as small as the effect with peers. So far, we have little understanding of what conditions are necessary to achieve positive effects beyond the group of outgroup peers and how are identities of individuals in such contact manipulation experiments perceived in the first place.

Intensity of contact and spillovers. Does the intensity of contact affect outcomes? Ghosh et al. (2024) examine this by varying the proportion of Hindu and Muslim youth in an Indian youth camp, mixing either 20% or 50% of Muslims with Hindu participants. Results indicate that a 50% mix—an atypical composition given Hindu dominance in India—leads to worse effects of contact on Hindus, relative to the more moderate 20% mix. This suggests that exposure to an unusually high proportion of a minority group may trigger feelings of outgroup threat.

Long term effects of contact. Intergroup contact effects can persist over extended periods (Billings et al., 2021), sometimes lasting many years (Bursztyn et al., 2024; Ghosh et al., 2024), though this is not always the case (Dahl et al., 2021). It remains unclear why some instances of exposure lead to long-term shifts in attitudes while others do not. These seemingly inconsistent findings may, however, be attributed to variations in outcome measures used across studies.

Paths forward. While research on contact hypothesis has thrived, there are still some open questions. First, we still lack comprehensive evidence on the role of facilitating conditions beyond that of a role of cooperative contact and common goals. Lowe (2021) presents some suggestive evidence

speaking against equal footing being a necessary condition. We lack any study examining the role of support of authority. Second, research so far has been focusing on effects of contact on individuals directly targeted by the intervention. Spillover effects and general norm change onto individuals such as friends, family, or community members of targeted individuals would be an interesting further avenue of research. Documenting such spillovers and conditions for such spillovers to occur would be an important component when evaluating overall effectiveness of contact interventions. Third, while standardizing effect sizes allows for some comparison between different studies, use of harmonized measures such as behavioral games may be a useful complementary tool. For example, adding a dictator game measure to the outgroup peer versus stranger measures in Mousa (2020) would allow an even cleaner evidence of the (lack of) generalized effects of contact onto the entire outgroup, including strangers. Scacco and Warren (2018), Rao (2019), and Corno et al. (2022) are examples that stand out in this respect. Finally, use of administrative data may be another way of dealing with possible experimenter demand in measuring outcomes. Marmaros and Sacerdote (2006) use data on email exchange between university students to measure effects of a random roommate assignment procedure.

14.4.5 Changing attitudes

The previous section presented evidence showing that attitudes toward outgroup members can shift through intensive intergroup contact. If prejudice and discrimination were deeply ingrained, similarly intensive interventions might be required. However, recent evidence suggests that meaningful attitude change may be achieved through less intensive means. We now discuss the roles of perspective taking, exposure to counter-stereotypical examples, exposure to role models, evaluative conditioning, and the potential of narratives. We also consider how attitude changes within the majority group can produce positive spillover effects on the behavior and well-being of minority groups.

Perspective taking. An example is a study by Broockman and Kalla (2016) who study effects of 5 minute long door-to-door canvassing conversations on voting for a non-discrimination law. The underlying theory of change is that eligible voters who exert mental effort in remembering how they were judged negatively and relating such experiences to daily injustices endured by others, would be more sympathetic to transgender individuals. In turn, this would translate in an increased support for a proposed anti-discrimination law. Using a field experiment, the study finds that individuals exposed to perspective taking canvassing exhibit increased support for the non-discrimination law.

Such perspective-taking may increase the salience of counter-prejudice thoughts. This in turn affects behavior. One possibility is that perspective taking reduces implicit biases through increased attention to the topic that requires deliberation, rather than relying on automated stereotypes. This would be in line with implicit discrimination. We revisit this idea in the next paragraph. An alternative explanation for the results enhanced empathy, possibly through reduced social distance or through a shift in perceived similarity.

A recent paper testing the empathy and perceived similarity channels is Andries et al. (2024). They randomly elicit attitudes of museum goers towards unauthorized immigrants to the US before and after they experience a 15 minutes long professional virtual reality piece documenting the plight of migrants crossing the US Southern border directed by a renowned film director. The mere experience of the immersive video increases an index of attitudes towards migrants, including an incentivized donation to a pro-migrant charity measured either immediately after the experience or two months later. The effect is magnified when the virtual reality experience is preceded (not when it comes after the experience) by a provision of neutral statistical information about migration in the US. They consider a theoretical framework in which empathy towards members of an outgroup can be shocked by a salient event such as the immersive virtual reality experience. Such emotional event can have lasting effects on individual preferences, as documented in the two months follow-up. A shock to empathy can be further magnified when individuals feel closer to the outgroup, or putting themselves “in their shoes”. The closeness is manipulated by the provision of statistical facts that is shown to increase perceived closeness using a survey. A follow-up online study with a much less liberal-skewed sample confirms the reduction in social distance following the provision of statistical facts.²²

In Barsbai et al. (2022) we also suggest that emotions and perspective taking may affect intergroup relations. We study a setting with heavily skewed bargaining situation between employers of Filipino domestic workers in Hong Kong and Saudi Arabia and the workers. A simple behavioral intervention that nudges the workers to bring a small, non-monetary gift and a photo of own family when starting their employment improves the treatment they receive from their employers, measured up to two years after the intervention. This includes a revealed preference measure documenting an increased share of contract extensions initiated by the workers themselves. We document the positive effect even among household members staying behind in the Philippines, who remain blind to the treatment. An incentivized online experiment with potential employers suggests that the entire effect is driven by the

²²Others have used immersive virtual reality to study perspective taking, too (Oh et al., 2016; Hamilton-Giachritsis et al., 2018).

photo, suggesting reduced social distance is a driving factor.

The studies above show that emotions may affect cross-group perceptions even in the absence of optimal conditions for intergroup contact, and even in the absence of a direct contact in the first place. While the evidence is encouraging, a better understanding of exact mechanisms behind the lasting effects of emotions may be needed to offer a more general path forward. Theoretical behavioral economics and cognitive psychology offers some ideas in recently developed models of associative memory (Bordalo et al., 2024), and memory retrieval (Kensinger and Ford, 2020). Thinking about the underlying theory of change, it is plausible that emotions directly affect individual preferences through the effect on the penalty d_g .

Habit breaking. Another set of thinking in psychology assumes that discrimination is to some extent driven by habitual, unconscious behavior. Such habits can be broken either by increasing salience of own bias and possibly by learning contingency—or “if-then”—planning strategies in situations in which discrimination may occur. Pre-requisites for habit breaking are thus awareness of own bias, willingness to update own beliefs, and concern about the bias, conditional on awareness (Plant and Devine, 2009). This links the theory with implicit discrimination, statistical, and taste based models.

Alesina et al. (2024) show that randomly exposing Italian teachers to their implicit bias against immigrants, as measured by an IAT, reduces their bias. Awareness of own bias—which was prevalent in their sample—resulted in reduced immigrant-native grading gap, as reflected in administrative school records. This effect is driven mostly by teachers who were surprised by their IAT results, suggesting that belief updating may be the underlying mechanism. Similar effects are documented among basketball referees following a media exposure of an academic study documenting racial bias of referees (Pope et al., 2018). However, the personalized feedback in Alesina et al. (2024) is more powerful among teachers with most negative stereotypes, relative to a debiasing treatment merely highlighting the evidence of bias among teachers in general. Increasing bias awareness thus offers a promising, low cost policy of effective bias reduction. Use of IATs in employee training is already becoming an increasingly popular tool for raising awareness of individual biases in organizations and educational institutions. A better understanding of exact mechanisms through which the behavioral effect operates and whether the effect is merely driven by increasing salience of unconscious bias or whether the attitude change is also affecting individual preferences, for example through perspective taking, needs to be further understood. Finally, understanding belief formation about own implicit bias may call for further inquiry.

Although training requires more time and resources, teaching contingency planning can provide individuals concerned about their own biases with additional tools to overcome them. This, in turn, may enhance the effectiveness of bias awareness interventions. After participants were made aware of their biases through their IAT results, Devine et al. (2012) randomly assigned psychology students to a session discussing potential habit-breaking strategies: recognizing own stereotyping responses and actively replacing them with non-stereotypic responses, engaging in counter-stereotypical imaging by increasing salience of minority role-models, imagining minority group members as individuals with own characteristics, rather than thinking about them in group-specific stereotypical terms, perspective taking, and increased intergroup contact.²³ Those in the treatment group exhibited reduced implicit bias and increased concerns about bias and this reduction remained stable over a two month period. (Forscher et al., 2017) replicate these results in a larger sample and even include a very long term follow-up lasting over two years. There are two interesting features of this study. First, they use an obfuscated survey in the follow-up, relaxing the concerns of experimenter demand. Second, they introduce a clear revealed preference measure by measuring willingness to express public disapproval of a controversial statement related to ethnic stereotypes. Unfortunately, the follow-up sample is relatively small, reducing the statistical power. Training may thus offer additional benefits on top of merely increasing awareness of own bias. It thus remains to be understood which sample should be targeted by training to maximize its effectiveness.

Yet another form of defeating habitual stereotypes may be exposure to counter-stereotypical examples. The stark contrast between a non-stereotypical example and a held stereotype may act as another way of reducing bias, possibly through belief updating about the group quality, perspective-taking, or emotional cues. An example may be observing a male in a stereotypically female occupation or a member of a minority group in a managerial position. While many studies in psychology document positive attitudinal shifts when exposed to counter-stereotypical examples (Dasgupta and Greenwald, 2001; Dasgupta and Asgari, 2004), such attitudinal changes are not strong enough to encourage actual behavioral change in labor market setting. Delfino (2024) shows that male applications for stereotypically female work in social services remain unaffected when a male photograph is shown in actual recruitment messages. However, this study is a first of its kind, focusing on a very particular topic. More such studies are needed on the topic.

²³For references to studies that inspired different training dimensions, see p.p. 1270-1271 in Devine et al. (2012).

Self-fulfilling prophecies and stereotype threat. The total effect of bias reduction in majority groups may also be indirectly complemented by its effect on those who are the targets of bias. Steele and Aronson (1995) and subsequent work on *stereotype threat* (reviewed in Spencer et al., 2016) show that if individuals feel at risk of confirming negative stereotypes about own group in a stereotyped task, their performance and aspirations deteriorates under anxiety and stress. Carlana (2019) documents this in a high stakes setting of Italian schools where girls exposed—quasi-randomly—to gender-biased math teachers perform worse in standardized tests and are less likely to enrol to more demanding high-school tracks compared to girls assigned to non-biased teachers. Consistent with stereotype threat, girls with biased math teachers also exhibit lower self-confidence. Carlana does not find a similar effect for literature, where gender stereotypes are not as strong. Similarly, in a labor market setting, a quasi-random assignment to biased majority group managers increases absenteeism and, conditional on showing up, reduces performance of minority workers measured using administrative data in a French supermarket (Glover et al., 2017). The evidence is consistent with possible unconscious stereotype threat, even though a survey reveals that workers do not explicitly perceive that managers’ have less confidence in them. Spencer et al. (2016) document several studies in psychology that did not find awareness of own experience of stereotype threat. At the same time, biased managers were shown to be less likely to interact with minority workers, leading to reduced monitoring and subsequent worker performance. Stereotype threat adversely affects individual productivity of minorities under biased managers. Thus, minority group would be on average perceived as less productive as long as the firm employs some biased managers. Consequentially, if employer beliefs are well calibrated but do not account for unobserved bias, statistical discrimination would predict setting a higher bar for hiring of minority workers. Glover et al. (2017) show this is the case. Minority workers working under non-biased managers perform on average better, relative to majority workers.

Role models. Counter-stereotypical thinking and habit breaking in general may be triggered by exposure to minority individuals in nontraditional roles or professions. These individuals, who challenge stereotypes in a positive way can also be described as role models. Effects of role models may operate both through changing attitudes in the broader population and by boosting confidence and aspirations of minority groups, possibly counteracting stereotype threat. Beaman et al. (2009) exploit the staggered roll-out of a gender mandate in local councils in India, showing that exposure to female council leaders increases the likelihood of voting for a female leader in subsequent elections. This exposure also directly affects male implicit stereotypes of female leaders, as measured by IATs, and weakens traditional gender

norms. Exposure to role models can reduce uncertainty about the quality of individuals from minority groups, positively update beliefs, directly affect preferences, and change implicit attitudes. In a follow-up study, Beaman et al. (2012) found that the same policy had a direct impact on the aspirations of adolescent girls exposed to female leaders, leading to an increase in their educational attainment. Since labor market opportunities of young adults—a likely comparison group for adolescents forming their expectations about the future—did not change, the effect is unlikely to operate through updated beliefs about labor market prospects. Instead, it appears to operate through increased aspirations.

While aspirations seem malleable, if the gap between current position and the aspiration level grows too high, the individual may either form unrealistic expectations resulting in subsequent frustration or may give up immediately (Dalton et al., 2016; Genicot and Ray, 2017). Understanding the correct targeting that is motivational enough but not yet discouraging is especially difficult when using role models at scale, reaching large and heterogeneous audiences, for example in the media. Yet there is evidence of such successful media applications in changing general perceptions about norms regarding intergroup prejudice (Paluck, 2009), changing fertility norms (La Ferrara et al., 2012), as well as promoting entrepreneurship especially among female entrepreneurs in Tanzania (Bjorvatn et al., 2020). The scalability makes this approach particularly attractive, as a large audience can be reached at relatively low cost. What remains to be understood is how most effectively to tailor such interventions, especially in heterogeneous societies, and how extensive such exposure should be. Serra (2024) reviews the literature on role models for developing countries, but not exclusively.

Evaluative conditioning. Preferences—referred to as affects in psychology—may be modified through evaluative conditioning, a technique extensively studied in psychology (Moran et al., 2024), even in the absence of awareness. This concept suggests that attitudes toward an object can shift through repeated co-occurrence with an affective stimulus, even without explicit association. For instance, frequent negative news coverage that triggers strong emotional responses may, if frequently paired with the portrayal of an outgroup, influence implicit bias toward that outgroup and directly impact the penalty parameter d_{min} . Psychology research explores evaluative conditioning across various domains, including bias reduction (Olson and Fazio, 2006; French et al., 2013). However, these studies rely on small samples recruited among university students and measure only short-term effects (Paluck et al., 2021). Field experiments that employ larger, more representative samples and assess longer-term outcomes have yet to be conducted.

Narratives. Humans use stories and narratives to justify own actions, maintain own self-image, promote or reinforce social norms, define social identities, and possibly to change attitudes of others. Economists have recently started to take note of narratives (Shiller, 2017) and devised models of narratives (Bénabou et al., 2018). A first paper studying the role of narratives in discrimination literature is Barron et al. (2023). They document discriminatory behavior of Jordanian children against Syrian refugee children fleeing the war to Jordan using incentivized sharing games. The key result is that while there is little discrimination on average, there is substantial heterogeneity in the sample. The heterogeneity particularly revolves around own family history and narratives of childrens’ parents about the refugee crisis. If Jordanian children—all born in Jordan themselves and most of their parents, too—come from families with Palestinian roots and thus also share refugee history, they do not discriminate. In contrast, Jordanian children without such family experience discriminate, on average. However, children also learn from narratives of their parents. These narratives then shape the childrens’ preferences: If their parents are themselves supportive of policies helping refugees, their children do not discriminate. The opposite holds true for children of parents who view the refugee crisis rather as a burden. While the evidence is purely correlational and does not directly measure how exactly the intergenerational transmission works, it is suggestive. Since narratives and narrative persuasion may directly affect both beliefs and emotions, and offers ways of justifying a wide range of behaviors, understanding its effectiveness offers a so far under-explored path to discrimination reduction. Further, consistent with models of selective and associative memory, narratives seem to be stickier in peoples’ memories than statistics (Graeber et al., 2024) and hence researchers studying information provision treatments should take note. A first step would be establishment of a true causal effect between narratives and discriminatory behavior, followed by better understanding of underlying mechanisms. Finally, understanding what makes some narratives appealing and ”viral” may be key in understanding how beliefs are formed and transmitted within groups.

14.4.6 Manipulating group identity

People value their identity. This, however, is multifaceted and malleable (Akerlof and Kranton, 2000). Theories of social categorization suggest that discrimination may be reduced by directly manipulating own identity, our beliefs about the identity of others, and by recategorising individuals, including self, into a different identities. While the concept of group identity is central to the entire chapter, I only devote relatively little space to its discussion. The reasons are twofold. First, the role of identity

has been constantly in the background of the evidence discussed so far. For example, the effect of an increased costs of discrimination may be accompanied by a change in our perception of own identity relations in order to justify the differential action at different prices. Identity shapes beliefs that in turn affect stereotyping and exaggerate group divisions (Bonomi et al., 2021). Motivated reasoning offers a way to justify motives that may otherwise seem immoral identity behavior (Bénabou and Tirole, 2011a). Strategic information withholding may have adverse psychological effects through difficult trade-offs minority group individuals have to make between fitting in majority identity to reduce adverse effects of their prejudice, while simultaneously compromising on own identity. Intergroup contact and perspective taking may directly affect intergroup relations, perhaps leading to thinking in terms of shared identities with those we previously considered outgroup members. Role models may challenge the boundaries of identity norms that may be holding aspiration levels down. And narratives are directly linked to identity formation. Second, the literature on economics of identity has been recently extensively reviewed in a series of excellent reviews (Charness and Chen, 2020; Li, 2020; Shayo, 2020). Here, I focus on how exogenous shocks alter identity with implications for intergroup relations, as well as specific ways in which identity recategorization may help reduce discrimination—and cautionary examples of how it can sometimes go terribly wrong.

Exogenous shocks. External events, such as entry of new outgroups, shape majority attitudes towards existing outgroups (Fouka and Tabellini, 2022) and influence relationships among the outgroups themselves (Fouka et al., 2022). The perceived social status of previously homogeneous occupations is also malleable with an entry of an outgroup (Goldin, 2014). Experiences of intergroup conflict increase discrimination (Hjort, 2014) and parochialism (Bauer et al., 2016), which can accelerate post-conflict recovery and strengthen state capacity by promoting ingroup cooperation. However, it also heightens the perception of out-group threat and raises the likelihood of renewed intergroup violence. Additionally, identity choice also responds to monetary incentives. For example, changing prices of consumption of goods associated with own group norms affect identity choice, which in turn affects consumption patterns (Atkin et al., 2021). In other words, if following one identity’s norms become too unattainable, one responds by identity change in order to prevent a utility loss from failing to meet group expectations (Akerlof and Kranton, 2000).

Social categorization. Social categorization suggests that interventions can manipulate identities such that two groups either become closer to one another or start perceiving one another as members of

a larger, over-encompassing identity. For example, experimentally increasing salience of shared family experience of refugee status may improve majority group relations with a recent immigrating refugee minority group (Dinas et al., 2021). Correlational evidence from Barron et al. (2023) is consistent with this.

Identity re-categorization can also be done at scale using media. On a dark side, propaganda could be used to sow group division by simultaneous promotion of ingroup superiority and outgroup dehumanization (Della Vigna et al., 2014; Yanagizawa-Drott, 2014). Yanagizawa-Drott (2014) exploit Rwandan mountainous topography to study the effects of propaganda spread through the—no longer existing—RTL radio program. The radio program operated by Hutu extremists was responsible for most inflammatory anti-Tutsi messaging preceding the 1994 Rwandan genocide, including referring to Tutsi individuals as “cockroaches” and signaling that the government would not prosecute violence against the Tutsi. The study documents that radio signal availability at the time is linked to increased number killing by both militia and civilians. The violence is linked not only to direct effects on villages with signal availability, but also to neighboring villages, further magnifying the effect of hateful propaganda. However, positive reconciliation efforts using media have also been documented. Using a field experiment in Rwanda, Paluck (2009) shows that one-year exposure to a reconciliation soap opera presenting stories of positive intergroup behavior in Rwanda resulted in improved perceptions of intergroup social norms, generalized increased empathy towards other Rwandans, and cooperative behavior. On a larger scale, Blouin and Mukand (2019) study effects of a national reconciliation radio broadcast in Rwanda using a methodology similar to Yanagizawa-Drott (2014). Exposure to the positive propaganda increases willingness to personally interact with members of other ethnic group. The broadcast also reduces salience of own ethnic identity and replaces it with a strengthened national identity.

Paths forward. The examples of identity malleability imply that we should be extremely sensitive to political identity messaging. While people may not exhibit direct individual discriminatory behavior (Berge et al., 2020), it may be relatively easily activated through social contagion (Bauer et al., 2018). Understanding how hateful propaganda can be counteracted in its hot phase when violence can still be prevented is of great importance. Further, more research is needed in understanding which policies trigger which types of group recategorization. For example, is it possible that mixing of Christian and Muslim football players in Mousa (2020) (in Section 14.4.4) increased the salience of common footballer identity, while not necessarily improving attitudes towards the general Muslim population?

14.5 Conclusions

This chapter has reviewed recent literature on discrimination reduction. Although much progress has been made—perhaps in response to calls like that of Bertrand and Duflo (2017)—many questions remain unanswered. While some of these have been addressed throughout the text, I would like to conclude by highlighting a few from a broader perspective and discussing some overarching ideas.

The progress in modeling discrimination offers many ideas on how discrimination may be reduced. First, since preferences are malleable, it is possible to design interventions that reduce intergroup animosity directly. Second, recent research on beliefs and information processing has extended our understanding of how information provision or withholding needs to be carefully implemented to lead to desired outcomes and why some interventions may not work. Finally, theories from psychology can be directly mapped to economics thinking about discrimination, offering novel ideas. While some progress on formalizing economics of identity goes in this direction, there is still scope for formalization of other ideas such as implicit bias, intergroup contact, or direct changes in intergroup attitudes. Understanding the exact underlying mechanisms resulting in behavioral change would help in generalizing the findings from specific studies and reconcile seemingly contradictory results of similar interventions in different settings.

A welcome shift is the expansion of cleverly designed field experiments and quasi-experimental studies, which offer insights from broader populations beyond the narrow university samples often used in laboratory research. These types of studies enhance the reliability of findings, as they are less susceptible to issues like experimenter demand, social desirability bias, learning effects, or Hawthorne effects commonly present in lab settings. Also, these recent studies have introduced innovative, incentivized outcomes that mitigate concerns about misrepresented attitudes in surveys. However, the diversity of outcome measures can complicate comparisons across studies—and sometimes even within a single study. For instance, implicit bias is typically assessed using the IAT, callback rates are used in correspondence studies, and laboratory experiments on intergroup discrimination use standardized money allocation games like the dictator game. Adopting such standardized measures in field experiments and quasi-experimental studies could improve comparability, ultimately allowing for clearer assessments of intervention effectiveness.

To date, empirical and experimental research has largely focused on interventions rooted in a single theory of change often within a specific environment and sample, and using convenient outcome measures. As I argue above, there is still no perfect mapping between results and underlying

theories. However, it also remains unclear whether more concentrated, multifaceted efforts that use multiple tools simultaneously could be more effective—especially for specific sub-groups—compared to less intensive approaches. For instance, such multifaceted approaches have proven effective in poverty reduction (Banerjee et al., 2015). At the same time, existing evidence does not allow for a straightforward comparison of the effectiveness of different methods across similar samples, environments, and outcomes. Designing a mega-study to test various methods within a controlled setting could be a promising solution (Milkman et al., 2021), even though this method has primarily been applied to simpler interventions, such as nudges.

An intriguing area for deeper study is the role of narratives in combination with interventions focused on information provision or attitude change. Given recent interest among economists in the power of narratives and evidence that narratives are often more memorable than statistics (Graeber et al., 2024), drawing inspiration from fields like the arts, marketing, or political messaging could enhance the effectiveness of discrimination reduction interventions. I covered some studies in which narratives are the core of their interventions and all show relatively strong effects (e.g., Yanagizawa-Drott, 2014; Blouin and Mukand, 2019; Andries et al., 2024), even if they do not offer comparison to non-narrative type of messaging.

Two final areas would be worth exploring further. The first is whether the direct effects of discrimination reduction have the predicted impact on systemic discrimination (Bohren et al., 2023b). For instance, the findings of Beaman et al. (2012) may be consistent with this notion: young Indian girls exposed to acceptance of female leaders exhibit an increased willingness to invest in their education, potentially initiating a virtuous cycle of empowerment. The second area of interest pertains to the chapter’s focus on pre-existing discriminatory preferences. Understanding how to prevent these preferences from forming in the first place represents a promising avenue for research. Interventions aimed at this goal would need to target very young children (Fehr et al., 2008).

Although discrimination may never be fully eliminated, the strategies discussed here underscore that meaningful progress is attainable, guiding us toward a world where people are valued for their individual qualities rather than through predefined categories.

Bibliography

- Acquisti, Alessandro and Christina Fong**, “An experiment in hiring discrimination via online social networks,” *Management Science*, 2020, *66* (3), 1005–1024.
- Agan, Amanda and Sonja Starr**, “Ban the box, criminal records, and racial discrimination: A field experiment,” *Quarterly Journal of Economics*, 2018, *133* (1), 191–235.
- Akerlof, George A and Rachel E Kranton**, “Economics and Identity,” *Quarterly Journal of Economics*, 2000, *115* (3), 715–753.
- Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti**, “Revealing Stereotypes: Evidence from Immigrants in Schools,” *American Economic Review*, 2024, *114* (7), 1916–1948.
- Allison, Scott T and David M Messick**, “The Group Attribution Error,” *Journal of Experimental Social Psychology*, 1985, *21*, 563–579.
- Allport, Gordon W**, *The Nature of Prejudice*, Reading, MA: Addison-Wesley, 1954.
- Andries, Marianne, Leonardo Bursztyn, Thomas Chaney, and Milena Djourelouva**, “In Their Shoes,” Technical Report, NBER Working Paper No. 32569 2024.
- Arai, Mahmood and Peter Skogman Thoursie**, “Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings,” Technical Report 1 2009.
- Arrow, Kenneth**, “The theory of discrimination,” in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Vol. 3, NJ: Princeton University Press, 1973.
- Åslund, Olof and Oskar Skans Nordström**, “Do anonymous job application procedures level the playing field?,” *Industrial and Labor Relations Review*, 2012, *65* (1), 82–107.

- Atkin, David, Eve Colson-Sihra, and Moses Shayo**, “How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption,” *Journal of Political Economy*, 2021, 129 (4).
- Ballance, Joshua, Robert Clifford, and Daniel Shoag**, ““No more credit score”: Employer credit check bans and signal substitution,” *Labour Economics*, 4 2020, 63.
- Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Pariente, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry**, “A multifaceted program causes lasting progress for the very poor: Evidence from six countries,” *Science*, 2015, 348 (6236), 1260799–1260799.
- Barlow, Fiona Kate, Stefania Paolini, Anne Pedersen, Matthew J. Hornsey, Helena R.M. Radke, Jake Harwood, Mark Rubin, and Chris G. Sibley**, “The Contact Caveat: Negative Contact Predicts Increased Prejudice More Than Positive Contact Predicts Reduced Prejudice,” *Personality and Social Psychology Bulletin*, 2012, 38 (12), 1629–1643.
- Barnhardt, Sharon**, “Near and Dear? Evaluating the Impact of Neighbor Diversity on Inter-Religious Attitudes,” Technical Report 2009.
- Barr, Abigail, Tom Lane, and Daniele Nosenzo**, “On the social inappropriateness of discrimination,” *Journal of Public Economics*, 8 2018, 164, 153–164.
- Barron, Kai, Heike Harmgart, Steffen Huck, Sebastian O Schneider, and Matthias Sutter**, “Discrimination, narratives, and family history: An experiment with Jordanian host and refugee children,” *The Review of Economics and Statistics*, 2023, 105(4), 1008–1016.
- Barsbai, Toman, Vojtech Bartos, Victoria Licuanan, Andreas Steinmayr, Erwin Tiongson, Dean Yang, Paseo De, and Roxas Legazpi**, “Picture This: Social Distance and the Mistreatment of Migrant Workers,” Technical Report, NBER Working Paper No. 30804 2022.
- Bartik, Alexander W and Scott T Nelson**, “Deleting a Signal: Evidence from Pre-Employment Credit Checks,” *Review of Economics and Statistics*, 2024, *Forthcoming*.
- Bartoš, Vojtěch, Michal Bauer, Jana Cahlíková, and Julie Chytilová**, “Youngism: Discrimination and Stereotypes,” Technical Report, CESifo Working Paper No. 11421 2024.

- Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka**, “Attention discrimination: Theory and field experiments with monitoring information acquisition,” *American Economic Review*, 6 2016, *106* (6), 1437–1475.
- , **Ulrich Glogowsky, and Johannes Rincke**, “Racial Discrimination in Seeking Advice,” Technical Report, Mimeo 2024.
- Bauer, Michal, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts**, “Can war foster cooperation?,” *Journal of Economic Perspectives*, 6 2016, *30* (3), 249–274.
- , **Jana Cahlíková, Julie Chytilová, and Tomáš Želinský**, “Social contagion of ethnic hostility,” *Proceedings of the National Academy of Sciences*, 2018, *115* (19), 4881–4886.
- Beaman, Lori and Jeremy Magruder**, “Who gets the job referral? Evidence from a social networks experiment,” *American Economic Review*, 2012, *102* (7), 3574–3593.
- , **Esther Duflo, Rohini Pande, and Petia Topalova**, “Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India,” *Science*, 2012, *335* (6068), 582–586.
- , **Raghavendra Chattopadhyay, Esther Duflo, and Rohini Pande**, “Powerful women: Does exposure reduce bias?,” *The Quarterly Journal of Economics*, 2009, *124*(4), 1497–1540.
- Becker, Garry S.**, *The Economics of Discrimination*, Chicago: University of Chicago Press, 1957; 2nd edition, 1971.
- Bénabou, Roland and Jean Tirole**, “Identity, morals, and taboos: Beliefs as assets,” *Quarterly Journal of Economics*, 5 2011, *126* (2), 805–855.
- Benabou, Roland and Jean Tirole**, “Laws and Norms,” Technical Report, NBER Working Paper No. 17579 2011.
- Bénabou, Roland, Armin Falk, and Jean Tirole**, “Narratives, imperatives, and moral reasoning,” Technical Report, NBER Working Paper No. 24798 2018.
- Benson, Alan and Louis-Pierre Lepage**, “Learning to Discriminate on the Job,” Technical Report 2024.

- Berg, Joyce, John Dickhaut, and Kevin McCabe**, “Trust, Reciprocity, and Social History,” *Games and Economic Behavior*, 1995, *10* (1), 122–142.
- Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel N. Posner, Bertil Tungodden, and Kelly Zhang**, “Ethnically Biased? Experimental Evidence from Kenya,” *Journal of the European Economic Association*, 2020, *18* (1), 134–164.
- Bernheim, Douglas and Antonio Rangel**, “Addiction and Cue-Triggered Decision Processes,” *The American Economic Review*, 2004, *94* (5), 1558–1590.
- Bertrand, M. and E. Duflo**, “Field Experiments on Discrimination,” in Abhijit Vinayak Banerjee and Esther Duflo, eds., *Handbook of Economic Field Experiments*, 2017, chapter 8, pp. 309–393.
- Bertrand, Marianne and Sendhil Mullainathan**, “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination,” *The American Economic Review*, 2004, *94* (4), 991–1013.
- Billings, Stephen B., Eric Chyn, and Kareem Haggag**, “The Long-Run Effects of School Racial Diversity on Political Identity,” *American Economic Review: Insights*, 2021, *3* (3), 267–284.
- Bjorvatn, Kjetil, Alexander W. Cappelen, Linda Helgesson Sekei, Erik Sørensen, and Bertil Tungodden**, “Teaching through television: Experimental evidence on entrepreneurship education in Tanzania,” *Management Science*, 2020, *66* (6), 2308–2325.
- Blouin, Arthur and Sharun W Mukand**, “Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda,” *Journal of Political Economy*, 2019, *127* (3), 1008–1062.
- Bohren, Aislinn J, Kareem Haggag, Alex Imas, and Devin G Pope**, “Inaccurate Statistical Discrimination: An Identification Problem,” *The Review of Economics and Statistics*, 2023.
- , **Peter Hull, and Alex Imas**, “Systemic Discrimination: Theory and Measurement,” Technical Report, NBER Working Paper No. 29820 2023.
- Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles**, “Empathy or Antipathy? The Impact of Diversity,” *The American Economic Review*, 2006, *96* (5), 1890–1905.
- Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini**, “Identity, Beliefs, and Political Conflict,” *Quarterly Journal of Economics*, 2021, *136* (4), 2371–2411.

- Bordalo, Pedro, Giovanni Burro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer**, “Imagining the Future: Memory, Simulation, and Beliefs,” *Review of Economic Studies*, 2024, *Forthcoming*.
- , **Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer**, “Stereotypes,” *Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794.
- Boucher, Vincent, Michelle Rendall, Philip Ushchev, and Yves Zenou**, “Toward a General Theory of Peer Effects,” *Econometrica*, 2024, *92* (2), 543–565.
- Broockman, David and Joshua Kalla**, “Durably reducing transphobia: A field experiment on door-to-door canvassing,” *Science*, 2016, *352* (6282), 220–224.
- Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott**, “Misperceived social norms: Women working outside the home in Saudi Arabia,” *American Economic Review*, 2020, *110* (10), 2997–3029.
- and **David Y Yang**, “Misperceptions About Others,” *Annual Review of Economics*, 2022, *14*, 425–452.
- , **Thomas Chaney, Tarek A. Hassan, and Aakaash Rao**, “The Immigrant Next Door,” *American Economic Review*, 2024, *114* (2), 248–284.
- , **Thomas Fujiwara, and Amanda Pallais**, “‘Acting wife’: Marriage market incentives and labor market investments,” *American Economic Review*, 2017, *107* (11), 3288–3319.
- Camacho, Gabriel, Mora A. Reinka, and Diane M. Quinn**, “Disclosure and concealment of stigmatized identities,” *Current Opinion in Psychology*, 2020, *31*, 28–32.
- Campos-Mercade, Pol and Friederike Mengel**, “Non-Bayesian Statistical Discrimination,” *Management Science*, 2024, *70* (4), 2549–2567.
- Carlana, Michela**, “Implicit Stereotypes: Evidence from Teachers’ Gender Bias,” *Quarterly Journal of Economics*, 2019, *134* (3), 1163–1224.
- Carrell, Scott E., Mark Hoekstra, and James E. West**, “The impact of college diversity on behavior toward minorities,” *American Economic Journal: Economic Policy*, 2019, *11* (4), 159–182.

- Charles, Kerwin Kofi and Jonathan Guryan**, “Studying discrimination: Fundamental challenges and recent progress,” *Annual Review of Economics*, 2011, *3*, 479–511.
- Charness, Gary and Yan Chen**, “Social Identity, Group Behavior, and Teams,” *Annual Review of Economics*, 2020, *12*, 691–713.
- Chuang, Yating and Laura Schechter**, “Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results,” *Journal of Development Economics*, 2015, *117*, 151–170.
- Coate, Stephen and Glenn C Loury**, “Will Affirmative-Action Policies Eliminate Negative Stereotypes?,” *The American Economic Review*, 1993, *83* (5), 1220–1240.
- Cornell, Bradford and Ivo Welch**, “Culture, Information, and Screening Discrimination,” *Journal of Political Economy*, 1996, *104* (3), 542–571.
- Corno, Lucia, Eliana La Ferrara, and Justine Burns**, “Interaction, Stereotypes, and Performance: Evidence from South Africa,” *The American Economic Review*, 2022, *112* (12), 3848–3875.
- Dahl, Gordon B., Andreas Kotsadam, and Dan Olof Rooth**, “Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams,” *Quarterly Journal of Economics*, 2021, *136* (2), 987–1030.
- Dalton, Patricio S., Sayantan Ghosal, and Anandi Mani**, “Poverty and Aspirations Failure,” *Economic Journal*, 2016, *126* (590), 165–188.
- Dasgupta, Nilanjana and Anthony G Greenwald**, “On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals,” *Journal of Personality and Social Psychology*, 2001, *81* (5), 800–814.
- **and Shaki Asgari**, “Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping,” *Journal of Experimental Social Psychology*, 2004, *40* (5), 642–658.
- Delavande, Adeline and Basit Zafar**, “Information and anti-American attitudes,” *Journal of Economic Behavior and Organization*, 2018, *149*, 1–31.
- Delfino, Alexia**, “Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs,” *American Economic Review*, 2024, *114* (6), 1816–1853.

- Devine, Patricia G., Patrick S. Forscher, Anthony J. Austin, and William T.L. Cox,** “Long-term reduction in implicit race bias: A prejudice habit-breaking intervention,” *Journal of Experimental Social Psychology*, 2012, *48* (6), 1267–1278.
- Dinas, Elias, Vasiliki Fouka, and Alain Schl pfer,** “Family history and attitudes toward out-groups: Evidence from the European refugee crisis,” *Journal of Politics*, 2021, *83* (2), 647–661.
- Doleac, Jennifer L and Benjamin Hansen,** “The Unintended Consequences of ”Ban the Box”: Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden,” *Journal of Labor Economics*, 2020, *38* (2), 321–374.
- Enke, Benjamin,** “What You See is All There is,” *Quarterly Journal of Economics*, 2020, *135* (3), 1363–1398.
- Esponda, Ignacio, Ryan Oprea, and Sevgi Yuksel,** “Seeing What Is Representative,” Technical Report 2023.
- Eyting, Markus,** “Why do we Discriminate? The Role of Motivated Reasoning,” Technical Report, SAFE Working Paper No. 356 2022.
- Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach,** “Egalitarianism in young children,” *Nature*, 2008, *454* (7208), 1079–1083.
- Feld, Jan, Edwin Ip, Andreas Leibbrandt, and Joseph Vecchi,** “Identifying and Overcoming Gender Barriers in Tech: A Field Experiment on Inaccurate Statistical Discrimination,” Technical Report 2022.
- Ferrara, Eliana La, Alberto Chong, and Suzanne Duryea,** “Soap operas and fertility: Evidence from Brazil,” *American Economic Journal: Applied Economics*, 2012, *4* (4), 1–31.
- Fiedler, Klaus and Matthias Bluemke,** “Faking the IAT: Aided and Unaided Response Control on the Implicit Association Tests,” *Basic and Applied Social Psychology*, 2005, *274* (4), 307–316.
- Finseraas, Henning and Andreas Kotsadam,** “Does personal contact with ethnic minorities affect anti-immigrant sentiments? Evidence from a field experiment,” *European Journal of Political Research*, 2017, *56* (3), 703–722.

- Forscher, Patrick S., Calvin K. Lai, Jordan R. Axt, Charles R. Ebersole, Michelle Herman, Patricia G. Devine, and Brian A. Nosek**, “A meta-analysis of procedures to change implicit measures,” *Journal of Personality and Social Psychology*, 2019, *117* (3), 522–559.
- , **Chelsea Mitamura, Emily L. Dix, William T.L. Cox, and Patricia G. Devine**, “Breaking the prejudice habit: Mechanisms, timecourse, and longevity,” *Journal of Experimental Social Psychology*, 2017, *72*, 133–146.
- Fouka, Vasiliki and Marco Tabellini**, “Changing In-Group Boundaries: The Effect of Immigration on Race Relations in the United States,” *American Political Science Review*, 2022, *116* (3), 968–984.
- , **Soumyajit Mazumder, and Marco Tabellini**, “From Immigrants to Americans: Race and Assimilation during the Great Migration,” *Review of Economic Studies*, 2022, *89* (2), 811–842.
- French, Andrea R., Timothy M. Franz, Laura L. Phelan, and Bruce E. Blaine**, “Reducing Muslim/Arab stereotypes through evaluative conditioning,” *Journal of Social Psychology*, 2013, *153* (1), 6–9.
- Fryer, Roland G. and Paul Torelli**, “An empirical analysis of ‘acting white’,” *Journal of Public Economics*, 2010, *94* (5-6), 380–396.
- Fudenberg, Drew and David K. Levine**, “A dual-self model of impulse control,” *The American Economic Review*, 2006, *96* (5), 1449–1476.
- Gaertner, S. L., J. F. Dovidio, B. S. Banker, M. Houlette, K. M. Johnson, and E. A. McGlynn**, “Reducing intergroup conflict: From superordinate goals to decategorization, recategorization, and mutual differentiation,” *Group Dynamics*, 2000, *4* (1), 98–114.
- Galinsky, Adam D. and Gordon B. Moskowitz**, “Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism,” *Journal of Personality and Social Psychology*, 2000, *78* (4), 708–724.
- Genicot, Garance and Debraj Ray**, “Aspirations and Inequality,” *Econometrica*, 2017, *85* (2), 489–519.
- Ghosh, Arkadev, Prerna Kundu, Matt Lowe, and Gareth Nellis**, “Creating Cohesive Communities: A Youth Camp Experiment in India,” Technical Report 2024.

- Glover, Dylan, Amanda Pallais, and William Pariente**, “Discrimination as a self-fulfilling prophecy: Evidence from French grocery store,” *Quarterly Journal of Economics*, 2017, *132* (3), 1219–1260.
- Goldin, Claudia**, “A Pollution Theory of Discrimination: Male and Female Differences in Occupations and Earnings,” in Leah Platt Boustan, Carola Frydman, and Robert A. Margo, eds., *Human Capital in History: The American Record*, University of Chicago Press, 2014, chapter 9.
- **and Cecilia Rouse**, “Orchestrating Impartiality: The Impact of ”Blind” Auditions on Female Musicians,” *The American Economic Review*, 2000, *90* (4), 715–741.
- Graeber, Thomas, Christopher Roth, and Florian Zimmermann**, “Stories, Statistics, and Memory,” *The Quarterly Journal of Economics*, 2024, *Forthcoming*, 2181–2225.
- Greenwald, Anthony G, Debbie E McGhee, and Jordan L K Schwartz**, “Measuring Individual Differences in Implicit Cognition: The Implicit Association Test,” *Journal of Personality and Social Psychology*, 1998, *74* (6), 1464–1480.
- Grigorieff, Alexis, Christopher Roth, and Diego Ubfal**, “Does Information Change Attitudes Toward Immigrants?,” *Demography*, 2020, *57* (3), 1117–1143.
- Haaland, Ingar and Christopher Roth**, “Beliefs about Racial Discrimination and Support for Pro-Black Policies,” *Review of Economics and Statistics*, 2023, *105* (1), 40–53.
- , – , **and Johannes Wohlfart**, “Designing Information Provision Experiments,” *Journal of Economic Literature*, 2023, *61* (1), 3–40.
- Hagenbach, Jeanne and Rachel Kranton**, “Competition, Cooperation, and Motivated Social Perceptions,” Technical Report 2023.
- Hamilton-Giachritsis, Catherine, Domna Banakou, Manuela Garcia Quiroga, Christos Giachritsis, and Mel Slater**, “Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment,” *Scientific Reports*, 2018, *8* (2975), 1–10.
- Haushofer, Johannes and Ernst Fehr**, “On the psychology of poverty,” *Science*, 2014, *344* (6186), 862–7.
- Hedegaard, Morten Størling and Jean Robert Tyran**, “The price of prejudice,” *American Economic Journal: Applied Economics*, 2018, *10* (1), 40–63.

- Hjort, Jonas**, “Ethnic divisions and production in firms,” *Quarterly Journal of Economics*, 2014, *129* (4), 1899–1946.
- Hopkins, Daniel J., John Sides, and Jack Citrin**, “The muted consequences of correct information about immigration,” *Journal of Politics*, 2019, *81* (1), 315–320.
- Hübert, Ryan and Andrew T Little**, “A behavioural theory of discrimination in policing,” *The Economic Journal*, 2023, *133*, 2828–2843.
- Kahneman, Daniel and Amos Tversky**, “Subjective probability: A judgment of representativeness,” *Cognitive psychology*, 1972, *3* (3), 430–454.
- Katz, Daniel, Floyd Henry Allport, and Margaret Babcock Jenness**, “Students’ attitudes; a report of the Syracuse University reaction study,” Technical Report 1931.
- Kensinger, Elizabeth A and Jaclyn H Ford**, “Retrieval of Emotional Events from Memory,” *Annual Review of Psychology*, 2020, *71*, 251–272.
- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R Sunstein**, “Discrimination in the Age of Algorithms,” *Journal of Legal Analysis*, 2018, *10*, 113–174.
- Kudashvili, Nikoloz and Philipp Lergetporer**, “Minorities’ strategic response to discrimination: Experimental evidence,” *Journal of Public Economics*, 2022, *208*.
- Lane, Tom**, “Discrimination in the laboratory: A meta-analysis of economics experiments,” *European Economic Review*, 2016, *90*, 375–402.
- Lang, Kevin and Ariella Kahn-Lang Spitzer**, “Race Discrimination: An Economic Perspective,” *The Journal of Economic Perspectives*, 2020, *34* (2), 68–89.
- **and Jee Yeon K. Lehmann**, “Racial discrimination in the labor market: Theory and empirics,” *Journal of Economic Literature*, 2012, *50* (4), 959–1006.
- Lepage, Louis-Pierre**, “Experience-Based Discrimination,” *American Economic Journal: Applied Economics*, 2024, *16* (4).
- Levy, Gilat and Ronny Razin**, “The coevolution of segregation, polarized beliefs, and discrimination: The case of private versus state education,” *American Economic Journal: Microeconomics*, 2017, *9* (4), 141–170.

- and —, “Echo Chambers and Their Effects on Economic and Political Outcomes,” *Annual Review of Economics*, 2019, 11, 303–328.
- Li, Sherry Xin**, “Group Identity, Ingroup Favoritism, and Discrimination,” in Klaus F. Zimmermann, ed., *Handbook of Labor, Human Resources and Population Economics*, Springer International Publishing, 2020, pp. 1–28.
- Lippens, Louis, Siel Vermeiren, and Stijn Baert**, “The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments,” *European Economic Review*, 2023, 151, 104315.
- , **Stijn Baert, Abel Ghekiere, Pieter Paul Verhaeghe, and Eva Deros**, “Is labour market discrimination against ethnic minorities better explained by taste or statistics? A systematic review of the empirical evidence,” *Journal of Ethnic and Migration Studies*, 2022, 48 (17), 4243–4276.
- List, John A**, “The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field,” *The Quarterly Journal of Economics*, 2004, 119 (1), 49–89.
- Lowe, Matt**, “Types of contact: A field experiment on collaborative and adversarial caste integration,” *American Economic Review*, 2021, 111 (6), 1807–1844.
- , “Has Intergroup Contact Delivered?,” Technical Report 2024.
- Lundberg, Shelly J and Richard Startz**, “Private Discrimination and Social Intervention in Competitive Labor Market,” *The American Economic Review*, 1983, 73 (3), 340–347.
- Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt**, “Rational Inattention: A Review,” *Journal of Economic Literature*, 2023, 61 (1), 226–273.
- Marmaros, David and Bruce Sacerdote**, “How do friendships form?,” *The Quarterly Journal of Economics*, 2006, 121 (1), 79–119.
- Milkman, Katherine L., Dena Gromet, Hung Ho, Joseph S. Kay, Timothy W. Lee, Pepi Pandiloski, Yeji Park, Aneesh Rai, Max Bazerman, John Beshears, Lauri Bonacorsi, Colin Camerer, Edward Chang, Gretchen Chapman, Robert Cialdini, Hengchen Dai, Lauren Eskreis-Winkler, Ayelet Fishbach, James J. Gross, Samantha Horn, Alexa Hubbard, Steven J. Jones, Dean Karlan, Tim Kautz, Erika Kirgios, Joowon Klusowski, Ariella Kristal, Rahul Ladhanian, George Loewenstein, Jens Ludwig, Barbara Mellers,**

- Sendhil Mullainathan, Silvia Saccardo, Jann Spiess, Gaurav Suri, Joachim H. Talloen, Jamie Taxer, Yaacov Trope, Lyle Ungar, Kevin G. Volpp, Ashley Whillans, Jonathan Zinman, and Angela L. Duckworth**, “Megastudies improve the impact of applied behavioural science,” *Nature*, 2021, *600* (7889), 478–483.
- Miller, Dale T.**, “A century of pluralistic ignorance: what we have learned about its origins, forms, and consequences,” *Frontiers in Social Psychology*, 2023, *1*.
- Moran, Tal, Yahel Nudler, and Yoav Bar-Anan**, “Evaluative Conditioning: Past, Present, and Future,” *Annual Review of Psychology*, 2024, *74*, 245–269.
- Mousa, Salma**, “Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq,” *Science*, 2020, *369*, 866–870.
- Neumark, David**, “Experimental research on labor market discrimination,” *Journal of Economic Literature*, 2018, *56* (3), 799–866.
- Nyhan, Brendan**, “Facts and myths about misperceptions,” *Journal of Economic Perspectives*, 2020, *34* (3), 220–236.
- Oh, Soo Youn, Jeremy Bailenson, Erika Weisz, and Jamil Zaki**, “Virtually old: Embodied perspective taking and the reduction of ageism under threat,” *Computers in Human Behavior*, 2016, *60*, 398–410.
- Olken, Benjamin A.**, “Promises and perils of pre-analysis plans,” *Journal of Economic Perspectives*, 2015, *29* (3), 61–80.
- Olson, Michael A. and Russell H. Fazio**, “Reducing automatically activated racial prejudice through implicit evaluative conditioning,” *Personality and Social Psychology Bulletin*, 2006, *32* (4), 421–433.
- Onuchic, Paula**, “Recent Contributions to Theories of Discrimination,” Technical Report 2024.
- Paluck, Elizabeth Levy**, “Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda,” *Journal of Personality and Social Psychology*, 2009, *96* (3), 574–587.
- , **Roni Porat, Chelsey S Clark, and Donald P Green**, “Prejudice Reduction: Progress and Challenges,” *Annual Review of Psychology*, 2021, *72*, 544–560.

- , **Seth A. Green, and Donald P. Green**, “The contact hypothesis re-evaluated,” *Behavioural Public Policy*, 2019, *3* (2), 129–158.
- Pettigrew, Thomas F and Linda R Tropp**, “A Meta-Analytic Test of Intergroup Contact Theory,” *Journal of Personality and Social Psychology*, 2006, *90* (5), 751–783.
- Phelps, Edmund S**, “The Statistical Theory of Racism and Sexism,” *American Economic Review*, 1972, *62* (4), 659–661.
- Plant, E. Ashby and Patricia G. Devine**, “The Active Control of Prejudice: Unpacking the Intentions Guiding Control Efforts,” *Journal of Personality and Social Psychology*, 2009, *96* (3), 640–652.
- Pope, Devin G., Joseph Price, and Justin Wolfers**, “Awareness reduces racial bias,” *Management Science*, 2018, *64* (11), 4988–4995.
- Quidt, Jonathan De, Johannes Haushofer, and Christopher Roth**, “Measuring and bounding experimenter demand,” *American Economic Review*, 2018, *108* (11), 3266–3302.
- Rao, Gautam**, “Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools,” *The American Economic Review*, 2019, *109* (3), 774–809.
- Raphael, Steven**, “The Intended and Unintended Consequences of Ban the Box,” *Annual Review of Criminology*, 2021, *4*, 191–207.
- Scacco, Alexandra and Shana S. Warren**, “Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria,” *American Political Science Review*, 2018, *112* (3), 654–677.
- Schelling, Thomas C**, “Dynamic models of segregation,” *Journal of Mathematical Sociology*, 1971, *1*, 143–186.
- Schildberg-Hörisch, Hannah**, “Are risk preferences stable?,” *Journal of Economic Perspectives*, 2018, *32* (2), 135–154.
- Serra, Danila**, “Role Models in Developing Countries,” in Utteeyo Dasgupta and Pushkar Maitra, eds., *Handbook of Experimental Development Economics*, Edward Elgar, 2024.

- Shayo, Moses**, “Social Identity and Economic Policy,” *Annual Review of Economics*, 2020, pp. 355–389.
- Shiller, Robert J.**, “Narrative economics,” *American Economic Review*, 2017, 107 (4), 967–1004.
- Spencer, Steven J., Christine Logel, and Paul G. Davies**, “Stereotype threat,” *Annual Review of Psychology*, 2016, 67, 415–437.
- Steele, Claude M. and Joshua Aronson**, “Stereotype threat and the intellectual test performance of African Americans,” *Journal of Personality and Social Psychology*, 1995, 69 (5), 797–811.
- Stigler, George J and Gary S Becker**, “De Gustibus Non Est Disputandum,” *The American Economic Review*, 1977, 67 (2), 76–90.
- Taylor, Curtis R. and Huseyin Yildirim**, “Subjective performance and the value of blind evaluation,” *Review of Economic Studies*, 2011, 78 (2), 762–794.
- Vigna, Stefano Della, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya**, “Cross-border media and nationalism: Evidence from Serbian radio in Croatia,” *American Economic Journal: Applied Economics*, 2014, 6 (3), 103–132.
- Wozniak, Abigail**, “Discrimination and the effects of drug testing on black employment,” *Review of Economics and Statistics*, 2015, 97 (3), 548–566.
- Yanagizawa-Drott, David**, “Propaganda and conflict: Evidence from the Rwandan genocide,” *Quarterly Journal of Economics*, 2014, 129 (4), 1947–1994.
- Zussman, Asaf**, “Ethnic discrimination: Lessons from the Israeli online market for used cars,” *Economic Journal*, 2013, 123 (572), F433–F468.