

The Whys of Social Exclusion

Insights from Behavioral Economics

Karla Hoff

James Walsh



WORLD BANK GROUP

Development Research Group

Macroeconomics and Growth Team

December 2017

Abstract

All over the world, people are prevented from participating fully in society through mechanisms that go beyond the structural and institutional barriers identified by rational choice theory (poverty, exclusion by law or force, taste-based and statistical discrimination, and externalities from social networks). This essay discusses four additional mechanisms that bounded rationality can explain: (i) implicit discrimination, (ii) self-stereotyping and self-censorship, (iii) “fast thinking” adapted to underclass neighborhoods, and (iv) “adaptive preferences” in which an oppressed group views its oppression as natural or even preferred.

Stable institutions have cognitive foundations—concepts, categories, social identities, and worldviews—that function like lenses through which individuals see themselves and the world. Abolishing or reforming a discriminatory institution may have little effect on these lenses. Groups previously discriminated against by law or policy may remain excluded through habits of the mind. Behavioral economics recognizes forces of social exclusion left out of rational choice theory, and identifies ways to overcome them. Some interventions have had very consequential impact.

This paper is a product of the Macroeconomics and Growth Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at khoff@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

The Whys of Social Exclusion: Insights from Behavioral Economics

Karla Hoff (The World Bank)

James Walsh (University of Oxford and The World Bank)*

Keywords: behavioral economics, culture, implicit discrimination, stereotype, psychological and social influences on decision-making

JEL codes: O12, Z13

*The authors would like to thank Adrienne Pon for comments.

Table of Contents

The “schematizing power of institutions”

Barrier 1: Implicit discrimination

Interventions to reduce discrimination

Barrier 2: Self-stereotyping and self-censorship

Interventions to weaken the influence of negative stereotypes on self-assessment

Barrier 3: The difficulty of adapting fast thinking to two worlds

Rules of thumb and self-regulation

Barrier 4: “Adaptive preferences” of the oppressed

Role models

Conclusion

All over the world, people are prevented from participating fully in society for reasons that go beyond the structural and institutional barriers that are the focus of rational choice theory— poverty, exclusion by law or force, taste-based or statistical discrimination, and externalities from an individual’s position in social networks.¹ As the precision of economics has increased through field, lab, and as-if-random “natural” experiments, researchers have uncovered socio-psychological barriers to upward mobility. In India, low-caste boys solved mazes just as well as high-caste boys when their caste was not publicly revealed, but solved 23 percent fewer mazes than high-caste boys when caste identity was revealed in mixed-caste groups (Hoff and Pandey, 2014). In France, grocery store clerks of African origin were 9 percent *more* productive than other clerks *except* on days when they were supervised by managers implicitly biased against minorities; on these days, they were of average productivity (Glover, Pallais, and Pariente, 2017). A belief that a race, gender, caste, or other ascriptive group is inferior can affect how others treat members of the group and how members of the group feel about themselves, creating productivity differences that sustain the beliefs, although no inherent productivity differences exist (Hoff and Stiglitz, 2010; World Bank, 2015).

In the last 30 years, economics has taken a cognitive turn. Economists and psychologists have made break-throughs in understanding how people make

¹ The first three barriers are obvious, but the fourth may not be. An individual in a low-income class may rationally choose not to seek upward mobility—for example, not to work hard in school and not to delay child-bearing beyond adolescence—if doing so would create too great a distance from peers, connections to whom make a person happy (Akerlof, 1997) or if efforts to seek upward mobility (e.g., taking advanced courses in school or buying a computer) have a high return only if others in one’s social network make them, too (DiMaggio and Garip, 2012). Peer group effects create a coordination problem that can explain the stability of class structure.

decisions, and a new field has emerged—behavioral economics. Camerer (2005) and Hoff and Stiglitz (2016) distinguish two strands. With insights from psychology, the first strand views the individual as a *quasi-rational actor*: he thinks clearly under ideal conditions, but under most real-world conditions his judgment and behavior are affected by seemingly irrelevant contextual factors in the moment of decision (Thaler, 2016). With insights from social psychology, sociology, and anthropology, the second strand of behavioral economics views the individual as a *quasi-rational, enculturated actor*: experience and exposure to social patterns have persistent effects on judgment and behavior by shaping the cognitive toolkit with which information is processed. The toolkit includes categories, concepts, causal narratives, and other *mental models* (or, equivalently, *schemas*) (Douglas, 1966, 1986; Bruner, 1991; D’Andrade, 1995; Strauss and Quinn, 1997; Bicchieri, 2006). An individual may have multiple and inconsistent mental models to interpret a situation he is in, and cultural cues in the environment will guide the selection of the mental model (DiMaggio, 1997, 275). There is a two-way relationship between individuals and institutions. Individuals create institutions, but institutions shape the mental models of individuals and what primes certain behaviors: “In an ongoing cycle of mutual constitution, people are socioculturally shaped shapers of their environment” (Markus and Kitayama, 2010, 441).

The purpose of this synthesis essay is to provide a perspective from behavioral economics on the forces that maintain social exclusion and on interventions to overcome them. The essay is divided into five sections. We first give evidence that institutions influence how people think. Then we discuss four

mechanisms through which an institution that excludes an ascriptive group may persist long after the institution is reformed or abolished. The mechanisms are (i) implicit discrimination, (ii) self-stereotyping and self-censorship, (iii) “fast thinking” of individuals who live in two worlds—a violent world of disadvantaged groups and an orderly world of school or work, and (iv) “adaptive preferences,” in which the excluded come to view the exclusion as natural. We discuss interventions to overcome each mechanism, in some cases with hugely consequential impact.

The “schematizing power of institutions”²

A common definition of institutions is “rules of the game.” In traditional economics, the rules (in particular, property rights) affect only the opportunity sets over which people optimize; institutions have no impact on the way individuals think. In contrast, all other social sciences recognize that institutions have a “schematizing power”: they shape the knowledge structures (mental models or schemas) that an individual uses to process information—what he attends to, what he perceives, and how he interprets ambiguous signals. Bruner (1990, 11, 57) argues that the schematizing power of institutions creates and sustains social identities: The symbolic systems make the user a reflection of the community by influencing how others see him and how he sees himself.

² Bruner (1990, 58).

Social groups that differ in their experiences and exposure have different mental models and behave in systematically different ways in the *same* situation (*e.g.* Henrich and others, 2001; Beaman and others, 2012; Brooks, Hoff, and Pandey, 2017). Many scholars view shared mental models as a primary manifestation of culture. (See Douglas, 1986, especially pp. 46-48; Swidler, 1986; and the definition in DiMaggio, 1997, p. 281 of “culture as a network of interrelated schemata.”) Culture has a constitutive role, not merely a regulatory role. Economists and political scientists increasingly incorporate mental models or rules of thumb (based on a schematic view of a situation) as a variable to explain change or persistence of inequality; *e.g.*, Hoff, Fehr, and Kshetramade, 2011 (impact of caste identity on altruistic punishment in India to protect an ingroup); Alesina, Giuliano, and Nunn, 2013 (impact of plow cultivation in pre-industrial agriculture on modern gender roles); Acharya, Blackwell, and Sen, 2015 (impact of the level of historical dependence on slave labor on contemporary racism in the U.S.); Bedolla and Miachelson, 2012 and Carpenter and Foos, 2017 (impact of “learned disengagement” of marginalized U.S. citizens on the response to get-out-the-vote activities).

Categories, one kind of mental model, lay the foundation for social stratification. The psychologist Gordon Allport (1950, 20) argued that “[t]he human mind must think with the aid of categories. Once formed, categories are the basis for normal prejudgment. We cannot possibly avoid this process. Orderly living depends on it.” Institutions that create hierarchies of ascriptive groups (*e.g.*, by race, gender, or ethnicity) impair the ability of others to learn things about the person that do not fit the category, since mental models *filter* information in a way that tends to preserve categorical beliefs.

Psychologists are beginning to understand the neural basis of categorization and associative learning:

When neurons are consistently activated by co-occurring features of experience, physical changes in the neurons strengthen the connections between and among them... Thereafter, if one of those neurons is activated, it will be more likely to activate another in that group ... Growing up in an environment of a given cultured shape brings with it a distinctive pattern of experiences and corresponding neural changes.... *The synaptic changes... cannot be erased like sentences from a text...* Change in the world can lead to a new pattern of strong neural connections, but it does not completely destroy earlier learning. (Strauss and Quinn, 1997, 90; emphasis added)

A simple experiment by Bruner and Potter (1964) in visual identification gives suggestive evidence that categorization leads individuals to resist disconfirming evidence. Subjects were shown eight photographs of ordinary objects (*e.g.*, a dog standing on grass), projected one at a time, and gradually coming into better focus. One group of subjects saw the photos starting from almost complete blur. Another group of subjects saw the photos starting from medium blur. And a third group saw the photos starting from light blur. For all groups, the picture being exposed was stopped at the same point of focus, regardless of its starting point. At this common terminal point, the subject was asked what the object was. The surprising finding was that subjects who had seen the longer video, starting at greater blur, were less likely to identify the object correctly. Despite more exposure, they learned less. Slightly less than one-fourth of the subjects recognized pictures when they began their viewing with a very blurred image, but more than half recognized them when viewing began with light blur. Bruner and Potter suggest that people have difficulty rejecting mental representations. Individuals “hang on” to false hypotheses: “at any particular clarity of the display, those who see it for the

first time are more likely to recognize the object than those who started viewing at a less clear stage” (p. 425). This is called an *interference phenomenon*.

Kahan and others (2017) demonstrate a related finding in the political domain. They find that when people process scientific data that conflicts with their ingrained worldviews (*e.g.*, that gun control increases crime), they often misinterpret the data. The interference problem is more severe, the more numerate they are. This indicates that the problem is not inadequate mathematical skill. The authors suggest that mathematical skill may actually enhance the ability to filter out unwanted information.

People may be capable of suppressing their biases in clear-cut cases, but incapable of doing so in situations of ambiguity. The psychologists John Darley and Paget Gross (1983) investigated how the social class of a student influences others’ judgments of how well she is doing in school. The experimental subjects were randomly allocated to one of four groups. Group 1 saw a video of a nine-year old girl, called Hannah, in a low-income neighborhood and were informed that her parents had only a high school education. Group 2 saw a video of Hannah in a high-income neighborhood and were informed that her parents were college-educated. Groups 3 and 4 had the same information as groups 1 and 2, respectively, but in addition viewed a videotape depicting Hannah taking an oral test. There was only one version of the videotape. It depicted Hannah’s performance as inconsistent—she answered some challenging questions correctly and some easy questions incorrectly.

Traditional economics would predict that additional information could only increase the precision of participants’ assessments and narrow the gap in assessments between them, but the opposite was true. The first two groups, who had information only

about Hannah's socioeconomic background, differed very little in their assessment of how well she was doing in school. In contrast, groups 3 and 4 differed significantly in their assessments. The 'rich' Hannah was judged to be more able than the 'poor' Hannah. Expecting her to do better, viewers of the 'rich' Hannah compared to viewers of the 'poor' Hannah evaluated her performance in the oral test more favorably. The results support the hypothesis that mental models play a role in information-processing that is distinct from their role in pre-judgment.

Alesina, Guiliano, and Nunn (2013) provides an example of the persistent effect of the meanings that an historical institution gave to gender. Pre-industrial agriculture used either shifting cultivation or plow cultivation. Unlike the hoe or digging stick used to prepare the soil in shifting cultivation, the plow requires significant strength—either to pull it or to control the animal that pulls it. In areas topographically well-suited to crops for which plow agriculture is efficient (wheat, barley, and rye), men had an advantage in farming relative to women, and adoption of the plow created gendered occupations—*men in the field, women at home*—that have influence in modern times. In such areas, female labor force participation in the year 2000 was more than 20 percentage points lower than in other areas. This influence remains as individuals migrate: the gender norms of immigrants' children who live in the U.S. and Europe are influenced by whether their ancestors were members of an ethnic group that used plow cultivation in the pre-industrial period.

Mental models that an institution of social exclusion creates can be deliberately strengthened after the institution is abolished in order to make the old social pattern persist. In U.S. southern counties that in 1860 had a high proportion of slaves, whites are

more likely today to express racial resentment toward African Americans and to oppose affirmative action, compared to whites who live in otherwise similar areas that had lower population shares of slaves (Acharya, Blackwell, and Sen, 2015). In order to hold down agricultural labor costs after slavery was abolished (as well as for social and political reasons), whites in counties that had relied heavily on slave labor reinforced racist norms and racial hostility. This shaped attitudes that were transmitted across generations through culture and institutions, such as Jim Crow. Anti-black attitudes faded earlier in areas with a low historical dependence on slave labor.

Drawing on many social sciences, 21st century behavioral economics (which we have called Strand 2) introduces into economic theory a new variable for processing information—*mental models*. The new variable is shaped by institutions through experience and exposure. It operates through four mechanisms of social exclusion—four social barriers—discussed in the remainder of the paper.

Barrier 1: Implicit discrimination

Rational choice theory offers two explanations for discrimination: taste-based and statistical. *Taste-based discrimination* arises when individuals dislike members of a group. Becker (1957) argued that employers who discriminate based on social identity knowingly incur higher costs. Taste-based discrimination would therefore not survive in perfectly competitive markets.

Statistical discrimination arises when individuals have imperfect information about an individual and assess his expected productivity based on his membership in a

group (*e.g.*, a race, social class, or gender) (Arrow, 1973). Employers may refuse to hire a member of a group because of a rational determination that the person has lower expected productivity than job applicants who are members of another group.

A third kind of discrimination, left out of traditional economics, is *implicit* (*unconscious*) *discrimination* (Banaji and Greenwald, 1995; Greenwald and Krieger, 2006). Implicit discrimination differs from explicit discrimination in many ways: its sources, malleability, and effect on behavior. Explicit and implicit discrimination do not emerge from the same socialization process (Dovidio and others, 1997). Explicit discrimination is much easier than implicit discrimination to change (Wilson and others, 2000). It is consistent with this that self-reports of explicit racism show a large decline among whites, whereas racially discriminatory behaviors remain common among them (Dovidio and Gaertner, 2004). Implicit bias predicts important life outcomes. Nosek and others (2009) find that cross-country variation in implicit attitudes against women in science predict gender-based achievement gaps in eighth-grade science and math. As we mentioned above and discuss more below, implicit bias by supervisors against minority staff can directly impair their performance.

Beaman and others (2009) investigated how having women in leadership positions affected the attitudes and behavior of constituents. The investigators studied the impact of a policy in India in which the government randomly reserved—in one-third of the villages—the position of village council leader for women candidates. The study revealed that the impact was very consequential. The investigators used the Goldberg paradigm, which is a common way to measure bias. They asked subjects in one group to evaluate a taped speech by a man. They asked subjects in another group to evaluate the

identical speech made by a woman. In villages that had never had political reservations for women, both male and female respondents gave the male politician *higher* ratings than the woman for effectiveness. But in villages with political reservations for women for the past seven years, men evaluated the woman's speech as *just as effective* as the man's. Discrimination by the measure of the Goldberg paradigm had been removed.

Goldin and Rouse (2000) find evidence of gender bias in hiring for symphony orchestras in the U.S. Before 1980, none of the five highest-ranked U.S. orchestras had more than 12 percent women. Through the 1970s and 1980s, the share of women hired by the orchestras increased—from about 10 percent in 1970 to about 35 percent in the mid-1990s. During this time, most orchestras introduced screens that hid the identity and gender of applicants from the hiring panel when they auditioned. Using data from audition records, the investigators found that “blind” auditions increased the probability by 50 percent that a woman would advance from preliminary rounds. The researchers attribute about 30 percent of the gain in the number of female musicians in orchestras to the advent of blind auditions.

Crimes by African Americans are understood differently than crimes by whites. For example, Pager, Western, and Bonikowski (2009) find evidence of the much greater cost to African American job applicants than to white job applicants of having a prison record. The investigators recruited African Americans and whites to apply for the same set of jobs with similar fictitious résumés, except that one group had résumés with a prison record and the other group did not. The participants applied in person for the jobs. African Americans were only half as likely as equally qualified whites to receive a callback or job offer. Moreover, African American applicants *without* a criminal record

were no more likely to receive callbacks or job offers than white applicants *with* a criminal record. The study describes the experience of the participants (the “testers”):

In applying at an auto dealership ... testers met with very different reactions [by race]. Joe, the black tester, was informed at the outset that the only available positions were for people with direct auto sales experience. ... When the employer interviewed Keith, their white ex-felon test partner, he gave him a stern lecture regarding his criminal background. The employer warned, “I have no problem with your conviction, it doesn’t bother me. But if I find out money is missing or you’re not clean or not showing up on time, I have no problem ending the relationship.” Despite the employer’s concerns, Keith was offered the job on the spot. (p. 790)

Salience is a central theme of behavioral economics. If individuals were unboundedly rational, all facts about an individual and a situation would be equally accessible. But for boundedly rational individuals, situations are simplified, and *how* they are simplified reflects attitudes that can contribute to discrimination. Women are promoted at lower rates than men in STEM (science, technology, engineering, and math) fields. Sarsons (2017) shows that an important explanation for the low promotion rate of women economists is that coauthored research publications by female economists matter less for tenure than coauthored research publications by male economists. It is not known how much of this discrimination is implicit. But a recent University of California, Berkeley senior thesis by Alice Wu (2017) found a way to measure attitudes towards women in economics that suggests the importance of implicit discrimination. In an online, anonymous message board that had more than one million posts, “Economics Job Market Rumors,” female economists were much more often than male economists perceived in ways unrelated to their professional roles. The three most common words

used to describe female economists were “hotter,” “lesbian,” and “bb” (baby). For men, the most common words were “mathematician,” “pricing,” and “adviser.”

Implicit prejudicial attitudes can be self-fulfilling, just as explicit prejudice and discrimination can be (Akerlof, 1976). In the empirical study of grocery store cashiers in France mentioned at the beginning of this essay, workers of African origin were substantially less productive on the days they were supervised by implicitly biased managers, where implicit bias was measured by the Implicit Association Test (IAT)) (Glover, Pallais, and Pariente, 2017). Bertrand, Chugh, and Mullainathan (2005) argue that IATs tap unconscious attitudes, which activate a different part of the brain than the parts that engage in conscious deliberation. When the employees of African origin worked with unbiased supervisors, they were 9 percent more productive than clerks of non-African origin; but when they worked with biased supervisors, they had only average productivity. The authors find evidence that the productivity decline arose because biased managers avoided interactions with minorities. One reason for the low supervision was that the supervisors were worried that they would be accused of bias if they made an inappropriate remark in their interactions.

A two-stage experiment in Sweden provides additional evidence that implicit bias contributes to discriminatory behavior (Rooth, 2010). In the first stage of the experiment, employers received equivalent applications to advertised jobs from applicants with Swedish names and from applicants with Arab-Muslim sounding names. All applicants were represented as male. Rooth contacted the recruiters to ask if they would participate in tests for explicit and implicit attitudes towards Arab Muslims. The tests demonstrated that employers had strong, explicit negative attitudes toward Arab-Muslims, though not

on the basis of beliefs regarding lower productivity: A clear majority of employers (77 percent) stated that there were no performance differences between the two groups.

Implicit attitudes predicted the difference in callback rates between applicants with Arab-Muslim sounding and Swedish names much more reliably than did explicit attitudes. The probability of a callback for Arab-Muslim job applicants declined by 5 percent for each one standard deviation increase in negative implicit association of Arab-Muslim men.

If prejudice reflects implicit thoughts, not conscious tastes or statistical discrimination, discriminatory beliefs and attitudes may be sensitive to subtle contextual cues. In one experiment, individuals were asked which group they preferred—a group of well-liked African American athletes or a group of disliked white politicians.

Respondents preferred the first group when the context emphasized occupation, but preferred the second group when it emphasized race (Mitchell, Nosek, and Banaji 2003).

Context influences the salience and valence of categories. Shayo and Zussman (2011) investigate more than 1,500 judicial decisions in Israeli small claims courts, where cases are randomly assigned to Arab or Jewish judges. They find evidence of judicial in-group bias: a claim is 17-20 percent more likely to be accepted if assigned to a judge of the same ethnicity (Arab or Jewish) as the plaintiff. Consistent with the emphasis in behavioral economics on the effects of salience on attention, the ethnic bias increases with the population-adjusted number of fatalities in the year preceding the ruling that are from Palestinian, politically-motivated attacks in the vicinity of the court. Terrorism leads Arab judges to favor Arab plaintiffs, and Jewish judges to favor Jewish plaintiffs. Shayo and Zussman conclude that ethnic conflict, by intensifying ethnic identities, can erode trust in the rule of law:

there is rather little ethnic ingroup bias in the Israeli courts except during periods in which political violence intensifies ethnic identification. In other words, by heightening identification, ethnic conflict can dramatically undermine the proper functioning of an ostensibly impartial institution like the court system.

Interventions to reduce discrimination

Identification of the impact of implicit discrimination leads immediately to the question, can a policy intervention reduce it? Evidence suggests that information alone may not be enough to change beliefs, since individuals tend to resist corrective information that calls their priors into question (Nyhan and Reifler, 2010). (The defensive reaction is called *counter-arguing*.) However, a longstanding body of work on prejudice reduction emphasizes the beneficial impact of interactions between social groups (Allport 1954; Pettigrew and Tropp, 2006).³ Much of this work argues that prejudice results from a lack of experiential knowledge and understanding of outgroups and that it can be overcome by interactions between the groups. A recent quasi-natural experiment provides strong supportive evidence (Rao, 2014). Many private schools in New Delhi lease land from the government at heavily subsidized rates. In exchange, the law requires that the schools admit randomly chosen poor children tuition-free. The law was not enforced until 2007, when a Delhi High Court decision ordered almost 400 private schools to reserve one-fifth of their seats for students from households with incomes of less than about US \$2,000 per year. The schools would be partially compensated by the government. The order applied

³ However, some evidence suggests that repeated interactions between ethnic groups in conflict with each other results in exclusionary attitudes towards the outgroup (Enos, 2014).

only to new admissions of students, who usually enter in pre-school, and required the students to be integrated into the same classrooms as the non-scholarship students.⁴

Rao discovered that exposure for four years to poor classmates and, in particular, regular interaction with them in cooperative groups, changed the social preferences and behavior of the rich students. When offered opportunities to come to school over the weekend to support fund-raising efforts for a charity that served disadvantaged children, rich students in cohorts with poor classmates sent 10 percent more volunteers than those without poor classmates. In dictator games,⁵ rich students who had poor classmates shared 45 percent more of their endowment, and were much more likely to split their endowment equally with the other player.

To measure the impact on discriminatory behavior by the rich students, Rao devised an experiment involving a relay race with a prize to the winner. Before the relay teams were formed, each student ran a sprint individually, which revealed how fast he ran. Then participants selected their teammates in the relay race. As Rao explains, “By having participants choose between more athletic poor students and less athletic rich students, I create a tradeoff between ability and social similarity.” In the relay race with

⁴ As a result, within an affected school, the presence of poor children increased sharply in new cohorts but not in existing, older ones. Some schools delayed taking any action on the plan for a year because enrollment decisions had already been made. The variation – between cohorts within a school and between schools— in exposure of rich children to poor classmates makes it possible to identify the impact on the rich children of social interaction with poor children.

⁵ In a dictator game, there are two players—a dictator and a recipient. The dictator is given an endowment. In this experiment, it was 10 Indian rupees (20 U.S. cents). The dictator makes one decision—how to split the endowment between himself and the recipient. In this experiment, the students were invited to play as the dictator in two games. The recipients were anonymous but the dictator had information on the socio-economic status of their school. In one game, the recipient was from a school with poor children. In the other, the recipient was from an elite private school.

the smallest prize (about US\$0.85), almost one-third of the rich students *without* poor classmates discriminated against poor students. Those *with* poor classmates discriminated 12 percentage points less.⁶

Recent research suggests that even a brief interaction between individuals in in-groups and out-groups can sometimes shift tastes and beliefs. Evidence that bears this out is an experiment on attitudes towards transgender people, who face high levels of prejudice and discrimination in the U.S. In a randomized field experiment, voters were canvassed by activists for transgender rights. The canvassers, some of whom were transgender, told voters that they might be asked to vote on whether to repeal legal protections for transgender people. Voters were asked to discuss their views and were shown a video presenting the arguments on each side. After the video, the canvassers encouraged voters to engage in “perspective taking” – discussing an occasion when they felt judged negatively for being different and considering whether their own experience offered a window into the experiences of transgender people. The brief interaction had a large effect. As compared with a control group (voters who were canvassed on recycling issues), the intervention increased positive attitudes toward transgender people by 10 percentage points. To put this in perspective, the impact is larger than the increase in positive attitudes toward gay men and lesbians in the U.S. between 1998 and 2012 (Broockman and Kalla 2016).

⁶ Boisjoly and others (2006) also find that exposure creates pro-social attitudes and behavior by in-groups toward out-groups. Compared to white students who were not randomly assigned African American roommates in college, white students who were assigned African American roommates were between one-third and one-half of a standard deviation more likely to endorse affirmative action. They reported several years later that they interacted more often and more comfortably with minorities.

There is also evidence that a brief intervention can change behavior in a high-stakes interaction. Disciplinary problems of students are a strong predictor of negative life outcomes; school sanctions early in life can set off a long-lasting negative trajectory (Rocque and Paternoster, 2011). In an experiment in five racially diverse middle schools in California, a brief intervention encouraged math teachers to adopt an empathetic instead of a punitive mindset with regard to discipline (Okonofua and others, 2016). They were told that the purpose was to review “common but sometimes neglected wisdom about teaching and to collect their perspectives as experienced teachers on how best to handle difficult interactions with students, especially disciplinary encounters” (p. 3). The teachers read an article about reasons why students misbehave (*e.g.*, social and biological changes during adolescence, worries, stresses, and social anxiety), which discouraged teachers from labeling students as troublemakers. The article encouraged teachers to place value on students’ experiences and to develop and sustain positive relationships with the students. This was reinforced with stories from students. The intervention neither discouraged disciplinary actions, nor encouraged the view that students’ perspectives were necessarily reasonable. The intervention reduced suspensions from school. Students of teachers who received the treatment were half as likely to be suspended over the academic year.

An alternative approach to shifting behavior is to increase self-awareness. Publicizing the extent of implicit bias may alert people to their unintended discriminatory actions and thereby reduce, or even eliminate, the discrimination. Price and Wolfers (2007) reported that fouls in NBA games in 1991-2002 were more likely to be called against African Americans when the referee team was made up only of whites, and were

more likely to be called against whites when the referee team was made up only of African Americans. *The New York Times* covered the study on its front page. Other newspapers and TV stations, such as ESPN, covered it, too. Six years later, Pope, Price, Wolfers (2013) showed that the discovery and publicity ended the discriminatory behavior of the referees. The new study compared the period 2002-2007 (before the publication of the earlier study) to the period 2007-2010 (after the publication). In-group bias persisted until 2007, but stopped in 2007-2010. The researchers explored whether the change occurred due to a shake-up in the line-up of referees, but found that of the 66 referees who had officiated at least 100 games in 2003-2006, 55 continued to officiate at least 100 games in the period 2007-2010. The researchers also considered whether the NBA had changed the racial make-up of the teams, but found that the fraction of mixed-race referee teams had actually decreased. They also spoke to the NBA league administrators, who indicated that no policies had changed in response to the paper. Thus, no observed change in structural conditions, but only a change in awareness, seems to explain the end of discriminatory refereeing.

It is sometimes possible to make small changes in a decision-making environment to activate scripts that rely less on stereotypes and thereby reduce discrimination. Bohnet, Geen, and Bazerman (2016) constructed a game in which participants had to hire employees for mathematical or verbal tasks. Participants were paid based on the performance of the person they hired. Participants were given the following information about a candidate: (a) performance on a prior test, (b) gender, (c) his or her identity as an American student from the Boston area, and (d) the average performance of the pool from which the candidate was drawn. Participants could hire either someone offered to

them or a randomly selected person from the pool. There were two experimental conditions. In one condition, participants were offered only one candidate. In the other condition, participants were offered a male candidate and a female candidate. All the offered candidates had on a prior test performed at average or slightly below average level of the pool from which they were drawn. IATs show that people have implicit beliefs that men have greater math skills than women, and that women have greater verbal skills than men (Nosek, Banaji, and Greenwald 2002; Plante, Theoret and Favreau 2009). Consistent with this finding, in deciding on a single offered candidate, the participants were more likely to choose the offered male candidate for a math task and the offered female candidate for a verbal task. When evaluated on their own, stereotype-advantaged individuals were chosen 66 percent of the time. However, the stereotype advantage was present only when participants were deciding on a single offered candidate and not present when they had a choice between a man and a woman, whom they could compare to each other.⁷ A fine adjustment to the hiring process eliminated gender discrimination.

The examples presented in this section reduced prejudice for reasons that may be heavily contingent on context. The findings may not be parsimoniously applicable to new settings. They may require adaptation to work in new contexts, or they may not

⁷ This caused a preference reversal in sessions with a math-based task. When candidates were made available separately, 65 percent of participants chose lower-performing males and 44 percent selected higher-performing females. When both a male and female were available for selection, 3 percent chose the lower-performing male and 57 percent of chose the higher-performing female.

work at all. Research in this area is growing rapidly to identify the mechanisms that reduce discrimination.

Barrier 2: Self-stereotyping and self-censorship

Up to now, we have discussed stereotyping of *others*. But individuals also stereotype *themselves*. The stigma of social exclusion can be so profound as to “get into people’s heads” and degrade their self-concept (Goffman, 1963).

To function well, people need to understand themselves in terms that sustain their sense of personal integrity and value and that accurately reflect their abilities. Many institutions are designed to support the psychological and social needs of dominant groups by providing role models, narratives, and rituals to embolden individuals to make the effort needed to succeed. Groups at the margins of society, on the other hand, often must expend extraordinary effort to negotiate their place in environments not made for them or hostile to them. “Like a distracting alarm, psychological threat can also consume mental resources that could otherwise be marshaled for better performance and problem solving” (Cohen and Sherman, 2014, 335). Individuals may spend time assessing whether they belong, whether they are wanted, whether they are good enough, whether they are worthy (Walton and Cohen, 2007). For a given situation, two individuals – one in an in-group, the other in an out-group – may be engaging in very different amounts of mental energy.

Even small pressures on mental resources can reduce the ability to self-regulate. A trivial example demonstrates the point. Shiv and Fedorikhin (1999) asked a group of

students to memorize a number and recall it a few minutes later in another room. They randomly assigned the students to two groups. One group was given a seven-digit number to memorize. The other group was given a two-digit number. When the students left the room, they had a choice of snack as reward for their participation – a bowl of fruit salad or a piece of chocolate cake. Compared to those asked to memorize the two-digit number, those asked to memorize the seven-digit number were 50 percent more likely to choose the cake, the less healthful choice – 63 percent versus 42 percent.

How and when does a negative stereotype become a cognitive tax? The cues in a situation, the individual's perception of the situation, and the relevance attributed to social identity in that situation all influence whether or not the stereotype is activated (Okamura, 1981). For some individuals, a negative stereotype may be chronically activated. Pioneering studies on the effect of priming social identity find that merely checking a box to indicate race before taking an aptitude test lowers the performance of African American students, but not of white students (Steele and Aronson, 1995, 1998). The “race prime” appears to raise the consciousness of negative stereotypes among African Americans. Steele and Aronson (1998, 423) call this effect “stereotype threat”: “Participants who experience stereotype threat spend more time doing fewer items less accurately.”

Stereotype threat has been documented in many contexts. In India, individuals are born into castes, which in each locality are ranked. High-caste individuals are traditionally considered socially and intellectually superior to low-caste individuals (called Scheduled Castes). While discrimination against low caste is illegal, low-caste children encounter the traditional order of caste and untouchability in the fables they

learn and the continued insults, discrimination, and atrocities against upwardly mobile members of low castes. Hoff and Pandey (2006, 2014) assessed the effect of making caste identity public and of caste segregation on the performance of junior high school boys in rural north India. Caste segregation is a mark of the civic privileges of the high castes, and the social exclusion and inferiority of the low castes (Jodhka 2002). The participants were asked to solve mazes and paid for each maze they solved. Participants were randomly assigned to one of three conditions: (a) anonymous, (b) caste revealed in mixed-caste groups, and (c) caste revealed in groups segregated by caste status (high or low). In the first condition, three high-caste and three low-caste boys were placed in a session and their identity and caste were not made public in the session. Since, in general, the children came from six different villages, their caste would not be known to the other children in the session. In the second condition, three high-caste and three low-caste boys were placed in a session and their identity and caste were made public at the beginning of the session. The third condition was the same as the second one except that the six boys in a session were all from high castes or all from low castes.

The anonymous condition showed that low-caste boys solved mazes just as well as high-caste boys. However, publicly revealing caste in mixed-caste groups created a 23 percent caste gap in total mazes solved in favor of the high castes, controlling for other individual variables. A possible explanation is that the boys felt “I can’t or don’t dare to excel.” In the third condition, segregation depressed the performance of both high-caste and low-caste boys. If segregation evokes a sense of entitlement in the high caste, the high-caste boys may have felt, “Why try?”

The experiment to test the effect on maze-solving ability of making a stigmatized identity salient was replicated in Beijing, China, although in this experiment the identity treatment was stronger: in addition to revealing children's social identity, students completed a pre-experiment survey that asked questions about their social identity and about the characteristics of groups with their own and other social identities (Afridi, Li, and Ren, 2015). The subjects were elementary school children of age 8-12 drawn from two social categories: (a) households classified as urban Beijing households, a privileged category, and (b) households classified as rural non-Beijing, a disadvantaged category in Beijing. The household registration system in China, known as *hukou*, classifies citizens based on the birthplace of their parents or grandparents, and favors those categorized as local urban residents in housing, jobs, access to schools, and public benefits. Unlike categories of gender, class, and caste, *hukou* is a transparently man-made creation. The experimental findings show that priming *hukou* shifts performance in ways that mirror the way the groups are ranked. In this sense, the social identity "makes up people," and a group's alleged inferiority becomes "an equilibrium fiction" (Hoff and Stiglitz, 2010).

Most people have difficulty judging their own ability. Coffman (2014) implemented an experiment with U.S. college students that reveals that a person's judgment of his ability in a given domain depends on the interaction between his gender and the gender stereotype associated with the domain. In the experiment, which minimized discrimination and fear of discrimination, female participants under-contributed their ideas in male-typed domains; that is, they *self-censored*. If women are less likely to contribute their ideas, this will hinder their advancement. It is a self-

administered kind of social exclusion. The variable that explains it is the mental model of gender.

Another, related mechanism that can lead to the replication of inequality after structural and institutional barriers have been removed is *coordination* on unequally rewarded tasks. In research-oriented academic departments, individuals have some choice over how much of their time to spend on non-promotable tasks, for example, attending committee meetings, evaluating applicants, and advising undergraduates. There is extensive evidence in academia and industry that women spend more time on non-promotable tasks than men (see references in Babcock and others, 2017). While many factors could explain this—for example, gender differences in preferences and abilities—it could also be driven by shared expectations regarding the appropriate behavior of women and men. To investigate this, Babcock and others ran controlled experiments. They examined how gender affects the allocation between men and women of a relatively poorly paid task. In each of ten rounds, participants—all seated in one large room and each with a computer—were randomly divided into groups of three persons. Members of the group had to make one decision—to volunteer, or not, to be the poorly rewarded person in the group. Each round could last at most two minutes. To volunteer, a person clicked on his or her screen. As soon as a group member clicked, or two minutes had elapsed, the round ended. The incentives were as follows: Each player in a group got \$1 if nobody in the group volunteered to be the poorly rewarded person. If somebody volunteered, the volunteer got \$1.25 and the other two people in the group each got \$2.00.

When the participants in the room were roughly half men and half women, so that individuals knew that their group was very likely to be mixed gender, women volunteered twice as often as men. When the participants in the room were all men or all women, so that individuals knew that their group was single gender, men and women were equally likely to volunteer for the poorly rewarded task. The findings are consistent with the hypothesis that mental models of gender have a large effect on behavior and tend to replicate historical inequalities.

In a follow-up experiment, individuals played a similar game in which there was a fourth member of each group. His or her only role was to try to resolve the coordination problem by asking one member to volunteer to be the poorly rewarded person. Women received more requests than men by a factor of 2.50. The gap increased as the game was repeated over ten rounds. When asked to volunteer, women were 49 percent more likely than men to agree. The results suggest that shared expectations based on traditionally unequal gender roles replicate the inequality when individuals coordinate anonymously in novel situations on unequally rewarded tasks. The traditional gender roles carry over to the novel situations even though there is no structural basis for the carry-over.

Another area in which gender roles affect behavior in ways that replicate historically imposed inequalities is in salary negotiation. Lab and field evidence suggests that men are more likely to ask for higher compensation than women (Small and others, 2007). But seemingly minor situational factors loom large in decision making. In a field experiment, List and Leibbrandt (2014) replicate the result in Small et al.: they find that when salaries are not explicitly made negotiable, women are 23 percent *less* likely to

negotiate for higher salaries; but when it is clear that salaries are negotiable, women are 8 percent *more* likely than men to negotiate for higher salaries.

Interventions to weaken the influence of negative stereotypes on self-assessment

Minor interventions can insulate socially excluded groups from the “threat in the air” created by social stigma. The next five paragraphs discuss experiments with disadvantaged students in the U.S. Then we discuss an intervention in India that reduced the legitimacy of domestic violence.

Experiments in the U.S. show that interventions that inculcate feelings of belonging can improve academic performance among non-traditional college students (Yeager and others, 2016). In one experiment, disadvantaged high school students who had been admitted to two- and four-year colleges were invited to participate in an online module designed to dispel the belief that disadvantaged students are the only group that has difficulty in college or the only ones who question whether they belong in college. One year later, 45 percent of the students who had participated in the intervention were enrolled full-time in school, compared to 32 percent of the control group. A similar experiment, in which participants had already entered college, reduced the enrollment gap between disadvantaged and advantaged students by 40 percent.

An intervention that has been tested in multiple contexts is a values affirmation exercise. In one experiment, an essay assignment was given to students through their normal coursework two times during the school year (Cohen and others, 2006). The assignments asked students to think about their most important value. It took 10-15 minutes to complete an essay on this topic. Students knew their teachers would read their

essays, but the teachers did not know the identity of the students in the treatment. The treatment improved the grade point average of socially excluded groups (African Americans and Latino Americans) in academic courses. It reduced by 50 percent the proportion of African Americans receiving a D or and F in their first term. The effects persisted over the two years in which participants' grade point averages were tracked. African American students' grades improved by 0.24 grade points. Among low-achieving African Americans students, performance improved even more: grade point averages increased by 0.41 points, and the rate of remediation or grade repetition was less than a third of that of the control group (Cohen and others, 2009).

Another way to help socially excluded groups improve performance is to frame the idea of intelligence as a *malleable* trait that grows in response to hard work, rather than as a *fixed* trait (Hong, Chiu, Dweck, Lin, and Wan, 1999; Dweck, 2006; Nussbaum and Dweck, 2008). Disadvantaged groups may be more likely to believe that intelligence is fixed rather than malleable (Claro, Paunesku, and Dweck 2016). In a seminal study, Aronson, Fried, and Good (2002) tested the impact of fostering a “growth mindset” among African American college students.⁸ Students were taught the theory of malleable intelligence in three one-hour lab sessions and encouraged to explain the ideas in a letter to an at-risk middle school student. The intervention increased the participants' belief that intelligence was malleable, their enjoyment of academics, and their belief that academics are important. It increased their semester grade point average from 3.05 to 3.32.

⁸ White students also participated in the study. The intervention improved their performance but by a smaller amount than that of African American students.

Many interventions to counter the belief that intelligence is a fixed trait have been effective in small or lab settings, but can such interventions be effective at scale and for a heterogeneous population? Recent and ongoing work investigates this. Paunesku and others (2015) designed and delivered two online 45-minute-videotapes to over 1,500 students in 13 geographically diverse U.S. high schools. One video communicated the idea of malleable intelligence and growth mindset. The other video encouraged students to reflect on how working hard at school can help them accomplish meaningful goals. One-third of the participants were at risk of dropping out of high school.

Compared to a control group, both treatments raised students' grade point averages in core academic courses. The fraction of at-risk students who satisfactorily completed core courses was 6.4 percentage points higher for the treatment groups than for the control. The results show that interventions can be delivered at low-cost and at scale. The World Bank's behavioral science unit, the Mind, Behavior, and Development Unit (eMBed), has delivered paper-based growth mindset interventions to approximately 40,000 students in Peru and 200,000 students in Indonesia. The preliminary results from Peru are promising.

Narratives known in a community are a source of shared mental representations. Some narratives enable people to see situations in a way that spurs activity. Other narratives constrain agency by representing a person as unable to influence outcomes outside a narrow domain. Famous examples of positive narratives that have shaped many Americans' views of poverty are the rags-to-riches novels by Horatio Alger: no matter how dire the hero's straits at the beginning, every hero in Alger's stories escapes poverty by dint of effort, ability, and inner strength.

A Theater for Development program, *Jana Sanskriti*, active in villages in West Bengal, India since 1985, engages in fieldwork in which villagers describe their problems, including incidents of domestic violence. The artistic director has written plays that are performed in the villages to dramatize the problems many people experience in their daily lives. As Augusto Boal (1973), the Brazilian writer and politician who created Theater for Development, explains:

When the skit is over, the participants are asked if they agree with the solution presented. At least some will say no. At this point it is explained that the play will be performed once more, exactly as it was the first time. But now any participant in the audience has the right to replace any actor and lead the action in the direction that seems to him most appropriate... The other actors have to face the newly created situation, responding instantly to all the possibilities that it may present...

The goal of *Jana Sanskriti* is to enable villagers to change their shared representations, for example, of domestic violence, and collectively rehearse social change. To evaluate the impact, Hoff, Jalan, and Santra (in progress) surveyed random samples of registered voters in villages where the plays had been performed and in matched villages where plays had never been performed. The study finds that exposure of a village to the plays reduced from 14 percent to less than 4 percent the fraction of women who thought domestic violence was legitimate. Exposure also reduced the percentage of households in which domestic violence had recently occurred from 32 percent to 26 percent. By providing an entry point for communities to *collectively* contest traditional social norms, the theater program may expand individuals' cognitive toolkit for interpreting domestic relationships. In the new mental models, domestic violence is perceived as cruel and illegitimate.

Barrier 3: The difficulty of adapting fast thinking to two worlds

The title of Kahneman (2011), *Thinking, Fast and Slow*, captures a central tenet in behavioral economics: individuals have a dual cognitive process. “Fast” thinking is intuitive and automatic, and “slow” thinking is deliberate and effortful. Fast thinking is generally well-adapted to environments that one knows well: “Time and energy are saved, rumination and doubt are reduced, and nothing important is lost” (Ross and Nisbett, 1991, 77; see also Gigerenzer and others, 1999). But in less familiar environments, fast thinking may lead to systematic mistakes. Economically and racially segregated neighborhoods put disadvantaged individuals in a difficult position: they do not access as young children the more privileged settings in which they can learn the mental models, norms, and rules of thumb needed to thrive in such environments.

When the U.S. government offered people living in a poor neighborhood vouchers to move to a higher-income neighborhood, young children’s earnings later in life improved by \$3,477 on average per year, an estimated \$302,000 in a lifetime (Chetty, Hendren, and Katz 2016). A likely benefit of moving to better neighborhoods includes access to the social and cultural capital to function well in middle-income groups (Wilson 1987, Sampson 2012).

Individuals in out-groups must often navigate from a young age two very different cultural worlds – their home environment, with its epistemology and norms, and the world of the in-group, with different epistemology and norms. Children who live in areas plagued by high rates of crime must learn one set of rules of thumb to survive in their neighborhood, and another set to thrive in school. School environments are heavily regulated by formal authority and norms of civility. In contrast, in high crime

neighborhoods, power and authority are fluid and negotiable. Anderson (1999, p. 15) writes that “one of the most salient features of urban life in the minds of many people today is the relative prevalence of violence.” For people living in neighborhoods that lack institutions to prevent crime, conduct is typically regulated by the threat of violence – the “code of the street.” Coates (2015, 23) wrote of his time growing up in Baltimore:

To survive the neighborhoods and shield my body, I learned another language consisting of a basic complement of head nods and handshakes. I memorized a list of prohibited blocks. I learned the smell and the feel of fighting weather. And I learned that “Shorty, can I see your bike” was never a sincere question, and “Yo, you was messing with my cousin” was neither an earnest accusation nor a misunderstanding of the facts. These were the summonses that you answered with your left foot forward, your right foot back, your hands guarding your face, one slightly lower than the other, cocked like a hammer.

Particularly for males who live in dangerous neighborhoods, there is generally no set of automatic responses to the assertion of authority that they can apply successfully in both their neighborhoods and more orderly places like schools. An automatic response of compliance would endanger them in their neighborhoods, whereas an automatic response of non-compliance in school could lead to suspension or expulsion. Middle-class youth normally do not face this problem. For them, the appropriate response to authority is the same in the home and school environment.

Rules of thumb and self-regulation

Traditional economics takes the person’s preferences as fixed. The standard policy prescription to deter disorder and violence is punishment. In contrast, behavioral economics suggests that the person can change his mental models and behaving if given adequate social and psychological supports. This section discusses interventions that have increased individuals’ life skills and raised their aspirations.

One of the simplest ways for socially excluded groups to learn the rules of thumb necessary to succeed in more privileged settings is mentoring and coaching. Bettinger and Baker (2011) evaluated a mentoring and coaching service for non-traditional U.S. college students. Coaches worked with students to help them clarify their aspirations, connect their daily activities to long-term plans, and build skills such as time management and self-advocacy. Coached students were 14 percent more likely to persist in school after 24 months, and four percentage points more likely to complete their degree within four years of receiving the treatment.

All-encompassing interventions are not always needed to change behavior in very positive ways. In some cases, rules of thumb lead to poor outcomes when they are applied to a particular domain. In South Africa, women who seek jobs tend not to ask for reference letters from former employers. When they do ask them to send letters, callback rates increase by 89 percent (Abel, Burger, and Piraino 2017). One might suppose that it is the more able women who seek letters; their superior abilities, not the letters, are the reason for the higher callback rates. But this was not the explanation. A randomized controlled trial (RCT) that encouraged women in the treatment group to seek and use reference letters doubled their employment rates. In contrast, getting reference letters had no effect in the case of male applicants. Data collected three months after the intervention shows that the intervention closed the gender gap in job-search success.

Another strategy to lead people to adopt better rules of thumb in a particular domain is to encourage goal-setting. Deliberating and focusing on specific, challenging goals stimulates goal-directed behavior (Locke and Latham, 1990). Goal-setting focuses

attention on aspired states. It makes salient and poignant the losses incurred if one does not reach one's goal, the relationship between steps necessary to achieve the goal, and the goal itself. An experiment in Canada recruited 85 low-performing university students and randomly assigned half of them to an intervention (Morisano and others, 2010). Students in the treatment group were invited to write down their aspirations, values, role models, priorities, and the ways that achieving their goals would affect the lives of other people. The students in the control group were asked to write about earlier positive experiences. Figure 1 shows that the students who participated in the goal-setting intervention had grades in the next semester almost half a point higher (on a scale of four) than the control group. A related experiment with similar results is Schippers and others, 2015.

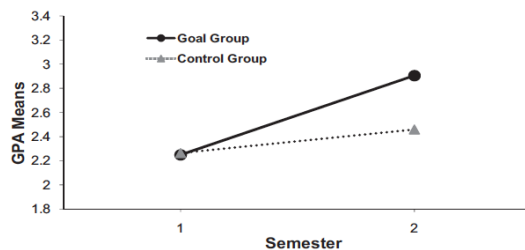


Figure 1. Low-performing students who were treated in the goal-setting intervention made gains in their grade point average (GPA) compared to the control group.

Source: Morisano and others (2010)

One variant of the goal-setting approach is called “WOOP” (wish, outcome, obstacle, plan). It combines *mental contrasting* with detailed *implementation intentions* (Oettingen, 2015). Mental contrasting entails visualizing your goals, the reasons they are important to you and the people around you, and considering the obstacles to achieving them. *Implementation intentions* involve making detailed “if..., then...,” plans to overcome obstacles (Gollwitzer 1999). Oettingen finds that contrasting their goals with

the barriers to achieving them enhances motivation. Duckworth and others (2013) implemented an intervention based on WOOP to eleven-year old children from disadvantaged backgrounds in the U.S. The students were given a worksheet packet and asked write down their most important wish or goal related to school work—“something that is challenging, but that you can achieve within the next few weeks or months,” the instructor explained. The children were also asked to write down “the one best outcome, the one best thing of fulfilling your wish or reaching your goal.” The children were given time to imagine the outcome they had written about, and then randomly assigned to one of two groups. In the treatment group, students were asked to imagine an obstacle they might face in achieving their wish and to create “if..., then...,” plans to overcome it. In the control group, students were asked to imagine a second positive outcome. The treatment group subsequently performed better than the control group: they had higher report card grades, higher attendance rates, and better conduct (Duckworth and others, 2013).

We next return to the central example in our discussion of barrier 3—automatic aggression in response to assertions of authority. Heller and others (2017) evaluated RCTs of cognitive behavioral therapy programs to reduce automatic aggressive responses by disadvantaged young males in the Chicago area. Two of the treatments were a program called “Becoming a Man.” The third treatment was a program in a Juvenile Temporary Detention Center.

A simple activity illustrates how Becoming a Man worked. The activity leaders used a provocative exercise to show how participants automatically followed one strategy rather than weighing all their options. Activity leaders broke groups of participants into

pairs and gave one person in each pair a ball. The other person was instructed to get the ball from him. He was given 30 seconds to do so. The automatic response of almost all the participants was to use force to take the ball. After the exercise, the activity leader pointed out that the participants could have asked for the ball. When prompted for an explanation as to why they had not done that, they usually responded that their partner would not have complied. The activity leader then asked the partner what he would have done if asked, and most partners said that they would have given the ball.

The participants in *Becoming a Man* had an average GPA of 2.0 (out of 4.0), had typically missed six to eight weeks of school in the year, and in many cases had a history of arrests. The intervention reduced participants' interactions with the criminal justice system. The first intervention, delivered to boys in 7th -10th grade, reduced violent crime by 44 percent and non-violent crime by 36 percent.⁹ Program participants also became more engaged in school, which the authors forecast could translate into increases in graduation rates of about 7-22 percent. The second intervention, delivered to boys in ninth and tenth grade, reduced arrests by 31 percent. The third intervention, delivered at the Juvenile Temporary Detention Center, reduced by 16 percentage points the re-admittance rate to the detention center within 18 months of release.

Cognitive therapy interventions have also been tested in Sierra Leone and Liberia. In Sierra Leone, Betancourt and others (2014) evaluated a youth readiness intervention that delivered psychosocial supports to war-affected youth. The goal was to help them regulate their emotions and improve their problem-solving skills. The

⁹The value of crime reduction alone is estimated to yield benefit-cost ratios that range from 5-to-1 up to 30-to-1 (Heller and others, p. 5).

treatment provided all participants with an education subsidy and randomly allocated the psychosocial support treatment. Students who received the psychosocial supports performed better and were more likely to stay in school.

In Liberia, Blattman, Jamison, and Sheridan (2017) partnered with a local organization to provide group-based therapy and/or \$200 cash (about three months' wages) to almost 1,000 criminally engaged men. Participants were randomly divided into four groups: one group received only the cash grant; a second group received eight weeks of therapy designed to foster self-regulation, patience, and a non-criminal identity; a third group received both the grant and therapy; and the control received neither. Those who had received only the cash transfer made no changes in criminal behavior or self-regulation. Among those who received only the therapy, violent and criminal behavior declined – in the short run, the individuals were 55 percent less likely to carry a weapon and 47 percent less likely to sell drugs. The effects were longer lasting and stronger among those who had received the therapy followed by the cash grant. The participants in this third group became less impulsive, had higher self-esteem, and were less likely to steal for at least a year after the intervention.

The studies discussed in this section show that short-term interventions can increase agency. One set of interventions helped individuals form goals and the strategies and emotional skills to achieve them. Another set of interventions gave individuals who grew up in violent neighborhoods the cognitive tools to adapt to a non-violent environment and build productive lives.

Barrier 4: “Adaptive preferences” of the oppressed

Traditional economics, as we have emphasized throughout, assumes fixed preferences, whereas the second strand of behavioral economics does not. In the past two decades, lab and field and “as-if-random” natural experiments (in which we include the work on gender and the plow) have provided evidence that is difficult to explain except through the impact of experience and exposure on preferences. Taking account of the malleability of preferences sheds light on a particularly perverse driver of social exclusion: “adaptive preferences” of the oppressed (Nussbaum, 2001). A subjugated group may come to see its subjugation as natural, normative, or even preferred. Bourdieu (2000, 217) writes that “the realistic, even resigned or fatalistic, dispositions which lead members of the dominated classes to put up with objective conditions that would be judged intolerable or revolting by agents otherwise disposed...*help to reproduce the conditions of oppression*” (emphasis added). Fatalistic attitudes and beliefs can be difficult to tackle in part because they call into question whether adults understand their own preferences. Duflo (2012) describes “hope” as a form of capability.

Examples of adaptive preferences in varying degrees of intensity exist in many contexts. Guyon and Huillery (2014) find evidence of a large social class gap in aspirations of 14-year old students in the area of Paris. In many countries, as shown in figure 2, a high proportion of women report that a husband is justified in beating his wife for refusing sex. One reason that preferences may be adaptive is that existing institutions serve as *reference points*—the baseline relative to which people imagine alternative

realities. People have limited ability to imagine counterfactuals.

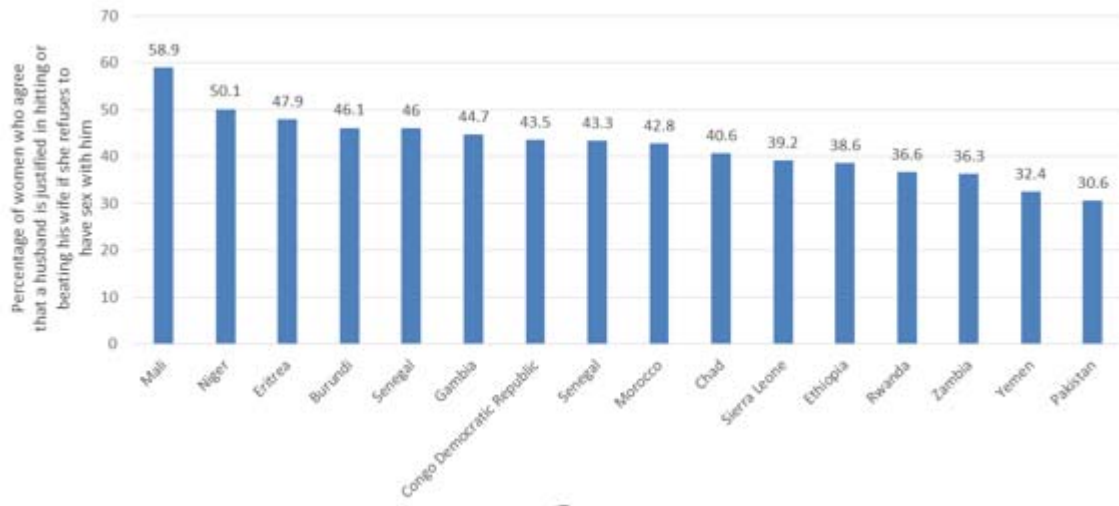


Figure 2. The legitimacy of wife-beating

Source: Based on data from DHS Statcompiler (<http://www.statcompiler.com>)

In parts of rural India, most girls leave school early, marry, and have children at a young age. The explanation in traditional economics would be that the behaviors reflect the preferred outcomes of the girls or their families, given their opportunity sets. However, a recent RCT points also to the role of limited imagination of alternative ways of live. In a large-scale field experiment, Jensen (2012) provided three years of recruiting services to help women in remote villages in India get jobs in the business process outsourcing industry. Although, on average, only three women per village were hired, young women in the treatment villages became less likely at a young age to get married or have children, and more likely to work or continue their education. The proportion of young women age 15-21 who were married dropped from 71 to 66 percent, and the proportion with children dropped from 43 to 37 percent. In addition, the treatment closed 30 percent of the gap in body mass index between girls in the villages and girls in the

wealthiest families in New Delhi. Was this simply a reflection of pre-existing preferences responding to new opportunities? A plausible interpretation is that there was also a change in preferences: Exposure to local village women who got good jobs helped young women imagine better lives for themselves.

One way for adaptive preferences to emerge is through a perverse trusting relationship that oppressed individuals may develop with their oppressors. This is called the Stockholm syndrome (Namnyak and others, 2008). The term emerged from a dramatic event in Sweden. On a summer morning in 1973, a prison-escapee entered a bank with a submachine gun and shot a police officer. In the failed bank robbery, he took four hostages and demanded that his prison mate be released from prison and join him. The government acquiesced. The two men barricaded themselves in the bank, with the hostages locked in the bank vault. Astonishingly, the hostages began to develop a bond with their captors and resisted cooperation with the police. A reverse Stockholm Syndrome also exists. In the so-called Lima Syndrome, captors develop sympathy for their hostages and release valuable hostages (Lama, 2017).

Examples related to the Stockholm Syndrome exist at the level of a whole society. In Sierra Leone, villages are ruled by “paramount chiefs” from families originally recognized by British colonial authorities. The number of ruling families varies across villages due to accidents of history. Villages with fewer ruling families have worse governance, child health, educational attainment, and incomes (Acemoglu, Reed, and Robinson, 2013). One might expect that villages with poorer development outcomes would have less respect for authority and be less satisfied with the ruling families. But

this is not the case. The researchers found that villages with fewer ruling families report higher respect for authority.

In the Indian state of Maharashtra, local government is by a variety of objective measures more oppressive in villages in which the high castes own most of the land. To increase the extent to which the landless depend on them, the high castes block many national pro-poor programs. Yet the perceived legitimacy of village government is higher in high-caste dominated villages: low-caste residents are 14 percent more likely to report trusting the landholders in the high-caste dominated villages as compared with villages where the high castes are not dominant and in which more pro-poor programs from the central government operate (Anderson, Francois, and Kotwal, 2015).

Role models

What can overcome the adaptive preferences of the oppressed? By observing others, individuals may develop higher aspirations and more positively assess their prospects for achieving them (Bandura, 1986, 1997). Successful role models can help individuals imagine new life paths and boost their perceived self-efficacy.

A natural experiment in India shows the impact of creating role models by securing places for women in positions of power. As discussed above, for the position of village council leader (*pradhan*), the government of India mandated in the 1990s reservations for women in a randomly selected one-third of villages. Exposure to women *pradhans* changed beliefs in at least four ways that reduced the social exclusion of women (Beaman and others 2009, 2012; Iyer and others, 2012): (a) The experience of living under a woman *pradhan* erased the prejudice, on average, of male villagers against women

leaders by many measures—the evaluation of political speeches (discussed above), an IAT, and the assessment of the quality of actual village *pradhans*. (b) In villages that had women leaders, parents’ aspirations for their teenage daughters, and teenage girls’ aspirations for themselves, were higher; and girls have gone to school somewhat longer and done somewhat fewer hours of housework. (c) After the reservations ended in a village, women have run for political office in higher percentages and in many cases won; see figure 3. (d) The greater presence of female political representatives produced an unexpected change in women’s reporting of crimes against women, and in the willingness of the police to record the reports. This occurred even though *pradhans* have no jurisdiction over the police. The increased reporting by female victims of crimes appears instead to reflect a change in their perception of the costs—psychic and otherwise—of reporting the crimes.

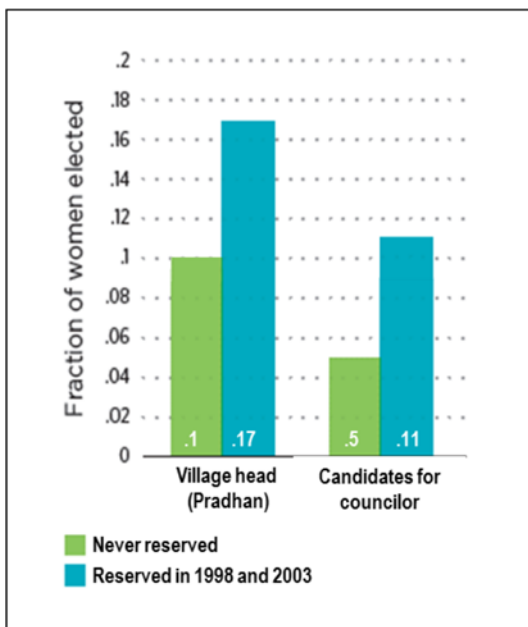


Figure 3. The fraction of women who won office in free elections

Source: Figure from World Bank (2015), based on data from Beaman and others (2009)

Another way to change the adaptive preferences of disadvantaged groups is through documentaries of people who escaped poverty and soap operas that expose viewers to new ways to build their lives. A field experiment conducted by Bernard and colleagues (2014) in Ethiopia examined the impact of showing a one-hour video with profiles of four Ethiopian villages who had escaped poverty or entered the middle class. The team randomly selected about 60 villages for the treatment and about 60 villages for the control group. The locations were remote—the researchers reached them by 4x4 vehicles or camels. Only 10 percent of the local populations watched TV at least once a week. The intervention raised aspirations and expectations, and the effect persisted. Five years later, aspirations were still significantly higher in the treatment than in the control group. The intervention also increased future-oriented behavior. School enrollment of children was 17 percent larger than for the control group six months after the intervention.

There is evidence that exposure to unfamiliar outcomes and social patterns even in fiction can give individuals new role models. The emission of soap operas in Brazil depicting agentic women with few or no children was a statistically significant factor in the recent decline in Brazil's fertility rate. By taking advantage of the plausibly exogenous spread of television across Brazilian municipalities, La Ferrara, Chong, and Duryea (2012) showed that a fertility decline was caused by the soap operas in the year that a municipality first received the emissions. The decline was most pronounced for women who were close in age to a protagonist of one of the soap operas, which is consistent with the idea that the characters were role models. To put this effect in a comparative perspective, the impact was similar to the effect of two additional years in women's education. Fertility fell by 11 percent of the mean among women aged 35–44.

Organizations such as BBC Media Action have brought this approach to scale across the world, tackling problems such as ethnic conflict, poor health, and oppressive gender norms.

Conclusion

A central theme of this essay is that institutions create concepts, categories, social identities, and other mental models that can persist long after the institutions are abolished. The mental models influence how boundedly rational people process information and what they want and aspire to become. Reforming or abolishing an institution to reduce social exclusion may not change the mental models that the institution has advanced. If they continue to influence behavior, social exclusion will persist. The realization of equal *substantive* opportunity may require interventions that target social exclusion at the social and psychological level.

The interpretation of the causes of social exclusion in traditional economics contrasts sharply with the perspective in behavioral economics. If a social group remains at the bottom of the social ladder long after procedural equality of opportunity has been established, the implication in traditional economics would be that the group has fixed characteristics that impede upward mobility, or that individuals in social networks at the low rungs of society will move up only if an event makes possible a coordinated change in behavior.

In contrast, behavioral economics shows that social exclusion is caused not just by structural and institutional barriers, but also by social and psychological factors, and

further, that interventions can change them or go around them. An ascriptive identity that stigmatizes a group affects its members in many ways besides procedural barriers to opportunity and explicit animus. Socially excluded groups face implicit bias. Negative stereotypes affect the group's performance, self-concept, and aspirations and can also drive self-censorship. Growing up in a segregated, disadvantaged neighborhood can give individuals rules of thumb that are poorly adapted to success in school and work. The policy implication in behavioral economics is that equality of *outcomes* between ascriptive social groups, such as those defined by race, gender, or ethnicity, should be a policy target along with formal equality of *opportunity*. Procedural equality as a target leaves unaddressed the schematizing power of the institutions that historically denied opportunity to certain groups. These institutions made social inequalities between ascriptive groups appear normal and possibly normative. A rapidly growing literature, which this essay selectively reviewed, shows that interventions can relax the constraints created by mental models that are a legacy of historical institutions. The impacts of many of the interventions described in this paper are difficult to explain in rational choice theory, but are not difficult to explain under the quasi-rational, enculturated actor framework of behavioral economics.

References

- Abel, Martin, Rulof Burger, and Patrizio Piraino. 2016. "The Value of Reference Letters - Experimental Evidence from South Africa." Working Paper.
- Acemoglu, Daron., T. Reed, and James Robinson. 2014. "Chiefs: Economic development and elite control of civil society in Sierra Leone." *Journal of Political Economy* 122(2): 319-368.
- Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "The Political Legacy of American Slavery." *Journal of Politics* 78(3): 621-641.
- Afridi, Farzana, Sherry Li, and Yufei Ren. 2015. "Social identity and inequality: The impact of China's hukou system." *Journal of Public Economics* 123:17-29.
- Akerlof, George, 1976. "The Economics of Caste and of the Rat Race and Other Woeful Tales." *Quarterly Journal of Economics* 90(4):599-617.
- Akerlof, George. 1997. "Social distance and social decisions." *Econometrica* 65 (5): 1005-1027.
- Alesina, Alberto, Paolo Giuliano, and Nathan Nunn. 2013. "On the Origin of Gender Roles: Women and the Plow." *Quarterly Journal of Economics* 128(2):469–530.
- Allport, Gordon. 1950. "Prejudice: A Problem in Psychological and Social Causation." *Journal of Social Issues* 6:4-23.
- Anderson, Elijah. 1999. *Code of the Street*. New York: Norton.
- Anderson, Siwan, Patrick Francois, and Ashok Kotwal. 2015. "Clientelism in Indian villages." *American Economic Review* 105(6):1780-1816.
- Anthony G. Greenwald and Linda Hamilton Krieger. 2006. "Implicit Bias: Scientific Foundations." *California Law Review* 94: 945
- Aronson, Joshua, Carrie B. Fried, and Catherine Good. 2002. "Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence." *Journal of Experimental Social Psychology* 38(2):113-125.
- Arrow, Kenneth J. 1973. "*The theory of discrimination*." Princeton University Working Paper No. 30A.
- Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart. 2017. "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability." *American Economic Review* 107(3): 714-47.

- Banaji, Mahzarin, R. and Anthony G. Greenwald. 1995. "Implicit social cognition: attitudes, self-esteem, and stereotypes." *Psychological review* 102(1):4.
- Bandura, Albert. 1986. *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall. Inc.
- Bandura, Albert. 1997. *Self-efficacy: The exercise of control*. Macmillan.
- Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. "Female leadership raises aspirations and educational attainment for girls: A policy experiment in India." *Science* 335(6068):582-586.
- Beaman, Lori, Raghavendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics* 124 (4): 1497–1540.
- Becker, Gary S. 1957. *The economics of discrimination*. University of Chicago Press.
- Bedolla, Lisa Garcia and Michele Michelson. 2012. *Mobilizing Inclusion: Transforming the Electorate through Get-Out-the Vote Campaigns*. New Haven, CT: Yale University Press.
- Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Taffesse. 2014. The future in mind: Aspirations and forward-looking behavior in rural Ethiopia. CSAE Working Paper WPS/2014-16 London: Centre for Economic Policy Research.
- Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. 2005. "Implicit Discrimination." *American Economic Review* 95(2):94-98.
- Betancourt, T.S., McBain, R., Newnham, E.A., Akinsulure-Smith, A.M., Brennan, R.T., Weisz, J.R. and Hansen, N.B., 2014. "A behavioral intervention for war-affected youth in Sierra Leone: a randomized controlled trial." *Journal of the American Academy of Child & Adolescent Psychiatry* 53(12):1288-1297.
- Bettinger, Eric and Rachel Baker. 2011. "The Effects of Student Coaching in College: An Evaluation of a Randomized Experiment in Student Mentoring." NBER Working Paper 16881
- Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge, UK: Cambridge University Press.
- Blattman, Christopher, Julian C. Jamison and Margaret Sheridan. 2017. "Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia." *American Economic Review* 107(4): 1165-1206.
- Boal, Augusto. 1973 *The Theatre of the Oppressed*: New York: Routledge Press.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." *American Economic Review* 96(5): 1890-1905.

Bourdieu, Pierre. 2000. *Pascalian meditations*. Stanford University Press.

Broockman, David and Joshua Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door Canvassing." *Science* 352(6282):220-224.

Brooks, Benjamin, Karla Hoff, and Priyanka Pandey. 2017. "Cultural impediments to learning to cooperate: An experimental study of high- and low-caste men in rural India." Manuscript, World Bank.

Bruner, Jerome S., and Mary C. Potter. 1964. "Interference in visual recognition." *Science* 144(3617):424-425.

Bruner, Jerome. 1990. *Acts of Meaning*. Cambridge: Harvard Univ. Press.

Bruner, Jerome. 1991. "The narrative construction of reality." *Critical Inquiry* 18:1-21.

Camerer, Colin F. 2005. "Comments on 'Development Economics through the Lens of Psychology' by Sendhil Mullainathan" in *Lessons of Experience: Annual World Bank Conference on Development Economics*, World Bank.

Carpenter, Joshua and Florian Foos. 2017. "Of UFOs and Politics: How Poor Voters Respond to Policy Promises." University of Alabama, manuscript.

Chetty, Raj, Nathaniel Hendren and Lawrence Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Project." *American Economic Review* 106(4).

Claro, Susana, David Paunesku and Carol Dweck. 2016. "Growth mindset tempers the effects of poverty on academic achievement." *Proceedings of the National Academy of Sciences* 113(31):8664-8668.

Coates, Ta-Nehisi. 2015. *Between the World and Me*. Text publishing.

Coffman, Katherine B. 2014. "Evidence on self-stereotyping and the contribution of ideas." *Quarterly Journal of Economics* 129(4):1625-1660.

Cohen, G.L., Garcia, J., Apfel, N. and Master, A., 2006. "Reducing the racial achievement gap: A social-psychological intervention." *Science* 313(5791):1307-1310.

Cohen, G.L., Garcia, J., Purdie-Vaughns, V., Apfel, N. and Brzustoski, P., 2009. "Recursive processes in self-affirmation: Intervening to close the minority achievement gap." *Science* 324(5925), pp.400-403.

Cohen, Geoffrey and David K. Sherman. 2014. "The psychology of change: Self-affirmation and social psychological intervention." *Annual Review of Psychology* 65:333-371.

D'Andrade, R.G. 1995. *The Development of Cognitive Anthropology*. Cambridge, UK: Cambridge University Press.

Darley, John and Paget Gross. 1983. "A hypothesis-confirming bias in labeling effects." *Journal of Personality and Social Psychology* 44(1):20-33.

DiMaggio, Paul, 1997. "Culture and cognition." *Annual Review of Sociology* 23: 263-287.

DiMaggio, Paul and Filiz Garip. 2012. "Network effects and social inequality." *Annual Review of Sociology* 38: 93-118.

Douglas, Mary. 1966. *Purity and Danger*. Routledge & Kegan Paul.

Douglas, Mary. 1986. *How Institutions Think*. Syracuse, NY: Syracuse University Press.

Dovidio, J.F. and Gaertner, S.L., 2004. "Aversive racism." *Advances in experimental social psychology*, 36: 1-52.

Dovidio, J.F., Kawakami, K., Johnson, C., Johnson, B. and Howard, A., 1997. "On the nature of prejudice: Automatic and controlled processes." *Journal of experimental social psychology*, 33(5):510-540.

Duckworth, Angela L., Teri Kirby, Anton Gollwitzer, Gabriele Oettingen. 2013. "From fantasy to action mental contrasting with implementation intentions (MCII) improves academic performance in children." *Social Psychological and Personality* 4(6):745-753.

Duflo, E. (2012) Tanner Lectures on Human Values and the Design of the Fight Against Poverty, manuscript, MIT, pp.28-52.

Dweck, Carol. 2006. *Mindset: The new psychology of success*. Random House.

Enos, Ryan D. 2014. "Causal Effect of Intergroup Contact on Exclusionary Attitudes." *Proceedings of the National Academy of Sciences* 111(10): 3699-3704.

Farjan, Gabriela, Alaka Holla, and Renos Vakis, in progress. "Do Teachers Use Their Biases to Assess Students' Performance?" World Bank, in progress.

Gaertner, S.L., 1973. "Helping behavior and racial discrimination among Liberals and conservatives." *Journal of Personality and Social Psychology*, 25(3), p.335.

Gigerenzer, Gerd, Peter M. Todd, and the ABC Research Group. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Glover, Dylan, Amanda Pallais and William Pariente. 2017. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." *Quarterly Journal of Economics* 132(3):1219-1260.

Goffman, Erving. 1963. Stigma: Notes on a spoiled identity. Jenkins, JH & Carpenter.

Goldin, Claudia and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review* 90(4): 715-741.

Gollwitzer, Peter. 1999. "Implementation Intentions: Strong Effects of Simple Plans." *American Psychological Association* 54(7):493-503.

Guyon, Nina and Elise Huillery. 2014. "The Aspiration-Poverty Trap: Why Do Students from Low Social Background Limit Their Ambition? Evidence from France." *LIEPP Working Paper*, Department of Economics, Sciences Po, Paris.

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, Harold A. Pollack. 2017. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." *Quarterly Journal of Economics* 132(1): 1-54.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In search of homo economicus: Behavioral experiments in 15 small-scale societies." *American Economic Review* 91(2):73-78.

Hoff, Karla, Jyotsna Jalan, and Sattwik Santra. In progress. "Can we tell a new story? The impact of theater for development on education aspirations and domestic violence." Presented at the International Economic Association conference, Mexico, 2017.

Hoff, Karla, Mayuresh Kshetramade and Ernst Fehr. 2011. "Caste and Punishment: The Legacy of Caste Culture in Norm Enforcement." *Economic Journal* 121: F449-F475.

Hoff, Karla and Priyanka Pandey. 2006. "Discrimination, Social Identity, and Durable Inequalities." *American Economic Review* 96 (2): 206-211.

Hoff, Karla and Priyanka Pandey. 2014. "Making Up People: The Effect of Identity on Performance in a Modernizing Society." *Journal of Development Economics* 106: 118-131.

Hoff, Karla and Joseph E. Stiglitz. 2010. "Equilibrium fictions: A cognitive approach to societal rigidity." *American Economic Review* 100:141-146.

Hoff, Karla, and Joseph. E. Stiglitz. 2016. "Striving for balance in economics: Towards a theory of the social determination of behavior." *Journal of Economic Behavior and Organization* 126, Part B: 25-57.

Hong, Ying-yi, Chu Yue Chiu, Carol S. Dweck, C.S., Derrick M.S. Lin, and Wendy W.N. Wan. 1999. "Implicit theories, attributions, and coping: A meaning system approach." *Journal of Personality and Social psychology* 77(3):588.

Iris Bohnet, Alexandra van Geen, and Max Bazerman. 2016. "When Performance Trumps Gender Bias. Joint Versus Separate Evaluation." *Management Science* 62(5):1225-1234.

Isabelle Plante, Manon Theoret, and Olga E. Favreau. 2009. "Student gender stereotypes: Contrasting the perceived maleness and femaleness of mathematics and language." *Educational Psychology* 29(4):385-405.

Iyer, Lakshmi, Anandi Mani, Prachi Mishra, and Petia Topalova. 2012. "The power of political voice: Women's political representation and crime in India." *American Economic Journal: Applied Economics* 4(4):165-193.

Jensen, Robert. 2012. "Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India." *Quarterly Journal of Economics* 127:753-792.

Jodhka, Surinder S. 2002. "Caste and Untouchability in Rural Punjab" *Economic and Political Weekly* 37(19)

Kahan, Dan, Ellen Peters, Erica C. Dawson, and Paul Slovic. 2017. "Motivated numeracy and enlightened self-government." *Behavioral Public Policy* 1(1):54-86.

Kahneman, Daniel, 2011. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap operas and fertility: Evidence from Brazil." *American Economic Journal: Applied Economics* 4(4):1-31.

Lama, Abraham, 2017. "Peru: Tale of a kidnapping—from Stockholm to Lima Syndrome." Inter Press Service News Agency.

Leibbrandt, Andreas and John A. List. 2014. "Do women avoid salary negotiations? Evidence from a large-scale natural field experiment." *Management Science* 61(9):2016-2024.

Locke, Edwin A., and Gary P. Latham. 1990. *A theory of goal setting & task performance*. Prentice-Hall, Inc.

- Markus, Hazel and Shinobu Kitayama. 1991. "Culture and the self: Implications for cognition, emotion, and motivation." *Psychological Review* 98(2):224-253.
- Markus, Hazel and Shinobu Kitayama. 2010. "Cultures and selves: a cycle of mutual constitution." *Perspectives on Psychological Science* 5(4): 420-430.
- Massey, Douglas S. 2007. *Categorically unequal: The American stratification system*. Russell Sage Foundation.
- Mitchell, Jason P., Brian Nosek, and Mahzarin R. Banaji. 2003. "Contextual variations in implicit evaluation." *Journal of Experimental Psychology: General* 132(3):455.
- Mokyr, Joel. 2009. *The Enlightened Economy: An Economic History of Britain 1700-1850*. Yale University Press.
- Morisano, Dominique, Jacob B. Hirsh, Jordan B. Peterson, Robert O. Phil and Bruce M. Shore. 2010. "Setting, elaborating, and reflecting on personal goals improves academic performance." *Journal of Applied Psychology* 95(2):255.
- Myers, Joseph R. 2003. *The search to belong: Rethinking intimacy, community, and small groups*. Zondervan.
- Namnyak, M., Tufton, N., Szekely, R., Toal, M., Worboys, S. and Sampson, E.L., 2008. "'Stockholm syndrome': psychiatric diagnosis or urban myth?" *Acta Psychiatrica Scandinavica*, 117(1):4-11.
- Nisbett, Richard E., and Lee Ross. 1991. *The Person and the Situation*. NY: McGraw Hill.
- Nosek, Brian, Mahzarin Banaji and Anthony Greenwald. 2002. "Math=Male, Me = Female, Therefore \neq Me." *Journal of Personality and Social Psychology* 83(1):44-59.
- Nosek, Brian, Frederick L. Smyth, Natarajan Sriram, Nicole Lindner, Thierry Devos, Alfonso Ayala, Yoav Bar-Anan, Robin Bergh, Huajian Cai, Karen Gonsalkorale, and Kesebir, S., 2009. "National differences in gender–science stereotypes predict national sex differences in science and math achievement." *Proceedings of the National Academy of Sciences* 106(26): 10593-10597.
- Nussbaum, David A., and Carol S. Dweck. 2008. "Defensiveness versus remediation: Self-theories and modes of self-esteem maintenance." *Personality and Social Psychology Bulletin* 34(5): 599-612.
- Nussbaum, Martha C. 2001. "Symposium on Amartya Sen's philosophy: Adaptive preferences and women's options." *Economics and Philosophy* 17(01):67-88.

Nyhan, B. and Reifler, J., 2010. "When corrections fail: The persistence of political misperceptions." *Political Behavior*, 32(2): 303-330.

Oettingen, Gabriele. 2015. *Rethinking positive thinking: Inside the new science of motivation*.

Okamura, Jonathan Y. 1981. "Situational ethnicity." *Ethnic and Racial Studies* 4(4):452-465.

Okonofua, Jason, David Paunesku, and Gregory M. Walton. 2016. "Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents." *Proceedings of the National Academy of Sciences*.

Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. "Discrimination in a low-wage labor market a field experiment." *American sociological review* 74(5):777-799.

Paunesku, David, Gregory M. Walton, Carissa Romero, Eric N. Smith, David S. Yeager, and Carol S. Dweck. 2015. "Mind-set interventions are a scalable treatment for academic underachievement." *Psychological Science* p.0956797615571017

Pettigrew, Thomas F., and Linda R. Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of Personality and Social Psychology* 90(5):751.

Pope, Devin, Joseph Price, and Justin Wolfers. 2013. "Awareness reduces racial bias." National Bureau of Economic Research Working Paper No. 19765.

Price, Joseph and Justin Wolfers. 2007. "Racial Discrimination Among NBA Referees." NBER Working Paper# 13206.

Rao, G. 2014. "Familiarity does *not* breed contempt: Diversity, discrimination and generosity in Delhi Schools." Job Market Paper.

Rocque, Michael and Raymond Paternoster. 2011. "Understanding the Antecedents of the 'School-to-Jail' Link: The Relationship between Race and School Discipline." *Journal of Criminal law and Criminology* 10(2):633-666.

Rooth, Dan-Olof. 2010. "Automatic associations and discrimination in hiring: Real world evidence." *Labour Economics* 17(3):523-534.

Sampson, Robert J. 2012. "Moving and the Neighborhood Glass Ceiling." *Science* 337(6101):1464-1465

Sarsons, Heather. 2017. "Recognition for Group Work: Gender Differences in Academia." *American Economic Review* 107(5):141-145.

Schippers, Michaéla, Scheepers, Ad W. A. Scheepers, and Jordan B. Peterson. 2015. "A Scalable Goal-Setting Intervention Closes Both the Gender and Ethnic Minority Achievement Gap." *Palgrave Communications* 1: 15014- 2015.

Shayo, Moses and Asaf Zussman. 2011. "Judicial Ingroup Bias in the Shadow of Terrorism." *Quarterly Journal of Economics*, 126(3): 1447-1484.

Sherman, D.K., Hartson, K.A., Binning, K.R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., Tomassetti, S., Nussbaum, A.D. and Cohen, G.L., 2013. "Deflecting the trajectory and changing the narrative: how self-affirmation affects academic performance and motivation under identity threat." *Journal of Personality and Social Psychology* 104(4):591.

Shiv, Baba and Alexander Fedorikhin. 1999. "Heart and mind in conflict: The interplay of affect and cognition in consumer decision making." *Journal of consumer Research* 26(3):278-292

Small, D. A., Gelfand, M., Babcock, L., & Gettman, H. 2007. "Who goes to the bargaining table? The influence of gender and framing on the initiation of negotiation." *Journal of Personality and Social Psychology* 93(4) 600-613.

Steele, Claude and Joshua Aronson. 1995. "Stereotype threat and the intellectual test performance of African Americans." *Journal of personality and social psychology* 69(5):797.

Steele, Claude and Joshua Aronson. 1998. "Stereotype threat and the test performance of academically successful African Americans". In Jencks & M. Phillips (Eds.), *The Black-White test score gap* (401-427). Washington, DC: Brookings Institute.

Strauss, Claudia and Naomi Quinn. 1997. *A cognitive theory of cultural meaning*. New York: Cambridge University Press.

Swidler, Ann. 1986. "Culture in action: Symbols and strategies." *American Sociological Review* 273-286.

Thaler, Richard H. 2016. "Behavioral Economics: Past, Present, and Future." *American Economic Review* 106(7):1577-1600

Walton, Gregory M., and Geoffrey L. Cohen. 2007. "A question of belonging: race, social fit, and achievement" *Journal of Personality and Social Psychology* 92(1):82.

Wilson, T.D., Lindsey, S. and Schooler, T.Y., 2000. A model of dual attitudes. *Psychological Review* 107(1), 101.

World Bank. 2015. *World Development Report 2015: Mind, Society, and Behavior*. Washington, DC: World Bank.

Wilson, William J. 1987. *The truly disadvantaged: The inner city, the underclass, and public policy*. University of Chicago Press.

Wu, Alice H., 2017. “Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum.” Available at SSRN: <https://ssrn.com/abstract=3051462>

Yeager, David S., Gregory M. Walton, Shannon T. Brady, Ezgi N. Akcinar, David Paunesku, Laura Keane, Donald Kamentz, Gretchen Ritter, Angela L. Duckworth, Robert Urstein, and Eric M. Gomez. 2016. “Teaching a lay theory before college narrows achievement gaps at scale.” *Proceedings of the National Academy of Sciences* 113(24): E3341-E3348.