Psychological Foundations of Implicit Bias: Mechanisms and Mitigators

Geoff Kaufman
Human-Computer Interaction Institute
Carnegie Mellon University
gfk@cs.cmu.edu
What is implicit bias?
HOW WE MAKE DECISIONS

SYSTEM 1
- FAST
- PARALLEL
- AUTOMATIC
- EFFORTLESS
- ASSOCIATIVE
- SLOW-LEARNING

SYSTEM 2
- SLOW
- SERIAL
- CONTROLLED
- EFFORT-FILLED
- RULE-GOVERNED
- FLEXIBLE
Our brains are evolutionarily hard-wired to store learned information for rapid retrieval and automatic judgments. Over 95% of cognition is relegated to the System 1 “auto-pilot.”
Psychological Perspective on Implicit Bias

Stereotypes inevitably form because of the innate tendency of the human mind to:

- **Categorize** the world to simplify processing
- **Store** learned information in mental representations (called *schemas*)
- **Automatically and unconsciously activate** stored information whenever one encounters a category member
Stereotypes are internalized as associations through natural processes of learning and categorization.
Implicit biases are distressingly pervasive, operate largely unconsciously, and can automatically influence the ways in which we see and treat others, even when we are determined to be fair and objective.
Implicit Bias Can Have an Automatic (and Unrecognized) Impact on Judgments & Behaviors

Racism in a resume

Job applicants with African American–sounding names got fewer callbacks.

<table>
<thead>
<tr>
<th>Resume quality</th>
<th>White names</th>
<th>African American names</th>
</tr>
</thead>
</table>
| Low            | ![Bar chart showing callback rate for low-resume quality](image)
| High           | ![Bar chart showing callback rate for high-resume quality](image) |

Average callback rate, %

Fig. 1. Competence, hireability, and mentoring by student gender condition (collapsed across faculty gender). All student gender differences are significant ($P < 0.001$). Scales range from 1 to 7, with higher numbers reflecting a greater extent of each variable. Error bars represent SEs. $n_{male student condition} = 63$, $n_{female student condition} = 64$. 

Moss-Racusin et al. (2012)
Women considered better coders - but only if they hide their gender

Researchers find software repository GitHub approved code written by women at a higher rate than code written by men, but only if the gender was not disclosed.

A 2013 survey found only 11.2% of software developers are women. Photograph: Antonio Zazueta Olmos/Antonio Olmos

Terrell et al. (2016)
**Implicit Bias Manifests in Subtle Ways in the Form of Micro-inequities**

**Micro-inequities:** ephemeral, covert, unintentional, frequently unrecognized events that reinforce power dynamics or perceptions of “difference”

**Examples:** slights, exclusions, slips of the tongue, nonverbal signals, unchecked assumptions, unequal expectations, etc.
### Examples of Disability Microaggressions in Everyday Life (Keller & Galgay, 2010)

<table>
<thead>
<tr>
<th>Theme</th>
<th>Example</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Denial of personal identity:</strong> occurs when any aspect of a person's identity other than disability is ignored or denied</td>
<td>“Oh, you’re married?”</td>
<td>There is no part of your life that is normal or like mine. The only thing I see when I look at you is your disability.</td>
</tr>
<tr>
<td><strong>Helplessness:</strong> occurs when people frantically try to help PWDs</td>
<td>Someone helps you onto a bus or train, even when you need no help.</td>
<td>You can't do anything by yourself because you have a disability.</td>
</tr>
<tr>
<td><strong>Infantilization:</strong> occurs when a PWD is treated like a child</td>
<td>“Let me do that for you.”</td>
<td>You are not really capable. I know better.</td>
</tr>
<tr>
<td><strong>Spread effect:</strong> occurs when other expectations about a person are assumed to be due to one specific disability</td>
<td>&quot;Your other senses must be better than mine.&quot;</td>
<td>You’re not normal.</td>
</tr>
<tr>
<td><strong>Secondary gain:</strong> occurs when a person expects to feel good or be praised for doing something for a PWD</td>
<td>&quot;We're going to raise enough money tonight to get Johnny that new wheelchair.&quot;</td>
<td>I feel good and get recognition for being nice to you</td>
</tr>
</tbody>
</table>
Implicit Bias Undermines Self-belief, Achievement, and Well-being

• **Stereotype threat** (Steele & Aronson, 1995):
  – fear of confirming a negative stereotype about one’s group
  – Often leads to **anxiety** and **negative feelings** that can use up mental resources and undermine one’s confidence and ability to succeed
  – Exacerbated by repeated experiences with microaggressions reducing one’s sense of belonging or self-belief in a particular domain (e.g., women in STEM: Beasley & Fischer, 2012; Shapiro & Williams, 2012)
“[At Princeton] I felt like a visitor landing in an alien country. I have spent my years since Princeton, while at law school, and in my various professional jobs, not feeling completely a part of the worlds I inhabit.”

—Sonia Sotomayor, U.S. Supreme Court Justice

“I have written eleven books, but each time I think, Uh oh, they’re going to find me out now.”

Maya Angelou

impostorsyndrome.com
What about System 2?

• Think of System 1 as the “auto-pilot” of cognition and System 2 as the human pilot that takes over when necessary.

• System 1 and System 2 have complementary trade-offs.
  – System 1 directs thoughts, feelings, & behaviors quickly and effortlessly, but is vulnerable to errors (including implicit bias)
  – System 2 allows us to override or correct System 1 thinking and analyze a situation slowly, deliberately, and effortfully, but is cognitively expensive.
Devine’s (1999) Dissociation Model

• **System 1: Stereotype Activation**
  – Stereotypes are firmly implanted (and reinforced) by learning and exposure, cognitive processes of categorization, etc.
  – Thus, stereotypes are automatically activated whenever a cue is present, regardless of personal prejudice level
  – Devine characterizes stereotyping as a “mental habit”

• **System 2: Preventing Stereotype Application**
  – Once a stereotype is activated, people can use System 2 processes to overcome the influence of the stereotype
  – Because controlled processes take motivation and effort, they can’t (or won’t) always be used.
  – Must first be aware of the activation of stereotypes, then take steps to mitigate their impact or weaken their power…
Techniques for Mitigating Implicit Bias
Growing evidence that implicit associations are malleable and can be “unlearned.”

- Relies on the construction of new associations and the cultivation of new mindsets to override or overpower existing associations.
- Requires “intention, attention, and time” (Devine et al., 2012)
- Practice and repetition are key!

“Like stretched rubber bands, the associations modified... likely soon return to their earlier configuration. Such elastic changes can be consequential, but they will require reaplication prior to each occasion on which one wishes them to be in effect.” – Banaji & Greenwald (2013, p. 152)
Counter-stereotypic Training:
Deliberately and repeatedly negating stereotypes or associating individuals with counter-stereotypic traits or attributes

Blair et al. (2001); Kang et al. (2012); Kawakami et al. (2000); Wittenbrink, Judd, & Park (2001)
Mindset Training:
Cultivating a deliberative mindset, reminding oneself of egalitarian goals, reinforcing curiosity and constructive uncertainty about others

Beattie et al. (2013); Sassenberg & Moskowitz (2005); Stone & Moskowitz (2011)
Mind Full, or Mindful?

Meditation:
Mindfulness meditation and “loving-kindness” meditation training have been shown to reduce outgroup biases

Kang et al. (2014); Lueke & Gibson (2015)
Counter-stereotypic Exemplars:
Reminding oneself of or surrounding oneself with people who defy stereotypes

Dasgupta & Asgari (2004); Dasgupta & Greenwald (2001); Kang & Banaji (2006)
Repeated exposure to unexpected, atypical, and counterstereotypical exemplars in the party game *Buffalo* shown to promote more diverse, inclusive representations of social categories.

*Kaufman, Flanagan, & Seidman (2015, 2016)*
Intergroup Contact:
Requires equal status and common goals, a cooperative environment with frequent informal interactions, and the presence of support from authority figures or customs.

Allport (1954); Peruche & Plant (2006); Pettigrew & Tropp (2006)
Practicing Empathy and Perspective-taking: Trying to understand others’ unique subjective experiences and points of view

Benforado & Hanson (2008); Galinsky & Moskowitz (2000); Todd et al. (2011)
The party games *Awkward Moment* and *Awkward Moment at Work* present hypothetical occurrences of bias; players compete to submit the best responses to these “moments.”

Research has shown the games increase players’ understanding of the experience of bias and increase empathy and perspective-taking.

Kaufman, Flanagan, & Seidman (2015, 2016)
Contact and Perspective-taking through Fiction: Fictional worlds of stories, games, and VR environments have proven to be effective spaces for bias reduction

Kaufman & Libby (2012)
"Delayed Revelation" as a Bias Reduction Technique

Withholding a character’s “outgroup” membership until a reader or player has formed an affinity toward or psychological connection with a game character

**Example 1**: disclosing a character’s outgroup membership (sexual orientation or race) in a fictional narrative later (versus earlier) in the story led to higher levels of identification with the character and lower levels of prejudice afterward (Kaufman & Libby, 2012)

**Example 2**: adolescent male players of the game Monarch (e.g., which features all female characters) reported higher identification with the character and reduced gender bias if the revelation of character gender occurred later (versus earlier) in the game (Kaufman, Flanagan, & Freedman, under review)
Gone Home utilizes both misdirection and delayed revelation to great effect (which goes better left un-spoiled!).

Dream Daddy subverts expectations about what the game will feature (e.g., a send-up of gay stereotypes) with what it does deliver (e.g., “dad” stereotypes, exploration of father-daughter dynamics).
Reducing Bias through Simulation and Games:
Fictional worlds of stories, games, and VR environments have proven to be effective spaces for bias reduction

Maister et al. (2015)
Current Project

Developing and deploying an interactive storytelling platform for undergraduates to relay their issues impacting their belongingness and well-being.

Investigating the use of creative nonfiction and fictionalization in personal storytelling contexts and the employment of technology-enabled embodiment to encourage greater comfort with self-disclosure on the part of authors and facilitate greater empathy on the part of readers.
Bias Detection Software Tools:
The development of apps and platforms to detect bias (e.g., in language, nonverbal behaviors, etc.) has become a growing business.

Giang (2015)
Bias Detection Software Tools:
The development of apps and platforms to detect bias (e.g., in language, nonverbal behaviors, etc.) has become a growing business

Giang (2015)
Physiological Sensing of Bias:
Increasing evidence that implicit bias is associated with distinct patterns of physiological responses (heart rate, brain activity, galvanic skin response, respiration rate, eyeblink patterns, etc.)

Dambrun et al. (2003)
MICROINCLUSIONS

DAILY ACTS OF INCLUSION THAT MAKE A DIFFERENCE
Take-home Points

- Implicit bias is pervasive (but malleable)
- Implicit bias manifests in often subtle ways to affect perceptions, expectations, behaviors, social dynamics, etc.
- These subtle effects can profoundly affect their targets’ well-being and sense of belongingness in a particular context
- Awareness of bias is only the first step: essential to engage in individual and collective efforts to combat implicit bias
- Tremendous (but still largely unrealized) potential for technological tools to assist in the detection and mitigation of bias
Geoff Kaufman
Human-Computer Interaction Institute
Carnegie Mellon University
gfk@cs.cmu.edu

Thank you!