Debiasing in Education: Administration, Advising, & Teaching

This handout, online: byrdnick.com/archives/14563/debiasing

Nick Byrd
Florida State University
byrdnick.com byrd_nick byrdnick

Bio. Nick studies psychology and philosophy. In a recent paper, he shows how strong evidence suggests that implicit bias is probably associative, but that debiasing is not fully unconscious or involuntary (Byrd, 2019).

Abstract. When it comes to implicit bias, there is good news and bad news. Sustained changes in implicit bias seem to require regular exposure to experiences that last more than just a few minutes. So, the bad news is that researchers will rarely change implicit biases with brief, one-shot experimental manipulations. The good news, however, is that we can probably reduce implicit biases over time by being more careful about whether and how we include people in leadership, decisions, departments, and instruction. This presentation (1) reviews two methodologically strong debiasing experiments, (2) presents the qualitative results of an easy-to-use debiasing protocol for presentations and teaching, and (3) prompts discussion about how these findings apply to your work.

Background: What is implicit bias? Why should we care about it?

- Implicit bias: indirectly measured, not explicitly endorsed biases in behavior (e.g., reflexive responses to faces).
- Affect Misattribution Procedure: rate subliminal (100 ms) images (Miles, Charron-Chénier, & Schleifer, 2019)
- Implicit Association Test (IAT): categorize items into both descriptive categories and normative categories. (Greenwald, McGhee, & Schwartz, 1998). Example: https://implicit.harvard.edu/implicit/takeatest.html
- Common observation from Race IAT performance: People's performance is often implicitly Pro-White.
  - People are faster to categorize white faces as positive than black faces.
  - People are faster to categorize black faces as negative than white faces.
  - People are more prone to errors when instructed to pair black faces with Good and white faces with Bad than when instructed to pair white faces with Good and black faces with Bad.
  - …regardless of people's consciously endorsed racial preferences (e.g., Gaertner & McLaughlin, 1983)
- Even subtle, small, or rare phenomena can have striking, large, or broad social effects (Greenwald et al., 2015)
  - E.g., compounding consequences of subtle, small, or rare changes in interest rates, subsidies, tariffs, etc.

1. Debiasing Experiments

45-minute, In-person Debiasing Session ➞ Up to 56 Days of Decreased Implicit Bias (Devine et al., 2012)

- 91 non-Black introductory psychology students (67% female, 85% White)
- Pre-test Implicit Association Test (IAT), randomly assigned to Treatment or Control, post-test IATs
- *Control*: No task. Dismissed until 28 and 56 day follow-up IATs.
- *Treatment*: "Narrated …interactive slideshow" about implicit bias, 5 debiasing strategies (ibid., p. 7)
  - *Stereotype replacement.* Identify the stereotypes that inform our responses and replace them with responses that are not based on stereotypes (Monteith, 1993).
  - *Counter-stereotypic imaging.* Imagine counter-stereotypical exemplars (Blair et al., 2001).
  - *Individuation.* Focus on individual features of someone rather than stereotypes about them (Brewer, 1988).
  - *Perspective taking*. Imagine the first-person perspective of a member of a stereotyped group rather than the stereotypes about their group (Galinsky & Moskowitz, 2000).
  - *Increasing opportunities for contact.* Seek out positive experiences with members of other groups rather than imagine stereotypically negative experiences with members of that group (Pettigrew & Tropp, 2006).
- *Immediately after debiasing session*: Treatment significantly reduced implicit bias, $F(88) = 7.95$, $p = 0.006$
- *28 and 56 days later*: sustained reduction (i.e., not significantly different than 👆), $F(88) = 0.67$, $p = 0.42$

5-minute, Mostly Online, Debiasing Session ➞ 4 Days of Decreased Implicit Bias (Lai et al., 2016, Study 2)

- 4888 non-Black undergraduates (60.3% White, 69.4% female) from 17 universities in the US
- Pre-test Implicit Association Test (IAT), randomly assigned to Treatments or Control, post-test IATs
- *Control*: No task. Immediate follow-up IAT and 2 to 4 day follow-up IAT.
- *Treatment*: randomly assigned to one of 9 debiasing tasks. (Many similar to what Devine et al. taught).
  - *Vivid counterstereotypic scenario.* Read a vivid story about a White villain and a Black hero (Dasgupta & Greenwald, 2001) and keep that in mind during the post-manipulation IAT.
  - *Counterstereotypic IAT*. Practice 32 trials of the IAT in which Black is paired with Good and White is paired with Bad, including some famously positive Black figures such as Oprah and some famously negative White figures such as Hitler (Joy-Gaba & Nosek, 2010).
  - *Competition with shifted group boundaries*. Play a simulated dodgeball game in which one's own teammates are Black and play well and one's opponents are White and play poorly.
  - *Shifting group affiliations under threat*. Read a vivid story about the threat of postnuclear war in which one's closest friends are Black and helpful and one's enemies are White.
  - *Priming multiculturalism*. Read a pro-multiculturalism excerpt, summarize it in one's own words, and list reasons that multi-culturalism improves group relations (Richeson & Nussbaum, 2004).
  - *Evaluative conditioning*. Observe 20 Black faces paired with positive words and 20 White faces paired with negative words.
  - *Evaluative conditioning with Go/No-Go task.* Press a button when a Black face is paired with a positive word, do not press a button with a Black face is paired with a negative word, and count the number of Black-positive pairings (Nosek & Banaji, 2001).
  - *Implementation intentions*. Learn that one can override bias by thinking of conditional intentions like, "If I see a Black face, then I will respond by thinking 'good'" (Gollwitzer, 1999).
  - *Faking the IAT*. Learn about Pro-White biases on the IAT and how to intentionally manipulate one's responses times in order for the test to detect a Pro-Black bias (Cvencek, Greenwald, Brown, Gray, & Snowden, 2010).
- *Immediately post-task*: Treatment sig. reduced implicit bias, *F*s(1, 1000–1045) = 6.16–286.7, *p*s = 0.001–0.01
- *2 to 4 days later:* none of the treatments significantly reduced IAT scores relative to control

Big Picture: What Worked?

- Longer (45-minute), in-person debiasing worked better than shorter (5 minute), mostly online debiasing
- Counterconditioning: associating certain (e.g. racial) features with something good or non-stereotypic
  - Less reflective version: the good/non-stereotypic content is paired with (racial) features *for you*
  - More reflective version: *you imagine/seek* good/non-stereotypic content paired with (racial) features

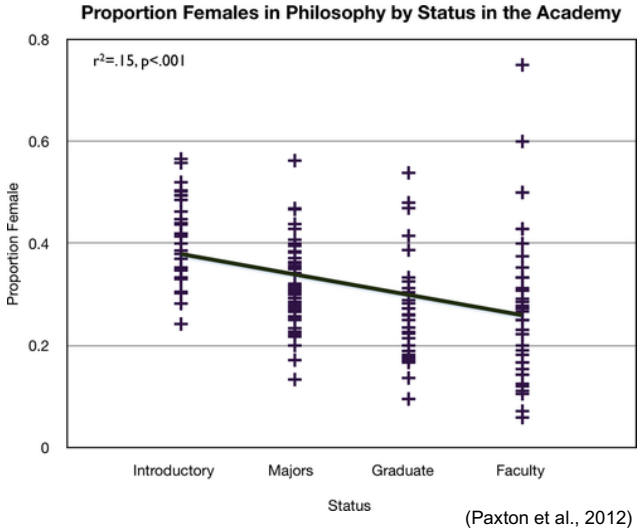
2. Example of Classroom Debiasing

Context: Representation of Women in Philosophy

Representation of women in all majors vs. in philosophy
- 60% of college graduates (Dept. of Education, 2013)
- 23.3% of philosophy graduates (APA, 2018)

Representation of women in philosophy:
- 2006: 18.7% of faculty (Buckwalter & Stich, 2011)
- 2009: 25.4% of faculty (ibid.)
- 2012: 32% of faculty (Paxton, Figdor, & Tiberius, 2012)
- 2017: 25.1 % of APA members (APA, 2018)

Could these data be related to implicit bias? (Saul, 2013)



(Paxton et al., 2012)

Content: Gender Representation in Philosophy Courses

Course: Introduction to Philosophy. Topics covered: Metaphilosophy, Epistemology, Ethics, Science

- Day 1 Activity: "Close your eyes and imagine a philosopher doing philosophy. Watch the philosopher for awhile and focus on what you see in this imagination."

Describe what you are imagining.	**What are they doing?**	**What do they look like?**
"An old guy."	"Wearing a toga?"	"…beard" "…white hair."
"The thinker statue…"	"…thinking." "reading."	"Bronze?"
"A bust of a dead dude…"	"…uhh…"	"marble"

- Debiasing protocol: show pictures of assigned/discussed philosophers only if they are women.
- Last Day (day before final exam): Repeat Day 1 Activity: "Close your eyes …imagine a philosopher…"

Describe what you are imagining.	**What are they doing?**
"Kate Rawles…"	"teaching me about biodiversity."
"Heather Douglas…"	"talking to philosophers and scientists."
"Peter Singer"	"telling us about effective altruism."
"Someone cool" "Someone relatable"	"playing devil's advocate about all of my intuitions."

Consequences: Public Stances Toward Philosophy

Imagine how course content shapes stances toward a field. E.g., philosophy course content answers queries like,

- "Is philosophy something that people still do? Does philosophy still matter? To who? In what ways?"
- "I enjoy philosophy, but can people like me do philosophy? Can people like me be philosophers?"
- "Even if I feel no ambition to study philosophy, how do I feel about philosophy? …about philosophers?"

3. Application

Three Cs of Debiasing For Inclusion & Diversity

Context: Include underrepresented sources and people, etc. in leadership, decisions, departments, training, etc.
Content: Include non-stereotypic and positive representations of negatively stereotyped experiences, people, etc.
Consequences: Assess leadership, decisions, training, etc. by (*inter alia*) the inclusivity of its context and content.

Example: Course Design (Some questions adapted from or inspired by Brantmeier et al.)

- Context (see also Course Evaluation questions 1-3)
  - What are some stereotypes about the topic, the field, the people in the field, the ideas in the field, etc.?
  - Who will be enrolled in the course? (Consider major, SES, race, gender, language, religion, ability, etc.)
  - What kinds of implicit bias might inform your students sense of connection to its field, practitioners, etc.?
  - How might students' experiences differ from one another or from instructor's as a result of implicit biases?
  - How might implicit biases or differences in experience impact decisions to enroll in (or drop) the course?
- Content (see also Course Evaluation questions 4-10)
  - What does this course do to create non-stereotypic or positive representations of underrepresented people?
  - How might this course create better representations of its field, the field's practitioners, its students, etc.?
  - How might this course reinforce biases about its field or its students? How might it countercondition them?
  - How can this course include more abilities, perspectives, learning preferences, modes of assessment, etc.?
- Consequences (see also Course Evaluation questions 11-13)
  - What will students do after this course? Who will students become? What can students do with the course?
  - How can the course's perspectives, representations, and sources impact students' futures for better? Worse?
  - How can representations created and reinforced by the course impact the future of the field? Society?
  - How might the course reinforce negative representations? How can it create better representations?

Example: Course Evaluation (Some items adapted from or inspired by Brantmeier et al.)

1. How much do you think stereotype(s) and implicit bias(es) impact interest and/or enrollment in this course?

   Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

2. How much does the course accommodate or appeal to the full range of students' abilities and needs?

   Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

3. To what extent does the course serve the full range of learning preferences, majors, career aspirations, etc.?

   Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

4. How many new, non-obvious representations of the course's field, methods, and practitioners did you find?

   None ---------- Not many ---------- A few ---------- Many ---------- It was all new to me

5. How much does the course allow, encourage, or reward creative, divergent, or non-stereotypic thinking?

   Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

6. How much does the course expose students to more than just the instructor's experience and perspective?

   Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

7. How much does the course allow students to share the responsibility of learning and teaching its material?

   Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

8. How varied are the course's materials, teaching instruments, assignment types, and modes of assessment?

   Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

9. How much can the course accommodate unexpected learning opportunities that arise in the classroom?

   Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

10. How easily can the course be customized to reflect changes in class size, class demographics, duration, etc.?

    Not at all ---------- Not very ---------- Somewhat ---------- Very ---------- Maximally

11. How much do you think the course reinforces negative stereotypes about or underrepresentation in its field?

    Maximally ---------- A lot ---------- A little ---------- Not much ---------- Not at all

12. How much do you think the course counterconditions implicit biases about stereotyped people in general?

    Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

13. To what extent do the course materials seem to be freely available to or sharable with those not enrolled?

    Not at all ---------- Not much ---------- A little ---------- A lot ---------- Maximally

Notes

Acknowledgements

This project has been improved by data from Florida State University's Fellows Society and Diversity & Inclusion in Research and Teaching Organization and by guidance from Hannah Byrd, Brittney Dennis, Kerestin Goodman, Kate Hill, Eve Humphrey, Lisa Liseno, Yancili Lozano, Kate Pierson, and Valeria Rigobon.

References

Alston, M., Darity Jr., W., Eckel, C., McNeil, L., & Sharpe, R. (In prep.). The Effect of Stereotypes on Black College Test Scores at a Historically Black University. Retreived from https://sites.google.com/view/mackenziealston/research

American Philosophical Association. (2018). *Demographic Statistics on the APA Membership, FY2016 to FY2018*. Retrieved from https://www.apaonline.org/page/demographics

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828–841. DOI: 10.1037/0022-3514.81.5.828

Brantmeier, E., Broscheid, A., & Moore, C. S.. Inclusion By Design: Survey Your Syllabus and Course Design. Retrieved from https://www.academia.edu/20197002/

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull, R. S. Wyer, & Jr. (Eds.), *A dual process model of impression formation* (pp. 1–36). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc. Retrieved from https://psycnet.apa.org/record/1988-98192-001

Buckwalter, W., & Stich, S. (2010). Gender and Philosophical Intuition. In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (Vol. 2, pp. 307–346). Oxford: Oxford University Press. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1966324

Byrd, N. (2019). What we can (and can't) infer about implicit bias from debiasing experiments. *Synthese*, (Online first), 1–29. DOI: 10.1007/s11229-019-02128-6

Cvencek, D., Greenwald, A. G., Brown, A. S., Gray, N. S., & Snowden, R. J. (2010). Faking of the Implicit Association Test Is Statistically Detectable and Partly Correctable. *Basic and Applied Social Psychology*, *32*(4), 302–314. DOI: 10.1080/01973533.2010.519236

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*(5), 800–814. DOI: 10.1037//0022-3514.81.5.800

Department of Education (2013). National Center for Education Statistics, Higher Education General Information Survey.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 48(6), 1267–1278. DOI: 10.1016/j.jesp.2012.06.003

Gaertner, S. L., & McLaughlin, J. P. (1983). Racial Stereotypes: Associations and Ascriptions of Positive and Negative Characteristics. *Social Psychology Quarterly*, *46*(1), 23–30. DOI: 10.2307/3033657

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*(4), 708–724. DOI: 10.1037/0022-3514.78.4.708

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493–503. DOI: 10.1037/0003-066X.54.7.493

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*(4), 553–561. DOI: 10.1037/pspa0000016

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. DOI: 10.1037/0022-3514.74.6.1464

Joy-Gaba, J. A., & Nosek, B. A. (2010). The Surprisingly Limited Malleability of Implicit Racial Evaluations. *Social Psychology*, *41*(3), 137–146. DOI: 10.1027/1864-9335/a000020

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., … Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology. General*, *145*(8), 1001–1016. DOI: 10.1037/xge0000179

Miles, A., Charron-Chénier, R., & Schleifer, C. (2019). Measuring Automatic Cognition: Advancing Dual-Process Research in Sociology. *American Sociological Review*, 0003122419832497. DOI: 10.1177/0003122419832497

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*(3), 469. DOI: 10.1037/0022-3514.65.3.469

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition*, *19*(6), 625–666. DOI: 10.1521/soco.19.6.625.20886

Paxton, M., Figdor, C., & Tiberius, V. (2012). Quantifying the Gender Gap: An Empirical Study of the Underrepresentation of Women in Philosophy. *Hypatia*, *27*(4), 949–957. DOI: 10.1111/j.1527-2001.2012.01306.x

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751. DOI: 10.1037/0022-3514.90.5.751

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*(3), 417–423. DOI: 10.1016/j.jesp.2003.09.002

Saul, J. (2013). Implicit bias, stereotype threat and women in philosophy. In K. Hutchison & F. Jenkins (Eds.), *Women in philosophy: What needs to change* (pp. 39–60). Oxford University Press.