# Biased Against Debiasing: On the Role of (Institutionally Sponsored) Self-Transformation in the Struggle Against Prejudice

Alex Madva

December 3rd 2015

# Counter-Conditioning

Prejudices and stereotypes can be transformed (or weakened) by counterconditioning.

Kerry Kawakami and colleagues:

- Counterstereotype training
- Approach training

Counterconditioning (or “debiasing”) reduces:

- Stereotype accessibility and application
- Implicit prejudice and unreflective social behavior
- Susceptibility to stereotype threat

# With No “Real-World” Application?

Cited to show that entrenched attitudes are capable of change,
but dismissed as lacking practical import.
→ nobody is doing field studies to test “real-world” effects.

Why are researchers (and activists!) pessimistic about debiasing?

3 reasons for pessimism:

- Relearning Worry: individuals will quickly relearn biases.
- Unfeasibility Worry: too laborious or time-consuming.
- Individualism Worry: counterproductive distraction from institutional problems

# Road Map

Survey 4 types of debiasing procedures

The 3 reasons for pessimism are unwarranted

- Relearning Worry: presupposes a discredited view of prejudice.

- Unfeasibility Worry: arises (ironically) from social and cognitive biases.

- Individualism Worry: changing institutions, without changing individual biases, can backfire

My view: a comprehensive program for combating prejudice should include a role for these debiasing procedures.

# Alcohol-Avoidance Training

Wiers et al. (2011) (replicated by Eberl et al. 2013)

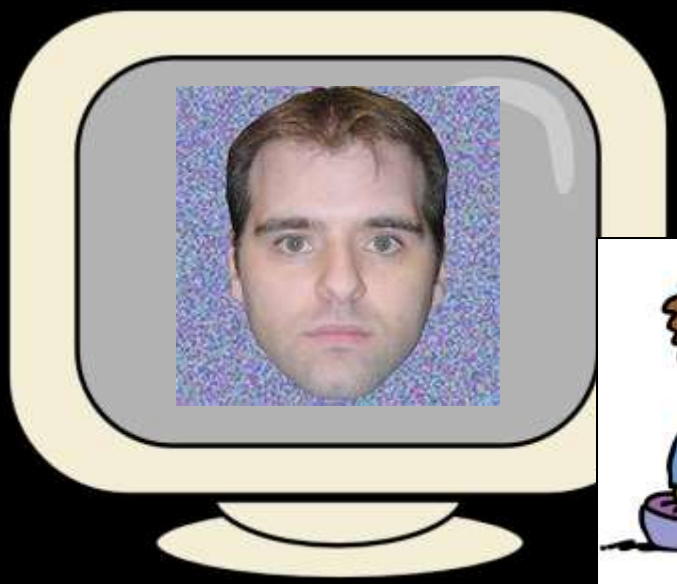

Prior to treatment, alcoholics worked through 4 sessions of 15 min.

→ Less likely to relapse at least one year later

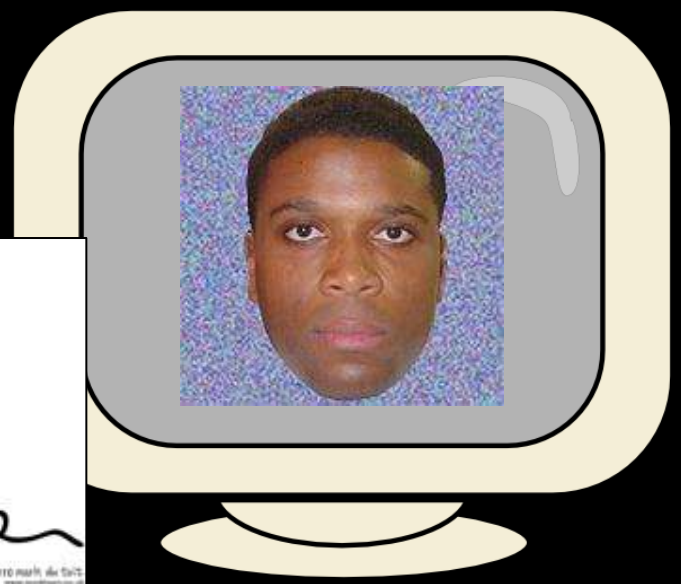
- Minimal addition to therapy with surprisingly durable effects
- Not done in isolation, but as one part of comprehensive treatment

# Approach-Avoidance Training

Kawakami, Phills, et al. (2007)



AVOID!
(push joystick away)



APPROACH!
(pull joystick in)

# Approach-Avoidance Training

Kawakami, Phills, et al. (2007)

→ Reduced implicit racial prejudice (on an IAT).

→ Changed unreflective social behavior during interracial interaction.

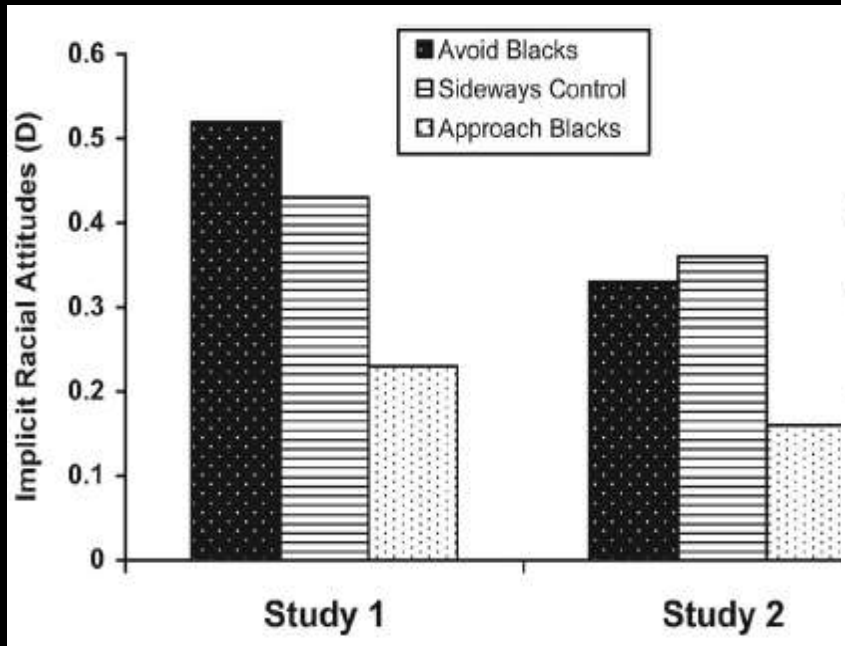



*Figure 1.* Effects of supraliminal (Study 1) and subliminal (St training procedures on implicit racial attitudes.

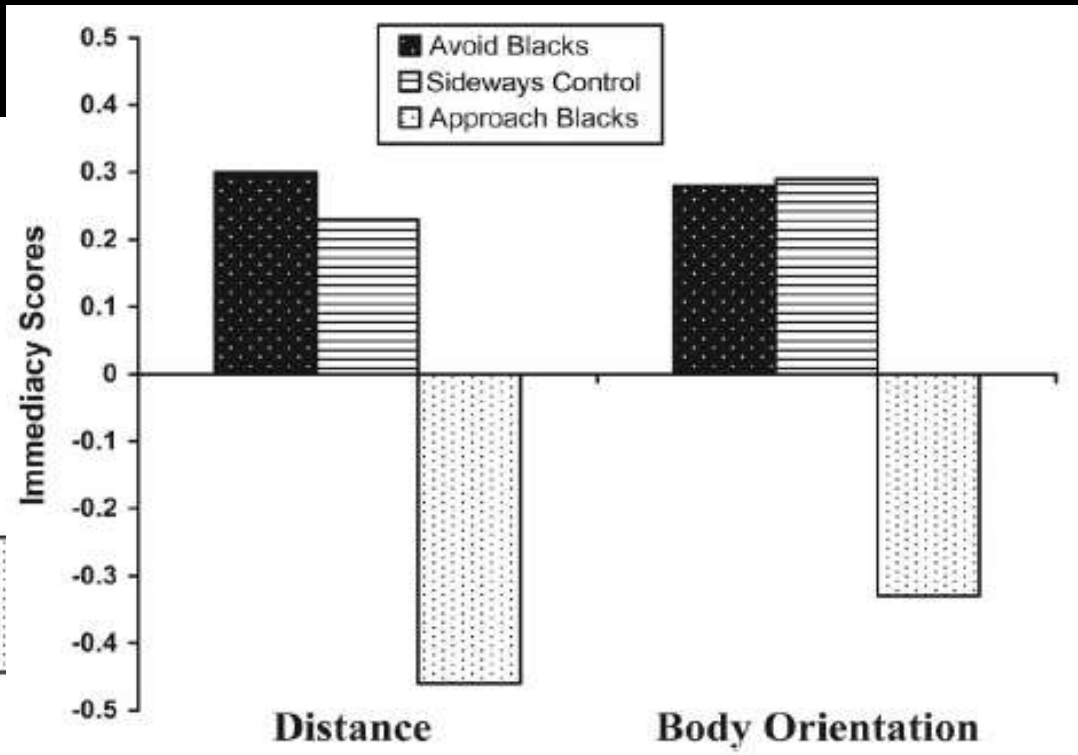



*Figure 3.* Effects of a subliminal training procedure in Study 4 on distance from and body orientation toward a Black confederate.

# Gender Counterstereotype Training

Kawakami, Dovidio, and van Kamp (2005, 2007)

# Gender Counterstereotype Training

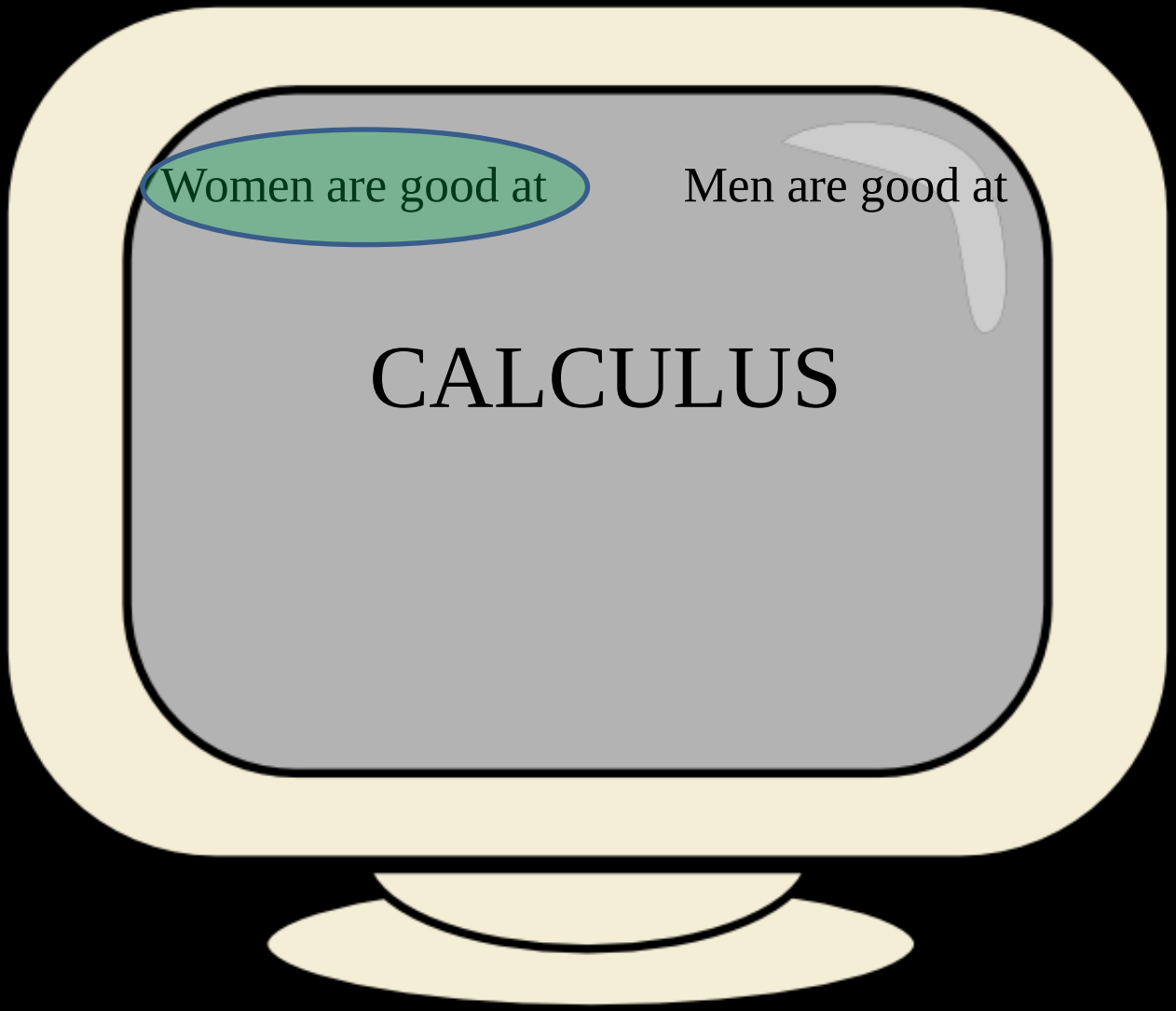
Kawakami, Dovidio, and van Kamp (2005, 2007)

Participants evaluated 4 job applications for a leadership position.

- All 4 applicants were qualified
- 2 women and 2 men (counterbalanced)
- Among those without training, only 35% chose a woman
- Among those with training, 61% chose a woman

# Countering Stereotype Threat

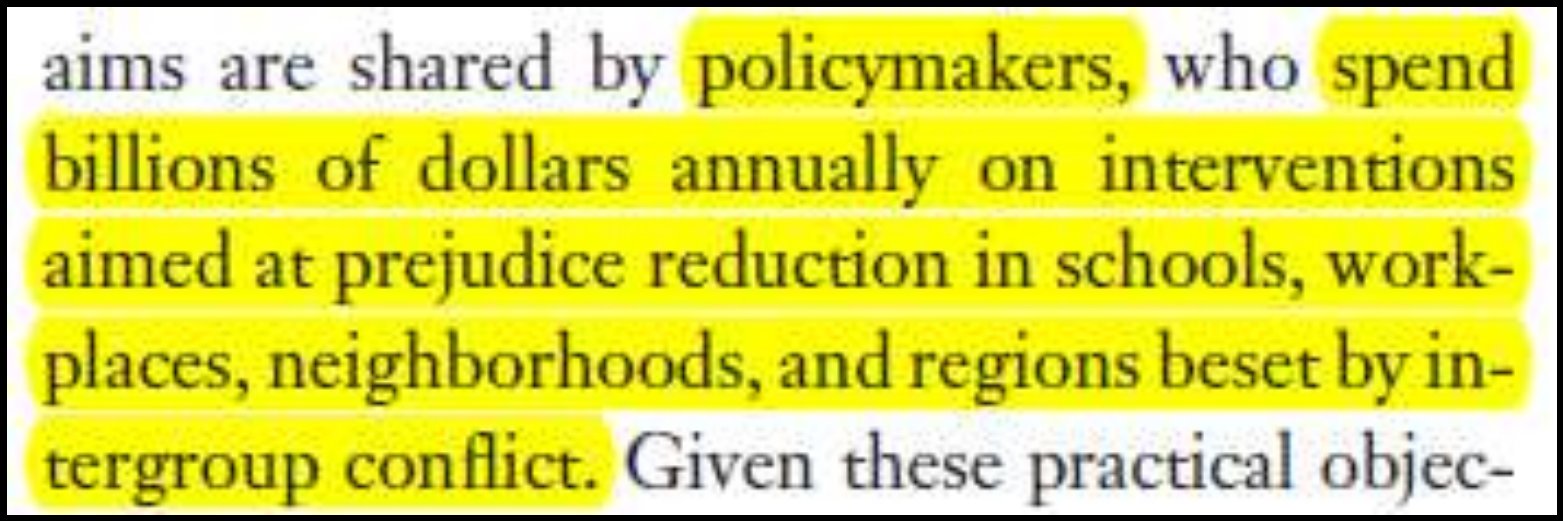
Forbes and Schmader (2010)




24-30 hours later:

women performed better on Math GRE and tests of working memory

# This seems like it should be a big deal...


aims are shared by policymakers, who spend billions of dollars annually on interventions aimed at prejudice reduction in schools, workplaces, neighborhoods, and regions beset by intergroup conflict. Given these practical objec-

Levy Paluck and Green (2009), "Prejudice Reduction: What Works? A Review and Assessment of Research and Practice"

Some common interventions show no evidence of prejudice reduction, and some are not empirically tested at all.

Might debiasing procedures be a boon to these efforts?

# David Schneider (2004)
# *The Psychology of Stereotyping*

"Obviously, in everyday life people are not likely to get such deliberate training,

but it is certainly possible that those who routinely have positive and nonstereotypic experiences with people from stereotyped groups

will replace a cultural stereotype with one that is more individual and generally less negative" (423).

→ Obviously people won't actually do these procedures,
but positive intergroup interactions can reduce bias.

# Only “indirect” practical import?

Scientists (and activists) assume we must “translate” these procedures out of their artificial lab context into an applied, “real-world” setting.

Even the scientists doing these studies! (Phills, Kawakami, et al. 2011)

more positive ways with a Black confederate. The next step for this research, however, would be to test these procedures in a more applied setting. For example, one possible strategy is to have schools implement morning welcome activities in which students from different ethnic/racial groups approach one another. These

# Reasons for Pessimism: Relearning Worry

How long do effects last?  Many think: not very long

Saaid Mendoza et al. (2010, 520): prejudice reduction from these procedures "may be more difficult to maintain upon reexposure to societal stereotypes outside the laboratory."

Bryce Huebner (forthcoming): "as we watch or read the news, watch films, rely on tacit assumptions about what is likely to happen in particular neighborhoods, or draw elicit inferences on the basis of the way in which a person is dressed, we cause ourselves to backslide into our implicit biases."

# Relearning Worry

Basic Conjecture: step outside the lab, get bombarded with stereotypes and reacquire biases.

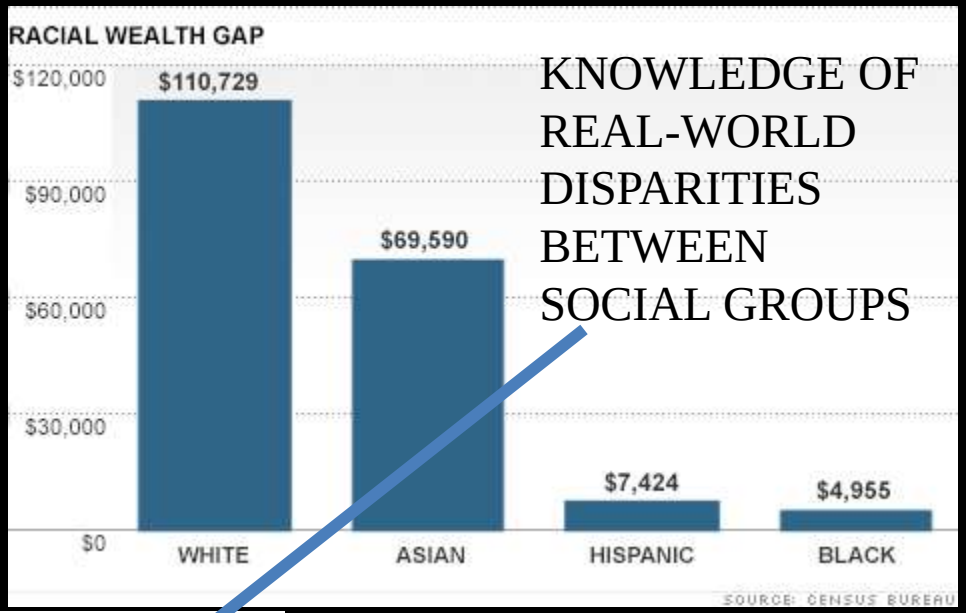
→ “Bombardment Basis” for Relearning Worry

Seems to arise from a Commonsense View of prejudice:

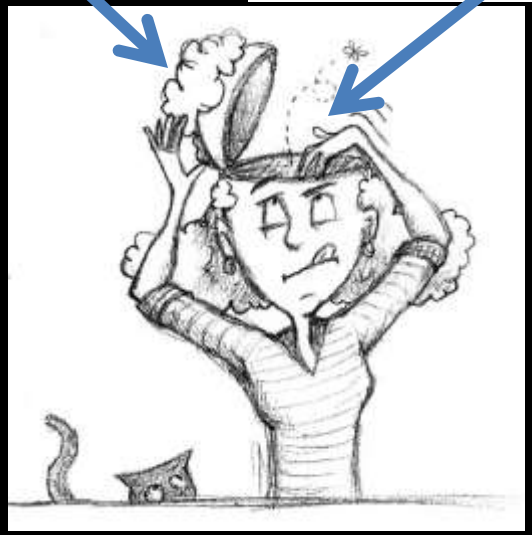
- Acquisition: acquire prejudices through repeated exposure to negative representations of social groups (esp. in early life).

- Ingrained: prejudices become deeply ingrained and subsequently difficult to change.

# Commonsense View of Prejudice


KNOWLEDGE OF REAL-WORLD DISPARITIES BETWEEN SOCIAL GROUPS



# Commonsense View & Bombardment Basis in Social Psychology

Dasgupta (2013)

Recall that we started with the assumption that implicit preference for some groups and bias against others are learned associations acquired by passive immersion in an unequal society where people are segregated into disparate roles, jobs, and geographies based on group membership. In every-

by conscious acts of individuals who possess them. Rather they are mirror-like reflections of local environments and communities within which individuals are immersed. Changes in these environments and communities (and sometimes emotions elicited by them) produce changes in implicit attitudes and beliefs about one's outgroup, ingroup, and the self. Put differently, I propose that implicit attitudes and beliefs are situational adaptations. As situations change, so too do implicit reactions.

# Effects of “Mass Media”

Implicit racial prejudice increases after...

listening to violent hip hop, but not pop
(Rudman and Lee 2002)

watching TV of whites displaying nonverbal bias toward blacks
(Weisbuch, Pauker, and Ambady 2009)

# Debiasing via Affirmative Action?

Since social attitudes "mirror" environments...

→ a "new" argument for affirmative action

→ affirmative action creates "debiasing agents" (Jolls and Sunstein 2006, Kang and Banaji 2006)

cult legal and policy questions. What if, instead, drawing on results such as Dasgupta and Asgari's study, an institution hires certain people because of their debiasing capacity on their students, customers, or employees? After

# But the Commonsense View is wrong!

Our minds are not perfect “mirrors” of social regularities

- We are not “empty heads” that get filled up with whatever bombards us.

Confirmation bias:
seek out and attend to what confirms our beliefs

Belief perseverance:
beliefs persist despite contravening evidence

# Individuals (Mis)interpreting Affirmative Action

A company with women and minorities occupying prominent positions. Will these individuals debias their peers?

If coworkers believe they have been promoted to satisfy a quota (or to be Debiasing Agents!), coworkers may perceive undue benefits, under-evaluate performances, etc.

Research suggests that diversity initiatives...

- can promote a perception that disadvantaged are unfairly advantaged
- can lead privileged individuals to become more discriminatory.

(Kaiser et al. 2012; "Report on the Status of Women Faculty in the Schools of Science and Engineering at MIT," 2011)

# Stereotyping Depends on Goals

Moskowitz (2010)

Implicit and explicit goals shape what we notice & how we interpret

Goals that promote stereotyping: boost self-esteem, justify inequality

Goals that inhibit stereotyping: be egalitarian, be creative, treat another as an individual, take an outsider's perspective on things

# Stereotypes in the World... Counterstereotypes in the Mind

(Nothing either good or bad but thinking makes it so?)

Suppose, after debiasing, we are bombarded with more stereotypical than counterstereotypical depictions of social groups.

Now confirmation bias, belief perseverance, and egalitarian goals might preserve debiased attitudes:

Seek out and notice counterstereotypes, ignore or discount stereotypes

# Effects of Commonsense View on Debiasing Research

My aim is not to insist *a priori* that the Relearning Worry is groundless.

→ Durability of debiasing: an open empirical question to be explored.

→ However, researchers are not exploring it.

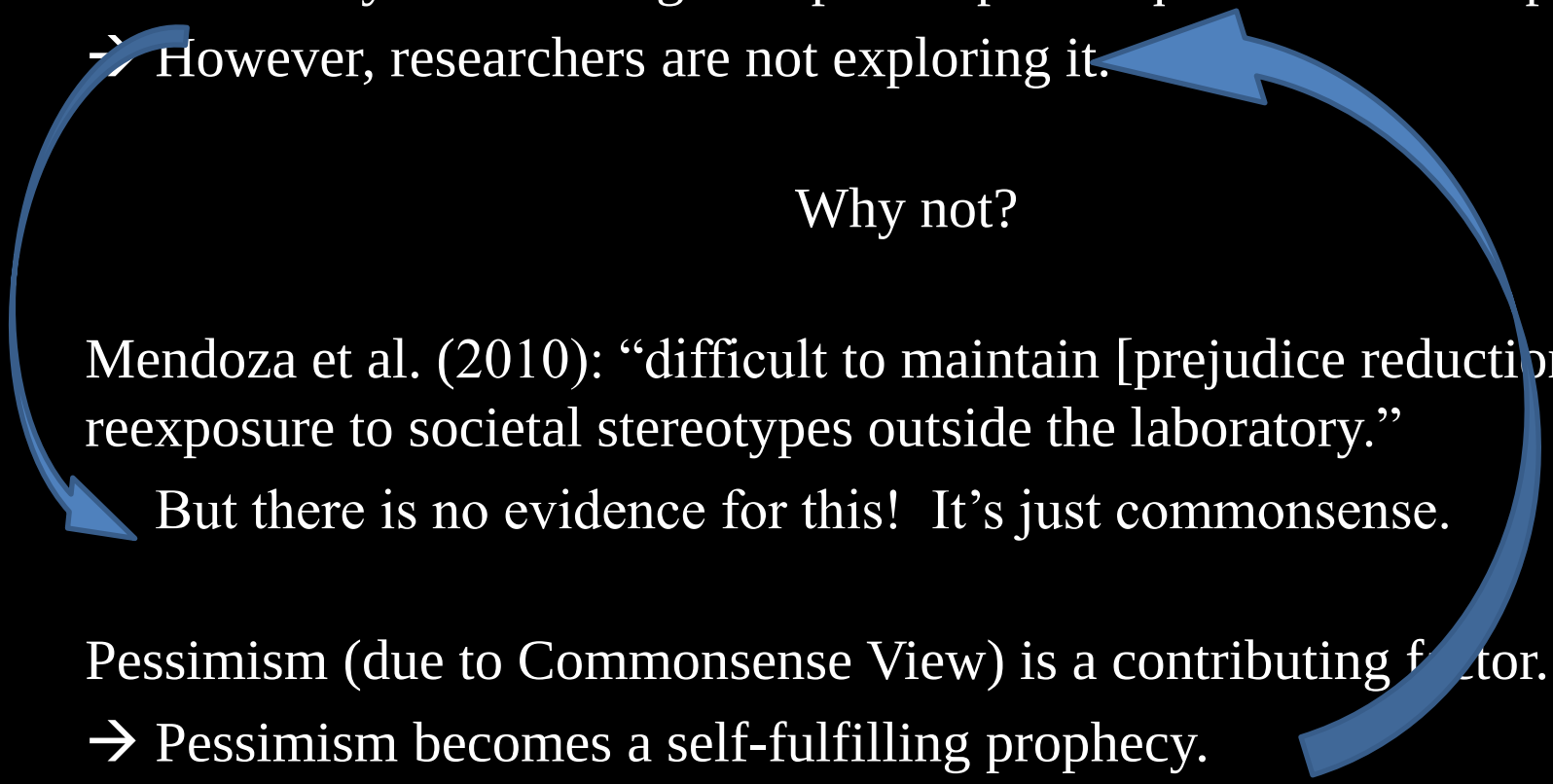
Why not?

Mendoza et al. (2010): "difficult to maintain [prejudice reduction] upon reexposure to societal stereotypes outside the laboratory."

But there is no evidence for this! It's just commonsense.

Pessimism (due to Commonsense View) is a contributing factor.

→ Pessimism becomes a self-fulfilling prophecy.



# Suppose debiasing is not permanent...

How long must debiasing effects last in order to be worthwhile?

Suppose debiasing worked like dental cleanings,
and it was best to debias ourselves once or twice a year.

→ Too much to ask?

Would it be a waste of time to debias ourselves once in a while
even if we didn't re-up as often as recommended?

(If you don't see the dentist often enough, should you never go at all??)

→ Concerns about practical UNFEASIBILITY

# Reasons for Pessimism: Unfeasibility Worry

Suppose these debiasing procedures are reasonably effective.

Are they too time-consuming and resource-intensive to be practically feasible?

Stewart and Payne (2008) write: Kawakami's "extensive training" requires "many, many repetitions to learn nonstereotypical responses."

Olson and Fazio (2006):

soning (Devine & Monteith, 1999). Indeed, it was only after a laborious 480-trial procedure that any change was observed in Kawakami et al.'s (2000) research. Yet, the

# Many, many, laborious repetitions?

Debiasing: significant effects after 200 trials (~20 minutes).

At most, participants work through 480 trials (~45 minutes).

(Alcohol-avoidance: 4 sessions, 15 minutes each)

15-45 minutes is nothing!

Widespread belief that prejudice is too deeply ingrained to uproot in a feasible way → undermined by these very findings.

Widespread belief that debiasing is unfeasible
→ may, ironically, be explained in part by social and cognitive biases.

# Biases Against Debiasing: The Framing Effect

"Hundreds" of repetitions to reduce a bias sounds like a lot.

Compare 45m to time and resources we already devote to prejudice reduction initiatives (many of which are untested).

45m is less than many spend per day on exercise and honing skills.

- Students spend 135 hours a year learning foreign languages.
- Children spend 4 hours a day watching television.

Nobody can spare a single afternoon to try out a prejudice reduction strategy that actually has empirical support?

# Subliminal Debiasing?

Suppose the training can be done subliminally.

→ debias ourselves while engaged in other tasks,
e.g., by “liking” things on social media or surfing the internet.

(Empirical research should explore whether video games
or other software can debias us subliminally.)

It wouldn’t matter how long debiasing took,

or even how durable the effects were.

But subliminal debiasing raises worries of its own!

# Biases Against Debiasing: Creepy Connotations & Associations

Talking seriously about counterconditioning sounds like "thought police" and brainwashing.

Katherine Spencer (p.c.): "Nobody wants to be responsible for the research that is the psychological version of the atomic bomb."

# The Creepiness Worry

We don't find it creepy to memorize flashcards, work through problem sets, practice sports drills and musical scales.

- We're averse to their tedium, not to their creepiness.

Debiasing is not about "implanting" alien beliefs in our minds.

- We're trying to fight back against the alien attitudes that we absorb from racist and sexist environments.

- The point of debiasing is to help us better live up to and embody the commitments we already have (or at least claim to).

# Big-Business Brainwashing

Bryan Gibson (2008) *Journal of Consumer Research*

- subtle procedure changed implicit preferences for Coke vs. Pepsi.
- Gibson: findings have important consequences for product placement.


Pepsi

"SOCIAL SECURITY ISN'T A PONZI SCHEME. IT'S NOT BANKRUPTING US. IT'S NOT AN OUTRAGE. IT IS WORKING."

– Rachel Maddow

LEAN FORWARD

msnbc

RACHEL MADDOW, msnbc host

# An Internal Tension?

## Creepiness Worry vs. Relearning Worry

Relearning Worry: debiasing is pointless because we'll relearn biases.

Creepiness Worry: debiasing is scary because it's like brainwashing.

Relearning presumably occurs through brainwashing,

A Puzzle: we're inured to powerful forces brainwashing us all the time, but feel uneasy about resisting these forces by debiasing ourselves

Why would we let big business have a monopoly on brainwashing!

# Concluding Remarks on Individualism Worry

Cannot assume that restructuring our institutions will have debiasing effects.

→ Diversity initiatives can lead to increased discrimination

→ Without attention to individuals' motivations and interpretations, interventions can backfire, heightening prejudice and discrimination.

Nevertheless, I support many profound structural interventions.

→ A Model of Mutual Reinforcement

Restructuring institutions: "environmental scaffolding" to counteract discrimination and combat bias

Debiasing procedures: "psychological scaffolding" to implement institutional changes without amplifying hostility.

# Thanks!


Dasgupta, N. (2013). Implicit attitudes and beliefs adapt to situations: A decade of research on the malleability of implicit prejudice, stereotypes, and the self-concept. *Advances in Experimental Social Psychology*, *47*, 233-279.

Forbes, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of personality and social psychology* *99*(5), 740.

Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the Implicit Association Test. *Journal of Consumer Research*, *35*(1), 178-188.

Kawakami, K., Dovidio, J.F., and van Kamp, S. 2007: The Impact of Counterstereotypic Training and Related Correction Processes on the Application of Stereotypes. *Group Processes and Intergroup Relations* 10 (2), 139-156.

Kawakami, K., Phills, C.E., Steele, J.R., and Dovidio, J.F. 2007: (Close) Distance Makes the Heart Grow Fonder: Improving Implicit Racial Attitudes and Interracial Interactions Through Approach Behaviors. *Journal of Personality and Social Psychology*, 92(6), 957–971.

Moskowitz, G.B. 2010: On the Control Over Stereotype Activation and Stereotype Inhibition.

*Social and Personality Psychology Compass* 4 (2), 140-158.

Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, *22*(4), 490-497.

# Debiasing in Daily Life

Devine et al. (2012)

5 strategies to employ in daily life :

- stereotype replacement,
- imagine a counterstereotypical exemplar,
- focus on “individuating” rather than “group-based” features,
- take the perspective of a stereotyped group member,
- and increase opportunities for positive social contact.

Reductions of implicit bias lasted at least 8 weeks.

Dasgupta & Asgari (2004): undergrad women who took multiple classes with female math and science professors showed less gender bias after one year.

(Presumably this worked despite the fact that participants were bombarded with stereotypes in the media of women as nurturing and men as assertive)

# Context-Specificity Worry

Background: an image of a black man in a prison might elicit negative responses if he is dressed like a prisoner, but positive responses if he is dressed like a lawyer (Barden et al. 2004).

Subtyping: pick up on distinctive features that mark off a subset of members from the larger group, but leave your default impression of the group unchanged.

Rydell & Gawronski (2009)
"I like you, I like you not:
Understanding the formation of context-dependent automatic attitudes"

Bob volunteers at an orphanage.

Rydell & Gawronski (2009)
"I like you, I like you not:
Understanding the formation of context-dependent automatic attitudes"

Bob steals from the orphanage.

Rydell & Gawronski (2009)
"I like you, I like you not: Understanding the formation of context-dependent automatic attitudes"

Attitudes toward Bob against a red background: negative

Attitudes toward Bob against blue or novel colors: positive

- Future encounters (against novel colors) reflect first impression.

But there's reason to think Kawakami's debiasing, unlike Gawronski's, will generalize.

Studies demonstrate how training in one "context" influences behavior in another "context"

- Retraining stereotypes led to changes in prejudice (novel stimuli)
- Subliminal approach training influenced IAT and social behavior

# Distinctive Features of Kawakami Debiasing?

Kawakami's studies involve meaningful actions

- Not just passively taking in information (like Banaji's screen saver)
- Embodied, meaningful performances

Wennekers (2013): responding only 50% of the time had no significant effect.

- Consistency in responses may be crucial (Cf. Olson and Fazio 2006, 431).

(Such consistency is rare in the "real world" - we cannot expect to approach or have positive interactions with every member of a social group. A reason to favor genuine debiasing procedures?)

(Contrast being exposed to pervasive religious messages on TV and billboards with *being in a cult*)

# Environmental cues in the classroom

# Environmental cues in the classroom

# “Sci-fi nerd” cues and computer science (Cheryan et al. 2009, 2011)

Comp-sci classrooms with sci-fi cues:

- reduce undergraduate women’s interest and expected success in computer science,
- but have no effect on men.

Replacing the stereotypical objects with neutral ones increased women’s interest.

Cheryan: environments “influence students’ sense of ambient belonging… or feeling of fit in an environment.”

# The Case of MIT

In 1999, MIT found it had been allocating much less lab space to women than men in faculty.

- President Charles Vest: “I have always believed that contemporary gender discrimination within universities is part reality and part perception... I now understand that reality is by far the greater part of the balance.”

The administration took responsibility... and made “stunning” progress.

- In 12 years,
  number of women faculty doubled,
  more women in leadership positions, and almost
  no remaining gender differences in salary, lab space, teaching loads.

# And yet...

Unintended Consequences at MIT (2011):

- perception that women at MIT have “unfair” advantage
  - but increase came from broadening job searches, not lowering standards or using “diversity reasons”

- every committee must include women, who are still in the minority among faculty
  - they may lose up to 50% of research time on committee work

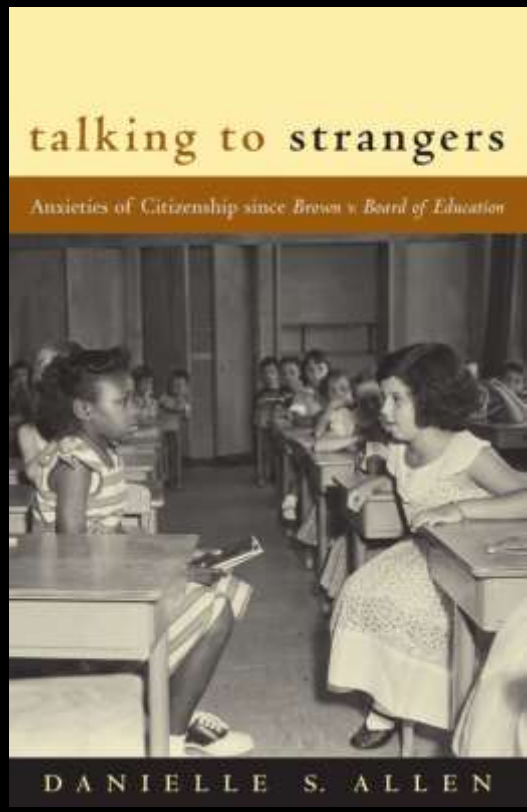
- women must still battle expectations for behavior that is “neither too aggressive nor too soft”

- and childcare is still perceived as an issue for women.

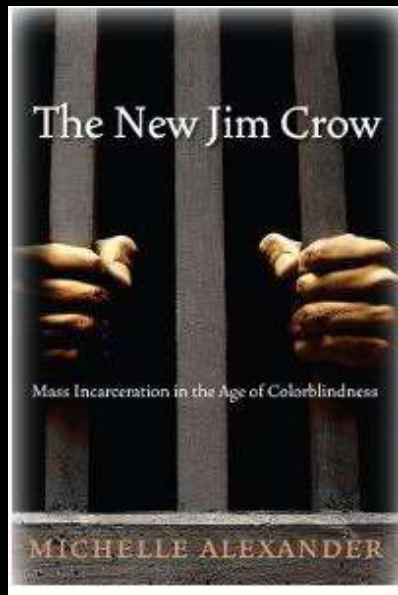
# ... to the many theorists and activists who think our biased habits contribute to injustice

Danielle Allen (2004):
Despite advances of *Brown v. Board of Education* and Civil Rights Act,
"we haven't yet managed to develop for this era new modes of citizenship to supplant domination and acquiescence.
At this juncture, our habits of citizenship
need reconstitution more than our laws" (175).

"our best-taught habit of citizenship is 'don't talk to strangers'" (101).



# Anti-Racist Reflective Commitments but Unreflective Racial Biases




Michelle Alexander (2010): "a clear majority of Americans were telling pollsters in the early 1980s... that they opposed race discrimination in nearly all its forms... there is no reason to believe that most of them were lying...
most Americans... had come to reject segregationist thinking and values, and not only did not want to be thought of as racist but did not want to *be* racist…" (203)

"Race plays a major role... in the current system, but not because of what is commonly understood as old-fashioned, hostile bigotry. This system of control depends far more on *racial indifference* (defined as a lack of compassion and caring about race and racial groups) than racial hostility…" (198)

# Suppose Allen and Alexander are right...

Most Americans sincerely reject racism and sexism.

But implicit racism and sexism persist,
and contribute to systemic discrimination and inequality.

→ Might Kawakami's debiasing procedures provide a useful, complementary tool for changing these problematic habits?

We know that cultivating good habits is necessary for a good life.

→ Self-betterment requires practice…

... going through the motions: flashcards, problem sets, sports drills…

→ Why should prejudice reduction be any different?

→ Debiasing procedures are <u>the basics</u>: simple forms of practice that help us cultivate habits to navigate "real-world" complexity.

# Aristotle, *Nicomachean Ethics*

It is well said, then, that it is by doing just acts that the just man is produced, and by doing temperate acts the temperate man; without doing these no one would have even a prospect of becoming good.

But most people do not do these, but take refuge in theory and think they are being philosophers and will become good in this way, behaving somewhat like patients who listen attentively to their doctors, but do none of the things they are ordered to do.

As the latter will not be made well in body by such a course of treatment, the former will not be made well in soul by such a course of philosophy.

# Cottage Industry Devoted to "Cognitive Biases and..."

Commonplace to speculate about the role of cognitive biases in, e.g., the widespread indifference or failure to act in response to climate change and global poverty and hunger.

- Failure to appreciate how we as individuals are implicated in large-scale problems
- Bystander effect
- Felt sense of "distance" between ourselves and those physically far away or socially removed (or non-human animals).

→ These are exactly the sorts of biases that approach training might help us overcome!

# The “Brainless” Worry?

Maybe these interventions just seem brainless or ridiculous.

A fantasy that the hard problems in our lives should be overcome by a deep, cathartic experience?

- Miranda Fricker (p.c.): “we assume vices like racial bias run *deep*. And this is connected with the comforting idea that virtues run deep in us too... [This] is threatening to the idea that we are morally deep beings.”

In the context of discrimination, this might manifest in the conviction that we need to really understand Marx or Foucault or MacKinnon before we get serious about prejudice reduction...

# The Individualism Critique

Dixon et al. (2012)
"Beyond prejudice:
Are negative evaluations
the problem and is getting us
to like one another more
the solution?"

### 4.2. Reconciling prejudice reduction and collective action models of social change?

The most important question that our article has left hanging is this: What are the prospects of reconciling a prejudice reduction model of social change, designed to help people get along better, with a collective action model of change, designed to ignite struggles to achieve social justice? There are a number of possible positions in this debate. One pole of the argument might assert that the two forms of social change are fundamentally complementary – that is, that getting people to like one another more will ultimately lead to social justice in a deeper sense. The other pole might assert that the two forms of social change are fundamentally incommensurable and that the drive for prejudice reduction has for too long marginalized, if not obstructed, more pressing concerns about core distributive justice (e.g., justice based on the fair distribution of resources such as wealth, jobs, and health). As readers will have gathered, we sympathize with the latter position, particularly when applied to the problem of improving intergroup relations in societies characterized by long-standing, systemic discrimination.[7] To conclude, we revisit the question of whether or not the two models of social change can be reconciled with the goal of opening up a wider dialogue.