Are IAT Effects an artifact of test design? Testing the impact of a rest period before reversing response assignments on the Implicit Association Test

Michael McCarthy (Michael.McCarthy@student.ufv.ca), David Hughes, & Dr. Sven Van de Wetering

---

Outline

- An overview of some concerning facts about implicit cognition research, and a possible remedy
- An application of the remedy—the present experiment
- Question and answer period

Some Facts…

- The **belief** that *implicit biases/attitudes/prejudices/etc.* exist somewhere in the brain unconsciously—causing all manner of ills towards certain groups—has become canonized in our culture (see Mitchell & Tetlock, 2017)
  - In government (e.g., Comey, 2015), business and technology (e.g., Google, 2013), and academia (e.g., Greenwald, Banaji, & Nosek, 2015)
- This belief is based solely upon "evidence" from so-called implicit (read: indirect) measures such as the **Implicit Association Test** (IAT; Greenwald, McGhee, & Schwarz, 1998)
  - "In studying implicit cognition, indirect measures are theoretically essential" (Greenwald & Banaji, 1995, p. 5)
  - **The IAT is claimed to be a diagnostic measure:** "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test" (Greenwald, et al., 1998)

Correction:
Some Concerning Facts

- The dual-process typology underlying the theory of implicit cognition is a *probably untestable meme with 30 years of available counter-evidence* (see Melnikoff & Bargh, 2018)

- The belief that the IAT is a diagnostic measure of anything, let alone attitudes or prejudice, is another *probably untestable meme with 20 years of available counter-evidence*

  - (see Blanton & Jaccard, 2008; Fiedler, Messner, & Bluemke, 2006; Mitchell & Tetlock, 2017; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013, 2015; or email McCarthy for a manuscript synthesizing this information)

- These belief systems are so widespread that they have become a self-perpetuating meme machine (see Mitchell & Tetlock, 2017)

A Possible Remedy

- The field of implicit cognition *desperately* needs a paradigm shift
- Fortunately, there is a huge empirical gap in this field to give room for a paradigm shift:
  - Research which treats the IAT (and other similar implicit measures) as a phenomenon unto itself (i.e., as a dependent measure to be studied experimentally), or
  - Research which manipulates the design of the IAT to study the design-performance relationship
- This shift will open the door to research that can help identify (rather than assume) the cognitive mechanisms that underlie performance on the IAT
  - Thus, this line of research would ultimately help evaluate what the IAT's empirical value *actually is* (rather than assuming its value without adequate evidence)

The Present Experiment

- Design
  - 2-groups (control, experimental)
  - Tested the effect adding a 15 minute rest period before category-switching would have on performance on the IAT
    - During the rest period participants completed a moderate or hard difficulty paper and pencil sudoku puzzle (participants chose which puzzle they wanted to do)

The Present Experiment

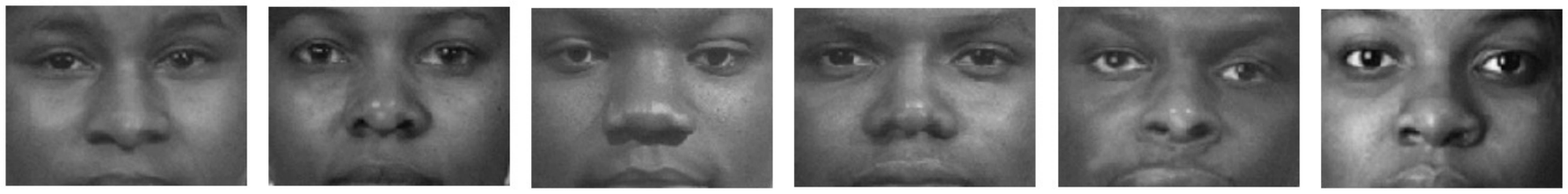
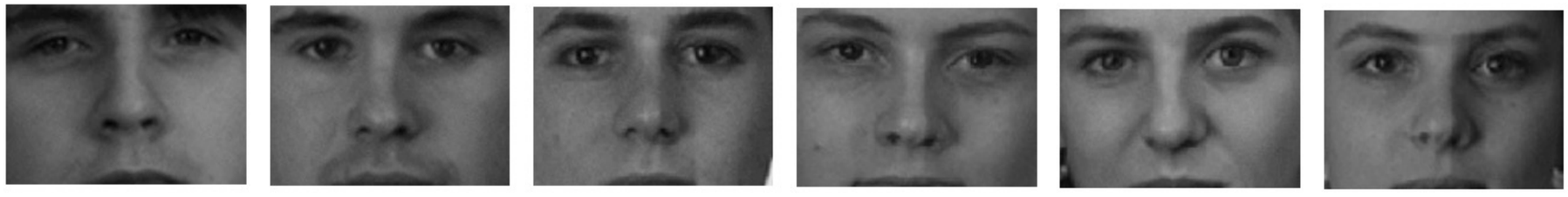
- Design Rationale
  - **Adding the rest period should decrease or eliminate interference of the earlier blocks on the later blocks**, so responding to the IAT should be easier
    - One of the typical IAT findings is an order effect (e.g., Greenwald et al., 1998), where completing compatible (vs. incompatible) blocks first produces larger IAT Effects
    - From personal experience, the test becomes quite confusing in the final blocks, as the response instructions get “jumbled together” as you progress through the blocks
  - **If there are differences between the conditions we can demonstrate that the magnitude of IAT scores are affected by procedural variables**, rather than by things such as attitudes/prejudices/etc.
    - Playing sudoku for 15 minutes *should not* make someone more positively or negatively prejudiced towards a group
    - If small changes to the IAT’s design create large differences in outcome, it demonstrates how sensitive performance on the test is to extraneous influences

The Present Experiment

- Participants
  - 88 students (52 female, 36 male; Age, Mean = 19.9) from the University of the Fraser Valley participated in exchange for 1% course credit and to enter a raffle for $100
    - **Note:** ~39 participants per condition needed for power on the IAT (see Greenwald et al., 2003)
- Participants were tested in groups of up to 24 at a time, and were **randomly assigned to a condition using a double-blind procedure**
- Data for all participants was included for the main analysis or interpretation
- Participants' **data was completely anonymous**, and **all aspects of the experiment were delivered via the computer**

IAT Stimuli

- *Note:* Stimuli taken from the Project Implicit website (Nosek et al., 2007)

Category	Items
Good	Fantastic, Cheerful, Love, Fabulous, Glorious, Spectacular, Celebrate, Enjoy
Bad	Detest, Failure, Rotten, Gross, Dirty, Selfish, Humiliate
Black	
White	

The IAT

- Example for those unfamiliar with the IAT
- *Note:* Inter-trial period = 250 ms

Press "E" for
Good
or
Black

Press "I" for
Bad
or
White

Part 3 of 7

Use the E key for Black faces and Good words.
Use the I key for White faces and Bad words.
Each item belongs to only one category.

If you make a mistake, an X will appear. Press the other key to continue.
Go as fast as you can while being accurate.

Press the space bar when you are ready to start.

Procedure	Left Category	Right Category	Congruency
Block 1	White Faces	Black Faces	
Block 2	Good Words	Bad Words	
Block 3	White + Good	Black + Bad	Congruent
Block 4	White + Good	Black + Bad	Congruent
Rest Block	Half the participants went directly to Block 5; the other half completed sudoku puzzles for 15 minutes		
Block 5	Black Faces	White Faces	
Block 6	Good Words	Bad Words	
Block 7	Black + Good	White + Bad	Incongruent
Block 8	Black + Good	White + Bad	Incongruent
Explicit Measures	(1) Suspicion Probes; (2) Race Preference Likert Scale & Race Feeling Thermometers; (3) Basic Demographic Qs		

Note 1: Blocks highlighted in red are not part of the typical IAT procedure and are unique to the present experiment. Congruency order was counterbalanced.

Note 2: The entire experiment was programmed and delivered to participants via E-Prime (Psychology Software Tools, Inc., 2016)

Results

- Data analysis was conducted in RStudio (RStudio Team, 2015) to ensure all analyses were reproducible

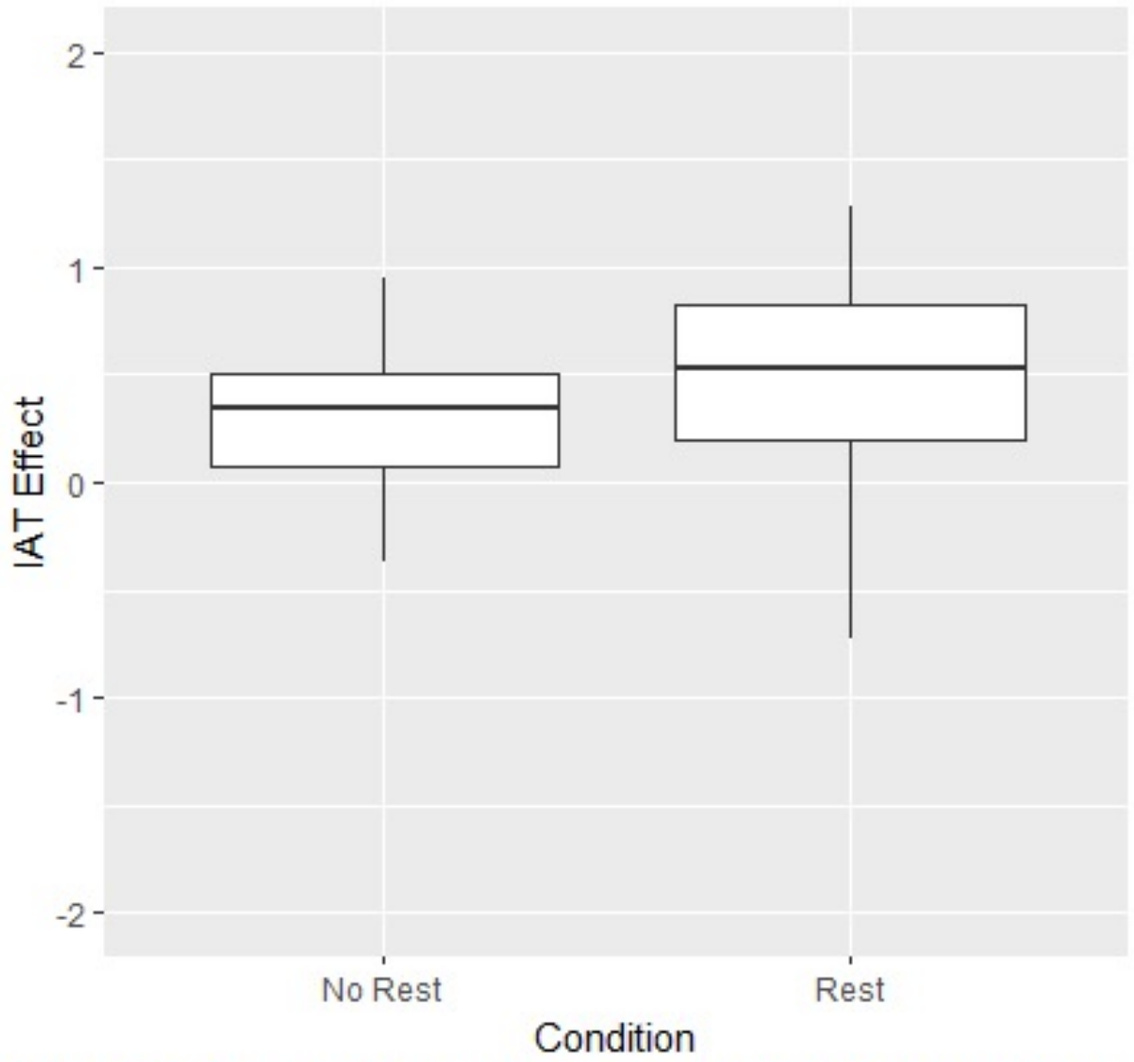
- IAT performance was scored using the IAT scoring algorithm recommended by Greenwald and colleagues (2003)

- **Note 1:** IAT scores are calculated as the *difference in response time and accuracy between the congruent and incongruent blocks* (it's actually slightly more complicated, but this is the basic calculation)

- **Note 2:** Positive IAT scores indicated faster and more accurate responses when the categories *White + Good* and *Black + Bad* were paired together relative to when the categories *White + Bad* and *Black + Good* were paired together

Was There an Effect?

Comparison of IAT effect scores between conditions



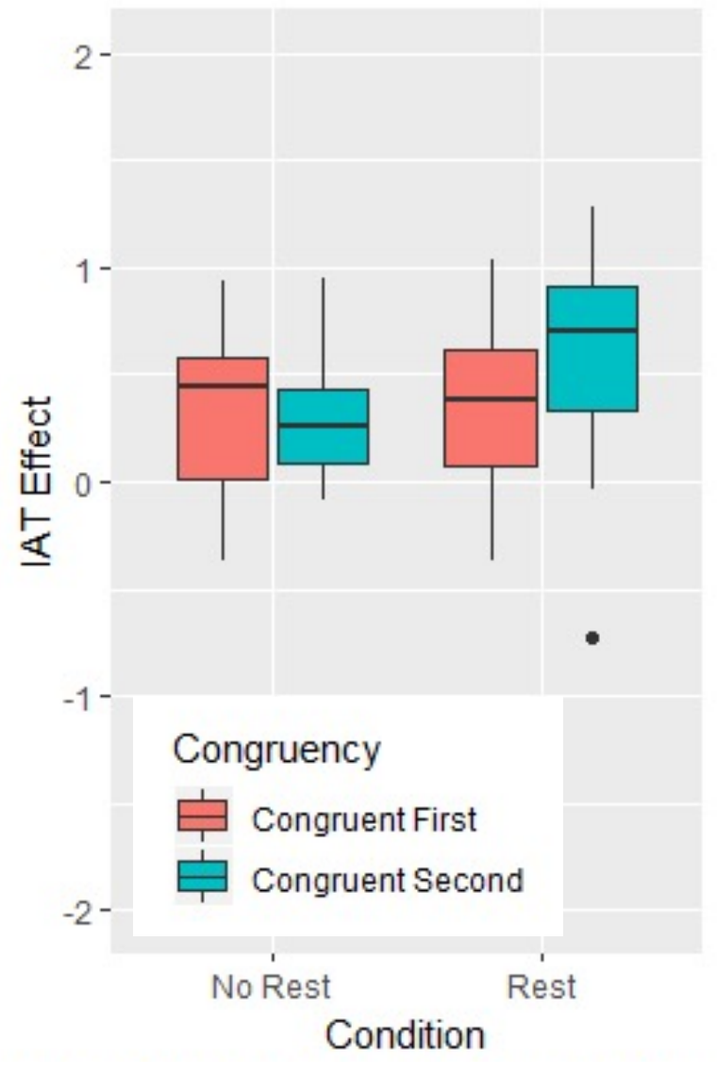
- An independent-samples t-test was conducted to compare IAT effect scores in the *no rest* and *rest* conditions

- There was a **significant difference** in the scores for the no rest (M= 0.33, SD= 0.35) and rest (M= 0.50, SD= 0.43) conditions; $t(86)= -2.04$, $p = .04$, $d = 0.44$

- The rest period (or possibly the sudoku itself) caused a performance difference on the IAT

Comparing Congruency Orders

Comparison of IAT effect scores between conditions and congruency orders

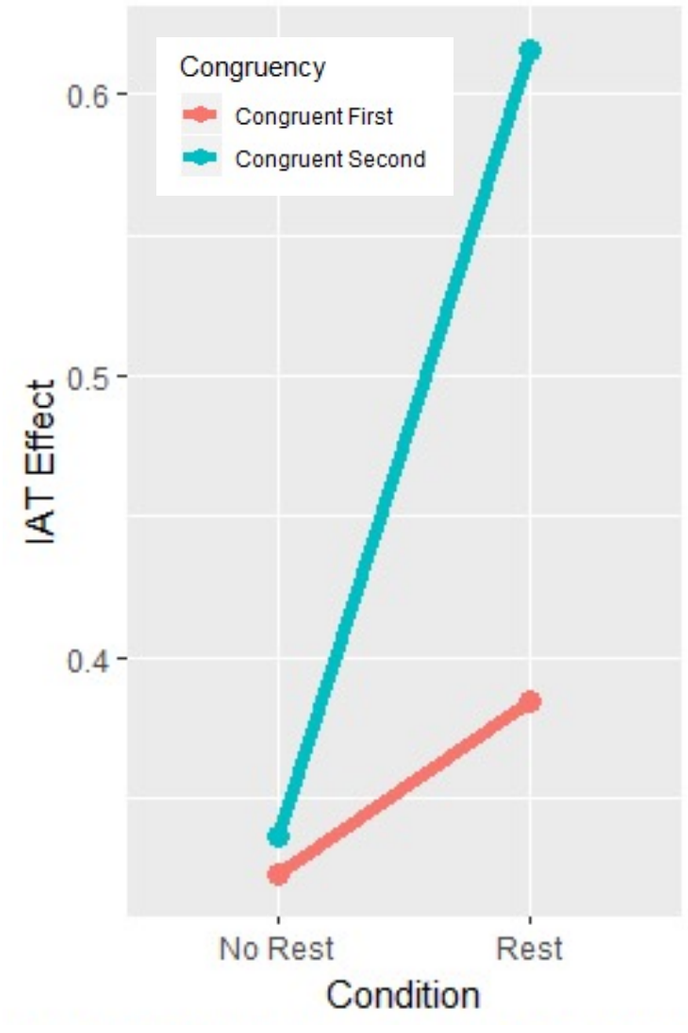


- **Note:** If the outlier from the rest congruent second condition is removed, the results from the previous t-test become:

- no rest (M= 0.33, SD= 0.35), rest (M= 0.53, SD= 0.39); $t(85) = -2.51$, $p = .01$, $d = 0.54$

Was There an Interaction?

Comparison of IAT effect score means between conditions and congruencies



- A 2 (condition) X 2 (congruency order) ANOVA was conducted to examine a possible interaction effect

- This analysis revealed **no significant interaction** between condition and congruency order, $F(1, 84) = 1.73$, $p = .19$, $\eta_p^2 = 0.020$

  - With the aforementioned outlier removed, $F(1, 83) = 3.32$, $p = .07$, $\eta_p^2 = 0.039$

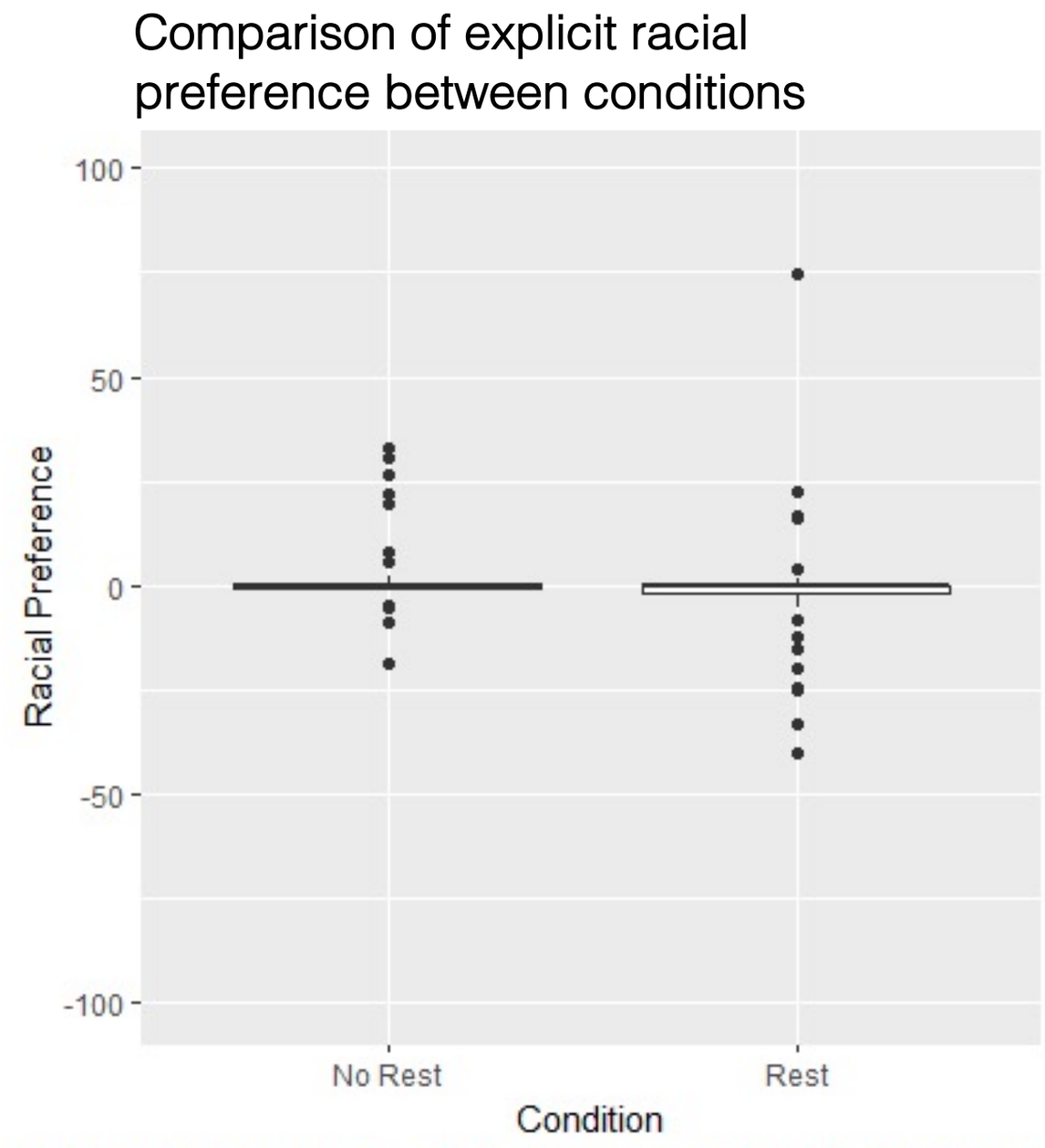
Were the Groups Unequal?

Latencies (ms)	Congruent First		Congruent Second	
	No Rest	Rest	No Rest	Rest
Block 3	859	859	1074	1078
Block 4	709	710	897	853
Block 7	1010	↓1087	846	↑727
Block 8	861	↓ 938	797	↑659

- The significant differences between conditions were caused mainly by different response speed and accuracy on Block 7 & 8

- The differences were 100 (+/- 40) milliseconds, and **there was a performance dissociation in the rest condition** (relative to no rest) where *congruent first performed slower*, and *congruent second performed faster*

Explicit Race Preference



- Data from the race preference likert and race feeling thermometers were combined by averaging participants' responses to both measures
- Clearly, **participants generally did not have any overall race preference** (counter to what the IAT scores would indicate, if we are to believe them)
  - We can rule this out as a possible cause for between-group differences

Discussion

- Key points
  - Procedural changes to the IAT that are unrelated to attitudes/prejudices/etc. can cause significant changes in performance on the IAT
  - Two relevant interpretations of this effect:
    - (1) Implicit attitudes *as measured by the IAT* are so unstable that they can be significantly changed by the most innocuous of things (i.e., playing sudoku can make people unconsciously associate white people with goodness…)
    - (2) The design of the IAT directly influences performance on the IAT, thus, the magnitude of IAT effect scores are an artifact of the test's design

Possible Causes?

- The performance dissociation between congruency orders in the rest condition suggests that **the results *were not* caused by fatigue/exhaustion/boredom/etc**.
  - Otherwise we would have expected the same pattern in both conditions (e.g., extremely rapid and inaccurate responses so as to finish faster)

- Participants would not be aware of the pairing reversal unless they had completed an IAT before; the performance dissociation could have been spuriously caused by familiarity with the IAT (*this is 100% speculative*)

Possible Causes? Cont'

- The rest period could have allowed participants to complete the remaining blocks "fresh" (similar to why there is typically an IAT order-effect) and with practice, causing performance gains

- This would explain the White + Good second condition's performance, but would not explain why the Black + Good second condition's performance was worse relative to control

  - Perhaps this suggests a dissociation in the psychological processes (and/or neural pathways) used to respond to the different pairings on the IAT?

Limitations

- The manipulation clearly had an effect, however, interpreting this effect is difficult (particularly because no one knows what the IAT actually measures)
  - Direct and conceptual replications needed to refine explanations
- The experimental paradigm of introducing the IV before the pairing reversal on the IAT seems promising though
  - It enables the comparison of *a priori* and *a posteriori* group differences in performance
  - To our awareness, no other experiments have used this paradigm, so **this experiment is best considered exploratory**

Conclusions

- Despite its popularity, the empirical value of the IAT and theories of implicit cognition are currently based up empirically unjustifiable claims and assumptions

- Scholarship treating the IAT (and thus implicit cognition) as a phenomenon unto itself is needed to shift the field's paradigms towards experimental research based on the scientific method

- This paradigm shift will improve scholarship in the field and aid our understanding of all the psychological processes/phenomenon underlying performance on the IAT (everyone wins!)

Questions, Comments, Suggestions?

P. S. If you are interested in this topic, check out McCarthy's other project during Poster Session II (12:45-2:15)!

References

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. Psychological Bulletin, 11, 1227-1274.

Blanton, H., & Jaccard, J. (2008). Unconscious racism: A concept in pursuit of a measure. Annual Review of Sociology, 34, 277-297.

Comey, J. B. (2015). Hard Truths: Law Enforcement and Race. Retrieved from https://www.fbi.gov/news/speeches/hard-truths-law-enforcement-and-race

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I", the "A", and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). European Review of Social Psychology, 17, 74-147.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4–27.

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.

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*, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003) Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal Of Personality And Social Psychology*, *85*(2), 197-216.

Google (2013). Unconscious Bias @ Work (video series). Retrieved from https://rework.withgoogle.com/guides/unbiasing-raise-awareness/steps/introduction/

Kang, J., Bennett, M., Carbado, D., Casey, P., Dasgupta, N., Faigman, D.,…Mnookin, J. (2012). Implicit bias in the courtroom. *UCLA Law Review, 59*, 1124-1186.

Melnikoff, D. E., & Bargh, J. A. (2018). The Mythical Number Two. Trends in Cognitive Sciences, 22(4), 280–293.

Mitchell, G., & Tetlock, P. E. (2017). Popularity as a poor proxy for utility: The case of implicit prejudice. In S. O. Lillienfeld & I. D. Waldman (Ed.), Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions.

References

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ratliff (Ranganath), K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology,18*, 36-88. https://www.projectimplicit.net/stimuli.html

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. Journal of Personality and Social Psychology, 105, 171–192. doi:10.1037/a0032734

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. Journal of Personality and Social Psychology, 108, 562–571. doi:10.1037/pspa0000023

Psychology Software Tools, Inc. (2016). E-Prime 3.0. Retrieved from http://www.pstnet.com.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. Retrieved from: *http://www.rstudio.com/*.