DATA PAPER

Data from the Race Implicit Association Test on the Project Implicit Demo Website

Kaiyuan Xu,[1] Brian Nosek[2] and Anthony G. Greenwald[1]

[1] University of Washington, Seattle, Washington, United States of America

[2] University of Virginia and Center for Open Science, Charlottesville, Virginia, United States of America

This data archive includes Race Implicit Association Test (IAT) scores of 2,355,303 Internet volunteers who completed educational/demonstration versions of the Race IAT at https://implicit.harvard.edu from 2002 to 2012. Data in this archive can be downloaded for all years, either separately by year or in a single file. Codebooks, indicating the variable labels and value labels, and changes of variables over years, are available for both individual-year data sets and the entire data set. Participation in the (still on-going) Race IAT "study" at the Project Implicit (PI) demonstration site includes completion of the Race IAT along with demographic questions, self-report measures of racial attitude, and various additional measures received by a portion of the participants. These data allow analyses involving changes in responding over time and interrelations among IAT and self-report measures of race attitudes, as well as the association of each of these with demographics. This archive is available at http://osf.io/project/52qxL/.

Dataset
The Data described in this paper is available from the Open Science Framework: https://osf.io/52qxl/ [1].

(1) Overview
Context
Temporal coverage
23 December 2002 to 31 December 2012 (to be updated after end of further calendar years).

Background
The Implicit Association Test (IAT) was introduced in 1998 [2]. The IAT is a dual categorization task in which, in the critical blocks, participants alternate between two categorization tasks involving four total target categories (e.g. flowers vs. insects; pleasant vs. unpleasant), but respond using only two response keys. The logic of the task is that if cognitively associated categories share a response key, responses will be facilitated, resulting in faster responding and less errors. For example, when categorizing pleasant words and flower images with one key, and unpleasant words and insects images with the other key, responding will generally be faster and have lower error rate than responding when the opposite pairings needs to be made (flower and unpleasant vs. insect and pleasant). The IAT is now the most well-validated measure of implicit attitude [3, 4], and the Race IAT (Black vs. White people; Good vs. Bad) is the most widely used of numerous IATs that have been used in research.

The IAT demonstration website at https://implicit.harvard.edu is maintained by Project Implicit (PI), which is a non-profit organization founded in 1998 by scientists at University of Washington (Anthony G. Greenwald) and Yale University (Mahzarin R. Banaji, now at Harvard University, and Brian A. Nosek, now at University of Virginia and the Center for Open Science). PI supports a collaborative network of researchers interested in basic and applied research concerning implicit social cognition and hosts data collection for many online research projects worldwide, in addition to providing demonstration IATs, such as the Race IAT, for educational uses.

(2) Methods
Sample
From 2002 to 2012, 3,664,508 sessions were initiated online by participants worldwide; of these 1,870,921 sessions were completed from initial consent to concluding debriefing, and 2,355,303 completed all 7 blocks of the Race IAT. Of sessions with demographic reports, representation within the total sample was: 57.2% female, 42.8% male; 86.1% US citizen, 13.9% citizens of other countries; mean age was 27.23. Prior to 28 Sep 2006, "Hispanic" was treated as a race category. After 28 Sep 2006, it was treated as ethnicity in the way that the U.S. Office of Management and Budget had started to track ethnicity separately from race. Prior to 28 Sep 2005, responses to the race question showed participation by 70.2% White, 9.5% Black, 5.6% Hispanic, 5.7% Asian or Pacific Islander, 0.9% Native

American, 5% multi-racial, and 3.1% "other". After 28 Sep 2006, responses to the race question (including no Hispanic option) indicated 70.1% White, 12.3% Black, 5.5% Asian or Pacific Islander, 0.7% Native American, 6.7% multi-racial, and 4.7% "other". The ethnicity question indicated 8.8% Hispanic, 83.6% Non-Hispanic, and 7.6% "unknown".

Materials
The implicit measure was a standard Race IAT [5] with 7 blocks. The 7 blocks are two 20-trial single categorization practice blocks (1: Black vs. White; 2: Good vs. Bad), followed by a 20-trial and then a 40-trial combined categorization block (e.g. 3&4: Black and Good vs. White and Bad), then another single categorization block with 40 trials and category sides switched (5: White vs. Black; it was also set up as a 20-trial block in 2002 and 2003), and finally another two more combined-task blocks of 20 and 40 trials, respectively (e.g., 6&7: White and Good vs. Black and Bad).

The IAT's Good and Bad stimuli were words, and the Black and White stimuli were face images. The Good words were "Joy", "Happy", "Laughter", "Love", "Glorious", "Pleasure", "Peace", and "Wonderful"; Bad words were "Evil", "Agony", "Awful", "Nasty", "Terrible", "Horrible", "Failure", and "Hurt". The face images were grey-scale images, cropped at the top (just above eyebrows), bottom (between the lips), and each side (just outside eyebrows). Each race group (Black and White) had 6 faces, three male and three female (These face images are available at https://osf.io/project/52qxL/node/JRvg8/).

Self-report measures were added, revised, or removed from the study over the years of its use (the codebooks describe these details). Some questions were shown to all participants during the periods of their use, whereas others were selected randomly for administration from various standard questionnaires. A Likert-format racial preference question (relative preference for Black vs. White) and two feeling thermometer questions (warmth of feeling toward Blacks and toward Whites) were shown to all participants. Four situational questions about racial inequality (2005 to 2012), and one question each assessing extraversion and happiness (2004 to 2006) were also shown to all participants. Measures administered randomly, each to approximately 2% of participants, were: Big Five Inventory (44 questions), Balanced Inventory of Desirable Responding (36 questions), Belief in a Just World (6 questions), Bayesian Racism Scale (15 questions), Humanitarianism/Egalitarianism (10 questions), Need for Cognition (18 questions), Need for Cognitive Closure (42 questions), Protestant Ethic (9 questions), Personal Need for Structure (12 questions), Right-Wing Authoritarianism (20 questions), Right-Wing Authoritarianism (short) (15 questions), Social Dominance Orientation (12 questions), and Self-Monitoring (18 questions).

Demographics questions were asked of all participants. Over years, questions asked participants their age, gender, race, citizenship, education level, field of study, occupation, political identity, religion, religiosity, and zip code/postcode. United States zip codes were recoded to U.S. State, County, or Metropolitan Statistical Area (MSA) in the data sets to eliminate the possibility of identifying individual participants. Non-US postcodes were recorded for 21% of all participants, but are not included in the public archive, again to preserve anonymity. Data with zip codes/postcodes are available by request to those that have ethics board approval to view and analyse the data.

Procedures
Those who volunteered to participate had opportunity to complete the Race IAT, the full-sample subset of self-report and demographic questions, and 10 randomly selected self-report questions (since September 2006). After completion, participants received their IAT score and some interpretative information, and were invited to complete some debriefing questions. The session concluded by closing the browser window in which it was run or switching the browser window to a different URL, which could be for a different IAT task at the PI site.

Quality control
Data sets were retrieved from Project Implicit severs at Harvard University in .txt format as "explicit.txt", "iat.txt", "sessions.txt" and "sessiontasks.txt". Data sets were then read into SPSS Windows Version 19, and archived as .sav files. All of those who started a study session appear in the archived data sets, regardless of their completion status. This includes many sessions that did not actually begin because of technical error, pop-up window blockers, or for other reasons. Those who did not fully complete the IAT task do not have any IAT-related variables in the public archive.

Ethical issues
All volunteers received an initial information statement and were asked to signal consent to participate before having an opportunity to provide data. No typical personal identifiers, such as name, or email address were obtained. Participants were free to discontinue participation at any time, as well as to decline to respond to any self-report or demographic question. All were given opportunity to send email indicating difficulties or asking questions.

(3) Dataset description

Object name
Datasets and Codebooks from Race IAT 2002-2012

Data type
Primary data (self-report and demographic responses), and processed data (IAT scores or measures computed from self-report and demographic responses).

Format names and versions
The archived data are in SPSS .sav files created by SPSS Windows Version 19. The codebooks are .xlsx files created by MS Office Excel 2010.

Creation dates
Data were collected from 23 December 2002 to 31 December 2012. The data released on 24 October 2013 were processed in September 2013.

Dataset creators

Kaiyuan Xu retrieved the raw data from the PI database at implicit.harvard.edu, computed IAT scores and IAT-related variables, provided variable labels and value labels for all measures, and created codebooks for each year's data sets. Brian Nosek provided files that had archived procedural changes over the years. Tony Greenwald provided IAT scoring syntax for SPSS, and advised on formatting of datasets and codebooks.

Language

English

License

CC-BY

Embargo

None

Repository location

Open Science Framework: http://osf.io/project/52qxL/ [1].

Publication date

Some of the archived data (relatively small portions of the total) have been partially reported in various prior publications:

- Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). 27 February 2002 [6]
- Nosek, B. A., et al. (2007). 07 April 2008 [7]
- Schmidt, K., & Nosek, B. A. (2010). 18 January 2010 [8]

(4) Reuse potential

This archive can be used to study questions about the correlation between implicit and explicit racial attitude, the effect magnitudes, inter-individual variability, and regional comparisons as in the Nosek et al (2007) paper. Other research questions can be answered by analysing the correlations between IAT score and the 13 self-report questionnaires over racial, political, and other social topics. Given the temporal coverage of the data set, trend analysis can be done to study how the race attitude shifted over years or time of day in different regions among different groups of participants. Also, there are numerous potential uses of the data for investigating questions related or unrelated to the IAT such as attrition, extraction of representative samples from volunteer data, relations among other variables in the dataset, and response latency distributions.

References

1. **Xu, K, Nosek, B A** and **Greenwald, A G** 2014 Race IAT 2002-2013 [dataset]. *Open Science Framework*, Available at: https://osf.io/52qxl/
2. **Greenwald, A G, McGhee, D E** & **Schwartz, J L.** (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480. DOI: http://dx.doi.org/10.1037/0022-3514.74.6.1464
3. **Nosek, B A, Greenwald, A G,** & **Banaji, M R**. (2007). The Implicit Association Test at Age 7: A Methodological and Conceptual Review. In: Bargh, J A, editor. *Social psychology and the unconscious: The automaticity of higher mental processes.* New York: Psychology Press. p. 265-292.
4. **Greenwald, A G, Poehlman, T A, Uhlmann, E L** & **Banaji, M R.** (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17-41. DOI: http://dx.doi.org/10.1037/a0015575
5. **Nosek, B A, Greenwald, A G** & **Banaji, M R.** (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*(2), 166-180. DOI: http://dx.doi.org/10.1177/0146167204271418
6. **Nosek, B A, Greenwald, A G,** & **Banaji, M R.** (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*(1), 101-115. DOI: http://dx.doi.org/10.1037/1089-2699.6.1.101
7. **Nosek, B A, Smyth, F L, Hansen, J J, et al.** (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*(1), 1836-1888. DOI: http://dx.doi.org/10.1080/10463280701489053
8. **Schmidt, K** & **Nosek, B A**. (2010). Implicit (and explicit) racial attitudes barely changed during Barack Obama's presidential campaign and early presidency. *Journal of Experimental Social Psychology, 46*(2), 308-314. DOI: http://dx.doi.org/10.1016/j.jesp.2009.12.003

How to cite this article: Xu, K et al 2014 Data from the Race Implicit Association Test on the Project Implicit Demo Website. *Journal of Open Psychology Data,* 2(1): e3, DOI: http://dx.doi.org/10.5334/jopd.ac

Published: 18 March 2014

]u[The *Journal of Open Psychology Data* is a peer-reviewed open access journal published by Ubiquity Press

OPEN ACCESS