



An Exploration of Relationships Between Perceptual and Cognitive Racial Biases

Nathan C. Fisher, Victoria E. Maskas, Addison D. Sans, Peter D. Kvam, and Brian Odegaard

College of Liberal Arts and Sciences, University of Florida

Dr. Brian Odegaard, Department of Psychology

Abstract

People are often biased in how they evaluate characteristics of individuals of different races. Some of these biases are perceptual: for example, the “race-lightness effect” demonstrates that for grayscale photos at equivalent luminance values, the faces of black individuals are frequently judged to be darker than the faces of white individuals. Other biases are cognitive: the Implicit Association Test (IAT) and direct assessments of racial attitudes demonstrate that people hold both negative implicit and explicit racial prejudices. Despite extensive literature on both topics, no study has explored the relationship between perceptual and cognitive racial biases in the same individuals in a within-subjects design. In this experiment, these relationships were explored in 28 individuals using three tasks: in one task, deemed the Race-Lightness Task (RLT), participants completed a 2-interval forced-choice procedure where they were shown pairs of faces (some with European features, some with African features) and asked to judge whether the second face was lighter or darker than the first. In a second task, individuals completed the IAT to measure their associations between race and positive/negative adjectives. Lastly, subjects answered a series of questions to measure explicit attitudes about different racial groups. While results from the RLT and IAT were uncorrelated, a significant correlation was shown between the IAT and a survey question about systemic racism. These results provide preliminary support for the independence of perception and cognition for racially based tasks, and provide insight into the pervasive nature of implicit and explicit racial prejudice.

Keywords: Perception, Cognition, Implicit Association Test, Racial Bias, Systemic Racism

Introduction

What is the relationship between social attitudes we hold and how we evaluate physical characteristics of other people? Some argue that these two domains are completely dissociated from one another, in that parts of the brain involved in *cognitive* processing are distinct from elements in the brain involved in *perceptual* processing (Firestone and Scholl, 2016). Others argue that cognitive and perceptual processing interact in ways that strongly influence how people perceive the world (Levin, Baker, & Banaji, 2017). Spirited debates surrounding this topic have spurred an important question: what is the relationship between cognitive and

perceptual processes in regard to *racial biases* that are prevalent in individuals? This study aims to investigate that question.

To accomplish this aim, two well-established psychological tests were employed: (1) the Implicit Association Test (IAT) (Greenwald et al., 1998) whereby automatic associations between Black/White individuals and personal characteristics are probed, and (2) the Race-Lightness Task (RLT), where participants assess the luminance of grayscale facial images of individuals with different ethnic features (i.e., Caucasian and African facial features). The IAT measures the strength of associations between concepts (e.g., black people, white people) and evaluations (e.g., good, bad) or stereotypes (athletic, clumsy). In this test, speeded judgments are made to respond to photos of individuals or concepts, with specific key-response mappings (e.g., respond with F when judging black individuals and negative words, and respond with J when judging white individuals and positive words). The main idea is that making a response is easier when closely related items share the same response key.

The Race-Lightness Effect (Travers et al., 2020) is used to assess perceptual racial biases. This task focuses on evaluating the relationship between a face's morphological features (e.g., how "Africanized" a face is) and a subject's perceived lightness or darkness of that particular grayscale face. Pairs of faces are presented to subjects in succession, each face having a specific combination of luminance and morphological features.

Responses to direct questions concerning participants' racial biases tend to correlate with bias shown in IAT data, with directness in this case referring to the degree of explicitness concerning the content and interest of the question (Axt, 2018). In this investigation, correlations were investigated between the IAT, RLT, and explicit measures of racial biases in a small pilot sample at the University of Florida, to determine if any significant relationships would emerge and be worth exploring in a larger sample in the future. While correlation does not necessarily imply causation, finding relationships between any of these constructs would be worthy of additional follow-up studies to further uncover if they are replicable, and if so, how and why they occur.

Methods

The experimental methods in this study were approved by the University of Florida Institutional Review Board and enacted in accordance with the approved guidelines. All experiment protocols were conducted in accordance with the provisions of the Declaration of Helsinki and subjects were provided with informed consent.

Participants

Thirty-seven participants enrolled in this online study. It was coded in JavaScript and utilized plugins from the jsPsych library; real-time data was stored in a Google Sheet. Participants were recruited through the SONA Systems cloud-based participant management software through the University of Florida's Department of Psychology. Each participant was granted points towards course credit in exchange for his or her participation. Out of these thirty-seven enrolled participants, nine participants failed to complete all tasks and thus were excluded from the final analysis. 28 participants were included in the final analysis (eight male, 20 female) with an average age of 18.71 years and standard deviation of 0.9.

Procedure

Participants began the task by enrolling through SONA Systems cloud-based participant management software through the University of Florida's Department of Psychology. Upon starting the study, participants were presented with an online consent form and provided consent by checking a box next to the statement, "If you agree and wish to participate in this study, click here." They were then presented with a photograph to demonstrate and ensure their computer was an arm's length away from them. Following this, the participants were presented with basic demographic questions accounting for sex and age. After these initial questions, participants were presented with two separate tasks: the Implicit Association Test and Race-Lightness Task (whose order in presentation was randomly assigned to participants). The RLT was a replication of a study conducted by Travers, Fairhurst, and DeRoy (2020) on racial bias in facial perception, which investigated the factors that contribute to the perception of African features being perceived as darker than they actually are (Travers et al., 2020). The second task presented was a replication of the Harvard Project's Implicit Association Test (IAT). When the IAT was utilized in this study, the strength of associations between races (white people, black people) and evaluations presented in accordance with these groups (good, bad, etc.) was assessed. The order

in which these two tasks were presented were randomly assigned to participants, as a means to eliminate confounding variables. Finally, the study asked direct questions concerning the participant's self-evaluation of their racial biases.

Task 1: The Race-Lightness Task

Before beginning the Race-Lightness Task, participants were told that they would be shown two faces in succession, with the second face being the "target" face. They would then be asked two questions: first, if the target face was "darker" or "lighter" than the first face; then, how confident they were that their answer was correct, marked by moving a slider along a scale of zero to 100 (the exact numbers were not shown) with the lower end labeled "not certain at all" and the higher end labeled "totally certain." The wording of the instructions and questions were very similar to that used by Fairhurst, Travers, and DeRoy (2020). After reading the instructions, participants were given 20 practice trials and told how accurate their responses were. Then, they began the full task of 144 trials, which consisted of four sections of 36 trials, each split with a break screen.

The faces presented to participants were taken directly from Travers, Fairhurst, and DeRoy (2020). There were 72 images of faces, but each one originated from one of eight unique faces that were computer generated using FaceGen (Singular Inversions, 2016). They were each manipulated to have the same luminance, then duplicated twice to create two distinct groups of images, one displaying images with 90% of the original luminance, the other displaying images with 110% of the original luminance. The faces had varying levels of luminance and morphological differences, both scaled from negative two (darkest/strongest African morphology) to positive two (lightest/strongest European morphology).

In total, five conditions of 'change in luminance' and five conditions of 'change in morphology' were included (combined for a total of 24 conditions; we did not include trials with both faces having the same level of both luminance and morphology). Each trial had one given change in luminance (represented by the luminance value) and one given change in morphology (represented by the morphology value) following a specific set of rules. For example, if the second face presented was two levels lower in a certain domain in either luminance or morphology (e.g. the first face was 110% lightness or morphologically European, second face was 90% lightness or morphologically African), that domain's value was labeled as negative

two. Following suit, this logic was utilized for the four other levels of change in luminance and morphology, creating a range of negative two to positive two.

Task 2: The Implicit Association Test

The second task was a version of the Harvard IAT. It began by presenting a table (see Figure 1) containing four categories (good, bad, black people, and white people) with eight words in the item's sections of the "good" and "bad" categories along with grayscale pictures of eight different faces (taken directly from the Harvard IAT) in the item's sections of the "black people" and "white people" categories (for a total of 16 words and 16 grayscale images of faces). The faces were the same for every participant, but there were two distinct possible sets of words, one of which was randomly assigned to each given participant. The participants were told on this screen how the instructions would change for each part.



Category	Items
Good	Pleasure, Magnificent, Terrific, Joyous, Delightful, Fantastic, Laughing, Love
Bad	Despise, Hurtful, Tragic, Bothersome, Gross, Angry, Rotten, Selfish
Black people	
White people	

Figure 1. IAT Instruction Table. From Project Implicit, by University of Virginia, n.d. (<https://implicit.harvard.edu/>)

After clicking continue on the initial instructions, participants began the first of the seven trials that compose the IAT. Each trial had its own instruction screen. In the first trial, they sorted the faces by pressing "E" to sort black faces to the left and "I" to sort white faces to the right, or vice versa (which race was assigned to the left or right was randomly assigned to each participant, but "E" always assigned the stimulus to the left). If they gave an incorrect input, a red "X" would appear until they gave the correct input. The next trial consisted of the exact same setup, except instead of black and white faces participants sorted good and bad words to the right or left using "E" and "I" respectively (bad words were always sorted to the left and good to the right). The third and fourth trials were the same; participants sorted words and faces (in random

order), maintaining the left or right side used in the previous two trials. This showed how quickly a participant sorted a given race and negative or positive word to one side and compare. The fifth trial was the same as the first trial (sorting faces only) but reversed to which side the faces were sorted. The sixth and seventh trials were the same as the third and fourth other than sorting the faces to a different side.

Task 3: Individual Assessment

After completing both the IAT and Race-Lightness Task, participants were asked to complete a series of direct, explicit, questions to assess their perceptions of their own racial biases towards black people/African Americans and white people/European Americans. They were presented with a series of six questions, but due to a data formatting issue, correct answers were not able to be discerned, so therefore only two out of the six questions were utilized for analysis. It is also important to note that the race of each individual participant in this experiment was not obtained. The questions that were analyzed were: “Which statement best describes you?” followed with seven response choices ranging from “I strongly prefer European Americans to African Americans” (1) to “I strongly prefer African Americans to European Americans” (7) (Axt, 2018); and “Which statement is most accurate concerning the way in which African Americans are presently treated in the United States today?” followed by four response choices: “systemic racism is [1 = extremely prevalent, 2 = moderately prevalent, 3 = slightly prevalent, 4 = not prevalent] in the United States today.”

Modeling and Analysis

IAT Analysis with a Variation of the Diffusion Model

A new computational model was used to analyze data from the IAT. This model uses a variation of the diffusion model and applies it to the IAT (Kvam et al., 2022). It aims to disentangle many variables influencing a given participant’s response times, such as congruence, drift variation, and non-decision time. “Congruence” refers to which race was associated with good or bad words. “Drift variability” refers to the idea that correct responses tend to be faster than incorrect responses. “Non-decision time” refers to the component of response times not associated with evidence accumulation and decision making. Taking these variables into account, this model provides an “association metric” - a numerical value that illustrates whether

a given participant more easily associated black faces with bad words and white faces with good words (represented by a positive association value), or more easily associated black faces with good words and white faces with bad words (represented by a negative association value).

Task 1: The Race-Lightness Task

In order to investigate the relationship between morphology and luminance in the race-lightness task, an analysis that measured the percentage of participants that responded “darker” across all 24 conditions was employed. Following this, the average “darker” response across all 28 participants was plotted, and the slopes and standard deviations across all 24 data points were found, providing 5 slope values (see Figure 2). To evaluate the influence of these twenty-four conditions on an individual basis, the same plotting measures were utilized again (plotting the percentage of times “darker” was chosen throughout the study) for all 28 participants. Doing so provided five slope values for all 28 participants. The absolute value of these slopes was taken and then averaged, producing 28 individualized metrics that represent the degree in which morphology influences lightness judgment. This metric was then able to represent bias on the race-lightness task and utilized for a series of correlations done with the IAT association measure.

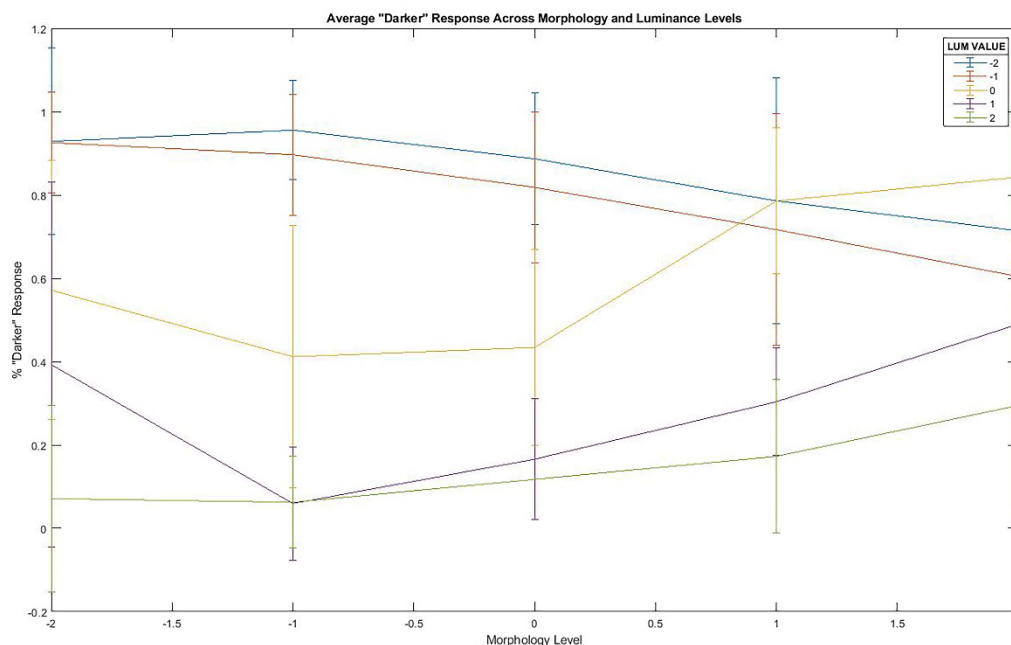


Figure 2. Graph of Average “Darker” Responses Across Morphology and Luminance Levels. Plot of average “darker” response for all 28 participants across all 24 morphology and luminance conditions. The Y axis is representative of the percentage of participants that responded “Darker” when shown the target face in succession to the first face. The X axis represents the degree of change in morphology. Each slope represents the degree of change of luminance. Error bars represent SEM.

Task 2: The Implicit Association Test

To analyze the IAT data in preparation for its comparison with the RLT data, a number of parameters was taken into consideration. Utilizing the metric of “drift variation”, response times for sorting words and images could be captured separately. This established a baseline for the effect that correctness had on a given participant’s response time for words and faces. Thresholds were then calculated - an estimate of the amount of evidence required to trigger a decision response, for each condition: sorting black/white faces only, sorting good/bad words only, congruent (black faces being sorted to the same side as bad words while white faces were sorted to the same side as good words) and incongruent (black faces being sorted to the same side as good words, etc.). Finally, each participant’s overall association was measured; this is a value which quantified to what extent the participant more easily sorted congruent trials (positive

association) or incongruent trials (negative association) and looked at non-decision time (Kvam et al., 2022).

Some examples include understanding that faces, and words are not processed at the same speed, association factors into the decision-making process, and of course, the associations between white/black and positive/negative stimuli. These associations are all grouped together in the paradigm, using vector-space semantic models, isolating racial bias from these other associative biases more effectively than past models, showing much greater reliability and predictive validity over time than D-Scores. D-scores are the typical metric used to quantify IAT performance, which quantifies IAT performance by estimating the difference in mean reaction times between two groups (Kvam et al., 2022).

The important metric for comparison with the results of other tasks is the Association Parameter. A positive score indicates that the participant was biased toward associating black faces with negative words and white faces with positive words, while a negative score indicates the opposite.

Results

Decision models of the IAT in the past have often failed to account for the variables presented by the associative and perceptive (reading/recognizing images) elements of the task (Kvam et al., 2022). In an attempt to disentangle these processes, a new analysis was utilized; the main metric of interest being the association parameter. This metric provides a numerical value that represents how easily participants associate black and white faces with good and bad words. In this study, the association values ranged from 0.25 to 0.3. The average association value across participants was .054, with a standard deviation of .15 (see Figure 3). This average positive score indicates that the average participant was slightly biased toward associating black faces with negative words and white faces with positive words. Looking at the absolute value metric created for the analysis of RLT, responses ranged from .022 to .15 across all 28

participants. The average value from this metric was .076, indicating an overall bias to respond “darker” incorrectly.

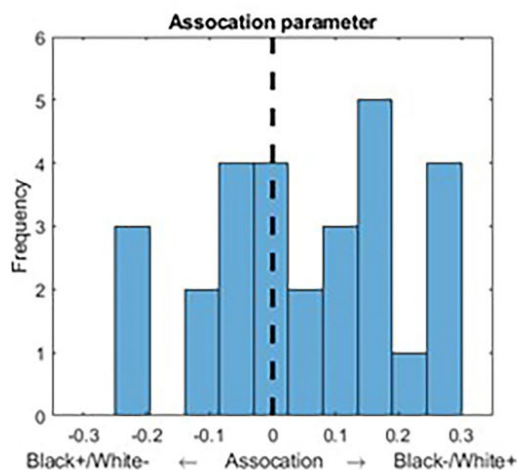


Figure 3. Distribution of Association Values Across all 28 Participants.

After utilizing the variation of the diffusion model to obtain the association metric and absolute value of the slope of darker rates for the RLT, correlations between these two individualized metrics were conducted to explore the relationship between cognitive and perceptual biases. The correlation between RLT performance and the IAT model parameter reflecting racial bias was not statistically significant, $r(26) = .18$, $p = .347$ (see Figure 4).

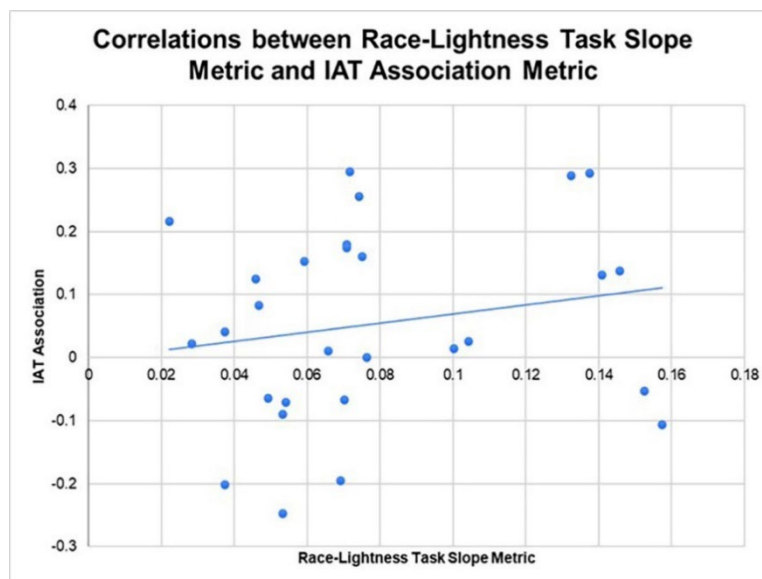


Figure 4. Correlations Between IAT Association Metric and RLT Slope Metric. Each data point in this figure represents a single subject.

In addition to this, correlations between both the two metrics above and two explicit questions were conducted: “Which statement best describes you” and “Which statement is most accurate concerning the way in which African Americans are presently treated in the United States today?” A statistically significant relationship between IAT performance and the explicit question on systemic racism prevalence was found (see Figure 5), $r(26) = .52$, $p = .004$. Specifically, the more likely an individual was to think that systemic racism was not prevalent in the US, the higher they scored on the IAT model parameter which characterized associating black individuals with negative words. However, there was no statistically significant relationship found between IAT performance and the explicit question on European vs. African preference, $r(26) = -.12$, $p = .54$. In addition to this, there was no statistically significant relationship found between RLT performance and the explicit question on European vs. African preference ($r(26) = .29$, $p = .13$) or systemic racism prevalence ($r(26) = .33$, $p = .08$). However, it is important to consider this study may be underpowered to detect an effect on these measures.

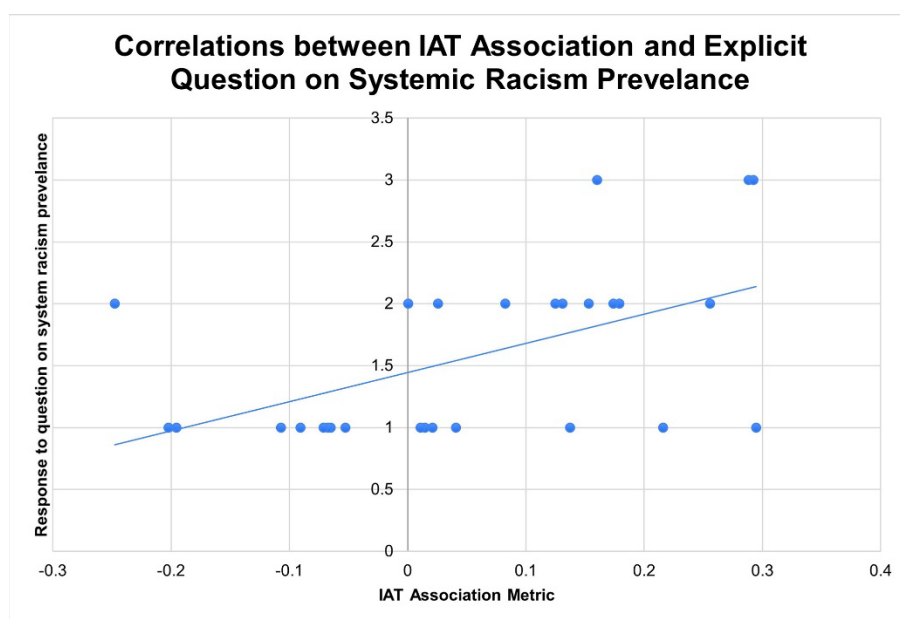


Figure 5. Correlations Between IAT Association Metric and Question on Systemic Racism Prevalence. Question responses are Likert-scaled from 1 to 4.

Discussion

The tasks completed in this study separated cognitive (IAT & explicit questions) from perceptual racial biases (in the Race-Lightness Task), allowing for their comparison within

participants. Correlations between an IAT model parameter characterizing the degree to which people associated black faces with negative words (the association parameter) and the RLT performance were not significant, providing preliminary support for the idea that these two domains may be mutually exclusive, especially for tasks concerning race. Although these results point to independence between cognition and perception when it comes to racial bias, it is important to note the significant correlation between IAT performance and self-reported views on the prevalence of systemic racism in the United States. This correlation is interesting since a comparison regarding these two constructs has not been made in the past literature. It may be useful to further investigate this relationship considering its relevance to modern efforts to combat and reduce systemic racism (Bleich et al., 2019). Further, the trends between RLT performance and explicit measures are intriguing. Overall, it is recommended that this pilot study be conducted in a larger sample to further tease apart the relationships between these constructs and tasks.

It is important to consider the many limitations posed to this study; specifically, the lack of age and socioeconomic diversity (all 28 participants were students enrolled at the University of Florida) and small sample size. However, this study brings light to a potential new method for studying the relationships between cognitive and perceptive biases. Future studies can include more participants and a more varied population (all participants were students at the University of Florida, and a significant amount were female) to gather more data on this relationship. Future studies should also ask the race of participants and run analyses based on those responses as well. The evidence for cognitive and perceptive biases being mutually exclusive domains is based on race identification in these tasks; future studies can investigate how this concept changes in other areas, such as gender, age, weight, or other implicit and perceptive biases. In conclusion, while the IAT and RLT may not correlate with one another, their surprisingly strong relationship between the IAT and an explicit question about systemic racism clearly warrants further study.

Acknowledgments

This study was funded by the University of Florida's College of Liberal Arts and Sciences and through the University of Florida's University Scholars Program.

References

- Axt, J. R. (2018). The Best Way to Measure Explicit Racial Attitudes Is to Ask About Them. *Social Psychological and Personality Science*, 9(8), 896–906.
<https://doi.org/10.1177/1948550617728995>
- Bleich, S. N., Findling, M. G., Casey, L. S., Blendon, R. J., Benson, J. M., SteelFisher, G. K., Sayde, J. M., & Miller, C. (2019). Discrimination in the United States: Experiences of black Americans. *Health Services Research*, 54(S2), 1399–1408. <https://doi.org/10.1111/1475-6773.13220>
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, 39, e229.
<https://doi.org/10.1017/S0140525X15000965>
- 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. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Kvam, P. D., Smith, C., Irving, L. H., & Sokratous, K. (2022). Improving the reliability and validity of the IAT with a dynamic model driven by associations. *PsyArXiv*. <https://doi.org/10.31234/osf.io/ke7cp>
- Levin, D. T., Baker, L. J., & Banaji, M. R. (2016). Cognition can affect perception: Restating the evidence of a top-down effect. *Behavioral and Brain Sciences*, 39, e250.
<https://doi.org/10.1017/S0140525X15002642>
- Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, 135(4), 501–512.
<https://doi.org/10.1037/0096-3445.135.4.501>
- Travers, E., Fairhurst, M. T., & Deroy, O. (2020). Racial bias in face perception is sensitive to instructions but not introspection. *Consciousness and Cognition*, 83, 102952.
<https://doi.org/10.1016/j.concog.2020.102952>
- Singular Inversions (2016). *FaceGen Modeller* [computer software]. Toronto, ON: Singular Inversions. [IAT Instruction Table]. Project Implicit. <https://implicit.harvard.edu/>