

Lee, Daniel J.

**Working Paper**

## Racial bias and the validity of the Implicit Association Test

WIDER Working Paper, No. 2016/53

**Provided in Cooperation with:**

United Nations University (UNU), World Institute for Development Economics Research (WIDER)

*Suggested Citation:* Lee, Daniel J. (2016) : Racial bias and the validity of the Implicit Association Test, WIDER Working Paper, No. 2016/53, ISBN 978-92-9256-096-6, The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki, <https://doi.org/10.35188/UNU-WIDER/2016/096-6>

This Version is available at:

<https://hdl.handle.net/10419/146244>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



UNITED NATIONS  
UNIVERSITY  
**UNU-WIDER**

WIDER Working Paper 2016/53

## **Racial bias and the validity of the Implicit Association Test**

Daniel J. Lee\*

May 2016

**Abstract:** Implicit associations and biases are carried without awareness of conscious direction. In this paper, I develop a model to study giving behaviours under conditions of implicit bias. I test this model by implementing a novel laboratory experiment—a Dictator Game with sorting to study both these giving behaviours, as well as a subject’s willingness to be exposed to a giving environment. In doing so, I adapt the Implicit Association Test (IAT), commonplace in other social sciences, for use in economics experiments. I then compare IAT score to dictator giving and sorting as a necessary test of its validity. I find that the presence of sorting environments identify a reluctance to share and negatively predict giving. However, despite the IAT’s ever-growing popularity, it fails to predict even simple economic behaviours such as dictator giving. These results are indicative that implicit bias fails to overcome selfish interests and thus the IAT lacks external validity.

**Keywords:** Implicit Association Test, implicit bias, race, prosocial behaviour

**JEL classification:** C91, D64, J15

**Acknowledgements:** This paper was made possible with the support of the UNU-WIDER. Special thanks to Susan Laury, Michael Price, John List, Charles Courtemanche, as well as ExCen laboratory assistants, numerous peers and seminar participants at Georgia State University and the University of Chicago.

---

\* Department of Economics, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, US; [dlee55@student.gsu.edu](mailto:dlee55@student.gsu.edu)

This study has been prepared within the UNU-WIDER project on ‘[Discrimination and Affirmative Action: What have we learnt so far?](#)’, which is part of a larger research project on ‘[Disadvantaged Groups and Social Mobility](#)’.

Copyright © UNU-WIDER 2016

Information and requests: [publications@wider.unu.edu](mailto:publications@wider.unu.edu)

ISSN 1798-7237 ISBN 978-92-9256-096-6

Typescript prepared by the Author and Liisa Roponen.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Denmark, Finland, Sweden, and the United Kingdom.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

# I Introduction

The Center for American Progress estimates the costs of discrimination at \$64 Billion *per year* or roughly 2 million annually displaced American workers (Burns, 2012). Discrimination is clearly costly. It is, almost universally, a unique and puzzling issue. And yet, though its existence is widely acknowledged, it is rarely discussed publicly. In particular, in Becker’s (1957) model of taste-based discrimination, animus is not only morally reprehensible, but also damaging to both social welfare and efficiency as animus necessarily burns money. However, evidence of animus is rarely observed in either naturally occurring data or field and laboratory experiments. This is perhaps due to the nature of such experiments, which tend to focus on non-visceral or unaroused decision making (*cold-phase*) when intuition dictates that personal distaste is more likely to be expressed in *hotter-phase* decisions.

This paper speaks to a recent trend in the social sciences—the claim that discrimination from animus stems from *implicit* biases and associations. The concept of implicit bias suggests that subtle cognitive processes govern our behavior. As a result, implicit biases are those that we carry without awareness of conscious direction (Kang, 2009). The development of the Implicit Association Test (IAT henceforth, discussed further below) has lent support to these claims by introducing a tractable measure of these implicit biases without having to rely on self-reporting mechanisms, which are known to be unreliable. The IAT is essentially several timed sorting tasks. In it, subjects match features, such as faces, to highly and lowly associated attributes, such as good or bad words. Allegedly, it is easier for an experimental subject to sort any feature with its more closely associated attributes. For instance, a picture of a chair is more closely associated with a word “furniture” than a word “food”, and hence more likely to be sorted faster as such. Thus, it is through this primitive of differential timing that one reveals his or her implicit biases.

There is some common-sense validation to this argument. Frequently cited example of these biases in decision making are men being more associated with management or white faces being more associated with pleasant words and feelings. As economists, we can think of these biases as coming through on the *hot-phase* of our decision-making process. However, to act on these biases in an IAT is costless, and can be thought of as a cheap-talk action. Furthermore, there has yet to be an in-depth economics experiment to test the validity of the IAT.

Regardless, meta-analyses seem to illustrate that these biases persist, (Bertrand et al., 2005; Greenwald et al., 2009) but should we care, and if so, to what

extent? The relevant question isn't merely one of existence, but whether an individual is both willing and able to act on these biases (e.g. in the case of a giving decision). To quote Dr. James Heckman, "The authors of these [discrimination] papers focus on the question of whether society is color blind, not on the specific question of whether there is market discrimination in realized transactions" (Heckman, 1998).

Given this critique and the damaging effects of bias, we want to know whether a well-functioning market can overcome implicit bias, or if it is robust to market interaction. Unfortunately, there appears to be some evidence that implicit bias is robust. For instance, Price and Wolfers (2010) claim that, due to the split-second nature of the occupation, implicit biases can explain their findings of discriminatory behaviors in NBA referees. This behavior and similar ones suggest a role for the IAT in economic research. What we first need, therefore, is a clean experimental test to see if implicit bias can predict economic behaviors. That is, is the IAT measuring the bias it claims to, and if so, does that bias influence behavior?

In this paper, I take a necessary first step in this line of research by writing a model of giving under implicit bias. I then conduct a laboratory experiment that examines the extent to which these IAT scores co-move with pro-social (giving) behaviors. Additionally, I allow subjects to sort in and out of giving environments to better identify the biases of different sharers and how they manifest in the market. I focus on giving behaviors because of a growing body of work in the social sciences discusses the relationship between bias and giving behaviors (Triplett, 2012). Furthermore giving behaviors are both non-strategic and non-spontaneous, and therefore easily controlled by the subject.

This paper proceeds as follows: The next section provides background on the IAT and relevant literature. Section III describes my model. Section IV outlines the experiment and describes the data. Sections V through VII present and discuss the results. A final section concludes.

## II Background

### II.1 The IAT

Bias cannot be randomly assigned, so the question remains, how can we measure it, particularly when we may be unaware of the biases we hold? Describing implicit bias as automatic, and analogizing the mechanics of it to those of a reflex, social psychologists Greenwald, McGhee, and Schwartz first claimed to be able to test for it using their Implicit Association Test, intro-

duced in 1998. The test is explained in their seminal paper as follows<sup>1</sup>:

An implicit association test (IAT) measures differential association of 2 target concepts with an attribute. The 2 concepts appear in a 2-choice task (e.g., flower vs. insect names), and the attribute in a 2nd task (e.g., pleasant vs. unpleasant words for an evaluation attribute). When instructions oblige highly associated categories (e.g., flower + pleasant) to share a response key, performance is faster than when less associated categories (e.g., insect + pleasant) share a key. (Greenwald et al., 1998)

Though screenshots of the IAT tasks are presented in Appendix A, this description merits further discussion. The reader will note that at its core, the IAT is essentially four (timed) sorting tasks. The first two tasks, also known as the “2-choice” tasks, are relatively simple, requiring the subject to sort either concepts or evaluation attributes. In this paper, I utilize a race (Black-White) IAT which has yet to be used in economics literature, despite the fact that black-white relations remain one of society’s most divisive issues.

Here, the measure of interest is implicit racial bias. Concepts in the 2-choice task are pictures of (black and white) faces, while the attributes are (good and bad) words<sup>2</sup>. To further illustrate, in the 2-choice tasks a subject may be asked to sort black faces on the left and white faces on the right. Similarly, another 2-choice task would be sorting associated attributes, in this case sorting good words on one side, and bad words on the other.

The other two stages combine these two sorting tasks in a “shared-response task”. Here the IAT might say good words AND white faces on the left. In this case *either* a face or a word will show up and you sort it accordingly. Then the task flips the association to say good words AND black faces on the left. This is a key distinction because the test is not eliciting a matching or opinion from subjects. Rather, this is simply a *joint* sorting task, designed to measure the strength of the association between concept and attribute. While this concept of associations may seem foreign to economists, it actually finds its roots in early utilitarian philosophies, wherein people seek not pleasure itself, but rather the objects associated with those pleasures (Mill, 1869). Table 1 shows the progression of IAT tasks.

These IAT tasks are conducted at a computer terminal where responses are measured by keystroke (e.g. E for left, I for right). The testing experience

---

<sup>1</sup>It may be difficult to visualize the assessment from this description alone, for further understanding I recommend visiting *Project Implicit*<sup>®</sup> at <http://implicit.harvard.edu>.

<sup>2</sup>examples of good words: *Joy, Love, Peace, Wonderful, Pleasure, Glorious, Laughter, Happy*; examples of bad words: *Agony, Terrible, Horrible, Nasty, Evil, Awful, Failure, Hurt*.

Table 1: Progression of IAT Tasks

Stage	Name	Description
Stage 1	Image Stimulus Learning Trial	In this trial, the custom stimulus (either images, when present, or custom words) will be presented and paired with the response to either the 'e' or 'i' key.
Stage 2	Word Stimulus Learning Trial	Most IATs that assess preference or stereotypes use positive or negative words as the associative stimuli. In this second trial, these words are presented.
Stage 3	Paired Test Trial #1	Stage 3 pairs the associations learned in Stages 1 and 2 and randomly presents a stimulus sampled from either of those sets of stimuli.
Stage 4	Reverse Image or Word Stimulus Learning Trial	Stage 4 is identical to Stage 1, except that the associations are learned with the opposite hand.
Stage 5	Paired Test Trial #2	Stage 5 combines the associations learned in Stages 2 and 4.

Source: Meade (2009)

is comparable to a human player interacting with a computer game, Empirical evidence has shown that the latent sorting time of black faces and good words in the same column is longer than it is with white faces and good words.

The standard scoring metric for the IAT is known as the *D-Score* (Greenwald et al., 2003). It is similar to Cohen’s measure for effect size,  $d$ , and is calculated as the difference in test block mean latencies by the standard deviation of latencies across test blocks. For the purposes of IAT scoring only paired trials (stages 3 and 5 in this experiment) are considered *test blocks*<sup>3</sup>. The equation for this *D-Score* is illustrated in equation 1 below:

$$D = \frac{\overline{x_1} - \overline{x_2}}{SD_{1\&2}} \quad (1)$$

Accordingly, the D-score can be either positive or negative. In the case of a

---

<sup>3</sup>See e.g. table 1.

Black-White IAT, a positive score indicates a positive (automatic) preference for whites, and vice-versa for a negative score<sup>4</sup>. A score of zero indicates little or no preference. The authors further classify and interpret D-scores using the conventional measures for effect size (Cohen, 2013), with break points at  $\pm 0.15, 0.35, 0.65$  for ‘slight’, ‘moderate’, and ‘strong’ associations, respectively.

These authors and others have then used the IAT to make several claims. For instance, these differences in latent sorting speed (and resulting D-scores) are measurements of implicit bias (i.e. the strengths of the associations we hold) and these biases are persistent—with extant anti-black biases even among minority groups (Nosek et al., 2002)! These are interesting claims, with the benefit that if true, we can observe a personal bias which people may not know, or may be unwilling to divulge. However, the claims are also dubious. While it is understandable why the IAT and similar tests use the primitive of timing, we need something stronger and more applicable in order to draw economic conclusions. As such, I ask what is the IAT actually measuring? For instance, Norton et al. (2012) suggest not wanting to appear biased (or wanting to appear race neutral) can cause a “race-paralysis” in this sort of task.

My critique is twofold, in that questions of both internal and external validity remain unanswered. Internally, consider the case of someone who may be particularly biased, but also finds sorting tasks enjoyable. Inversely, consider a subject who is unbiased but maladroit at sorting. Do we expect this sorting ability (or lack thereof) to offset the time differential?

I formalize this aspect of my critique by adapting the notation of Borghans et al. (2008):

$$T_{i,IAT} = h_i(f_i, V_i) \tag{2}$$

Let  $T_{i,IAT}$  denote person  $i$ ’s performance on the IAT task. Output in this task is generated by an individual’s implicit associations,  $f_i$ , as well as  $V_i$ , a vector of other determinants of task productivity, such as sorting ability.

Now consider, without loss of generality, the case of two individuals,  $i$  and  $j$ , with equivalent biases and productivity functions, yet one is better at sorting. That is  $f_i = f_j$ , and  $V_i > V_j$ , implying differential task performance  $h_i(f_i, V_i) > h_j(f_i, V_j) \Rightarrow T_{i,IAT} > T_{j,IAT}$ . When we allow for heterogeneity in either the bias or the production function (or both) it becomes evident that implicit bias remains unidentified.

---

<sup>4</sup>My critique notwithstanding, in this paper I will continue to use the terminology of preferences so as to remain consistent with the literature.



Further, some may be quick to point out that these subjects are unmotivated. As Grether and Plott (1979) note, this lack of motivation can be a true cause for concern in the validity of psychology experiments. However, when the outcome of interest is cheap talk (as in the IAT), unmotivated subjects may still be valid. The question of interest, which remains to be answered, is whether or not this IAT cheap talk predicts bias in marketplace behaviors. That is, is the IAT mapping into economically relevant decisions—is it externally valid? If so, what are the dosage implications? That is how much more do severe levels of implicit bias map into these decisions as opposed to moderate or even slight bias?

## II.2 Literature Review

As such, the true rub lies within the application of this test. Several studies suggest we should be interested in implicit bias by claiming that it has an effect on economic decision-making<sup>5</sup>. Again, Price and Wolfers (2010) argue implicit bias explains discriminatory behavior amongst NBA referees, although they do not use an IAT explicitly. Select few studies in economics do. Of them, Lowes et al. (2015) find evidence of ethnic homophily while Reuben et al. (2014) and Rooth (2010) find predictive evidence of negative hiring conditions. The former uses an experimental labor market for women in STEM fields, and the latter uses a correspondence study with an IAT follow-up. It finds that implicitly associated stereotypes (e.g. Arabs are lazy) forecast interview callbacks in Sweden. However, none of these papers use a Race IAT, which is the standard and most common. The alleged interaction between implicit bias and labor market decisions suggests a role for further economic analysis in other areas of decision-making, such as pro-social behavior.

Thus far, the economic study of bias has primarily dealt with competitive models—those in which individuals optimize their own behavior. These models date back to Becker (1957) as well as Phelps and Arrow (1972; 1973, respectively) who developed models relating taste-based (preferential) and statistical (informational) bias, respectively. Since these two models have different policy implications it is particularly important to properly identify the channels of bias. Briefly, in Beckers model employers may experience a disutility from hiring minority workers. Consequently, these workers may have to accept lower wages or similarly increase productivity to ‘compensate’ employers in-kind for this bias<sup>6</sup>. In Arrow’s model firms have limited information about poten-

---

<sup>5</sup>For a more thorough review of the recent psychology and management literature regarding the IAT see Jost et al. (2009).

<sup>6</sup>This does not necessarily imply an absence of discrimination, although with enough

tial employees and are forced to infer productivity information from primitive observables. In the following discussion, I will talk about discrimination as resulting from these biases.

In this vein of “primitive observables” our natural inclination as economists to identify the effects of bias is to plug some outcome of interest (e.g. wages, employment) into a regression with some likely covariates (e.g. sex, race), control for as many factors as possible, and interpret the results, or relegate bias to a residual. Comprehensive works by both Yinger (1998) and Altonji and Blank (1999) review this regression model of identification. The consensus is that there are markets in which discrimination both exists and is prevalent. The empirical challenges in these studies, however, are twofold. First, with these reduced-form models, we cannot identify the causal pathways for this discrimination (such as implicit cognition). Secondly, the observed outcomes may be severely biased due to missing data. Charles and Guryan (2011) further critique this regression approach by asking what is the ideal experiment the regressions are mimicking.

In this sense, a natural solution to these shortcomings is to run experiments. A popular method in this research has been fictitious tests in the form of either audit or correspondence studies. Both audit studies, which use trained testers (e.g. Gneezy et al., 2012) and correspondence studies, which use fabricated paper applications (e.g. Bertrand and Mullainathan, 2004; Hanson and Hawley, 2011; Hanson et al., 2011) provide further evidence of the existence of discrimination, though they are largely silent on the magnitude of the effect. In this paper, I help to identify the magnitude (or lack thereof) of any differential treatment. Furthermore, both audit and correspondence studies have the potential to produce spurious evidence of discrimination (Neumark, 2012), and furthermore are subject to the Heckman (1998) critique of auditor influence and inferences drawn that are based on otherwise unobservable factors.

These critiques suggest a role for other field and laboratory experiments. Evidence for bias is consistently found in the field. Furthermore, this evidence is persistent across a wide-variety of circumstances and domains, from excessive in-group cooperation amongst kibbutz members, when compared to Israeli city-dwellers (Ruffle and Sosis, 2006) to Pigouvian price discrimination amongst sports card traders (List, 2004).

However, laboratory experiments have not yet found significant consensus regarding the presence of bias, and an open question is the role of *implicit* bias. Several of these studies use the methodology of a Voluntary Contribution

---

unbiased employers, discrimination can be competed away.

Mechanism (VCM henceforth) public goods game (Brown-Kruse and Hummels, 1993; Cadsby and Maynes, 1998; Solow and Kirkwood, 2002; Castillo and Petrie, 2010). An outstanding issue is that VCM games study group behavior, and are not reflective of the one-on-one interactions of the audit and correspondence studies described above. Furthermore, laboratory experiments should be more reflective of the discriminatory practices that we view to be most damaging to society and welfare. In this vein we consider experiments that incorporate power asymmetries that a standard VCM game lacks, to mimic realms where bias is most present.

In response, several studies of note that have used 2-player games to measure discrimination. To study discrimination in culture, Ferraro and Cummings (2007) use the standard ultimatum game with Hispanic and Navajo subjects in New Mexico. They find significantly different behavior between the two groups. Furthermore, by eliciting subjective beliefs they claim these different behaviors are indicative of statistical discrimination. Similarly, Fershtman and Gneezy (2001) use a paired design to test for and disentangle channels of discrimination in Israeli society. In their experiment, significantly less money was passed to male Jews of Eastern origin in a trust game. However, this result was not replicated with a dictator game, indicating statistical discrimination.

Slonim and Guillen (2010) use the design of a trust game to detect gender discrimination. Further, to disentangle possible effects they include a treatment that allows for partner selection. They find (almost) no discrimination without selection, but significant taste-based discrimination with selection. Finally, Eckel and Petrie (2011) use a trust game with a costly option to see your partner's picture, and find both a demand for pictures, and increased first-mover earnings under pictures.

These 2-player designs allow for much cleaner identification than the group play of a VCM design, particularly when sorting or selection is used as a treatment cell<sup>7</sup>. The problem is this set of games still involves strategic interactions. Thus, instead of trust or ultimatum, I find a dictator game (the unique elements of which are described below) to be more appropriate to studying bias in pro-social behavior. Here, since the second player is passive, any giving is non-strategic and differences in giving can only be due to discrimination. This is discussed further in the section on experimental design below. In his review of the dictator game literature, Camerer (2003) notes that we tend to observe 10-30% of passed endowments. These rates are problematic if they are only artefactual of the lab, and could be indicative of experimenter demand effects

---

<sup>7</sup>Sorting refers to opting out of playing, whereas selection refers to picking ones partner.

or privacy concerns.

One check on the observed rates is to allow subjects to sort out of dictator giving, that is offering dictators a potentially different payoff,  $\$w'$ , to not play the dictator game (i.e. allocate  $\$w$ ). There are three notable papers that address sorting, and thereby motives for giving. In Dana et al. (2006) one third of subjects opted to take a private \$9 payoff instead of playing a \$10 dictator game. Broberg et al. (2007) extend this design by eliciting a subject's willingness to pay to exit using a BDM mechanism. They find more subjects are willing to exit, and for higher prices. Finally, and serving as the inspiration for this design, Lazear et al. (2012) (LMW henceforth) examine both costly (exit) and subsidized (entrance) sorting. Using a framework of social preferences, they find that sorting not only affects *how many* people share, but also what *kinds* of people share.

Though not yet used to examine bias, the motives for giving argument and the particulars of a sorting design apply nicely to this field of inquiry. It is my intent to describe these *kinds* of people not only by their giving behaviors, but potentially by their implicit biases as well. Further, I examine how these biases affect their decisions. This naturally follows from the behavioral finding that subjects are more likely to opt of cross-race environments necessitating a judgment of racial characteristics relevant to common stereotypes (Norton et al., 2012).

In addition to addressing the above problems, this paper contributes to the literature in several novel ways. It is the first to examine the psychological pathways of bias by using the IAT. This is important because as stated above different pathways may have different economic implications for behavior. This paper is unique in providing racial information of the recipients and allowing a sorting option with varying property rights in a dictator game. By comparing the observed rates giving and differential exits to IAT scores, this paper investigates validity of the IAT in a way the research was previously lacking. Accordingly the extent of racial bias and the external implications of the test are thereby assessed.

### III Model Description

Assuming the IAT actually measures bias, it should also be able to predict economic decisions reflecting that bias, such as giving and sorting behaviors. However, the directions and theory underlying these decisions have not been fully explored. To that end I formalize a model of giving under implicit bias.

LMW note that different kinds of sharers exist, and introducing a sorting

environment allows us to distinguish between these types, described as follows: Willing Sharers, who prefer to share and enter into sharing environments; Reluctant Sharers, who prefer not to share but do so to comply with social pressures, norms, or mores; and Non-Sharers who simply do not share<sup>8</sup>.

In this vein, LMW wish to detect a reluctance to share. I revisit this analysis and extend the definitions further by examining one potential pathway of this reluctance—*conditional* on bias.

For the purposes of this experiment, consider a utility maximizing individual, henceforth referred to as the dictator. The dictator is indexed by her level of bias,  $i$ , which I assume manifests as animus and perfectly correlates to the dictator’s IAT D-score<sup>9</sup>. The D-score which is drawn from a standard normal distribution, that is  $i \sim N(0, 1)$ <sup>10</sup>. In this model, the dictator may be in an economic environment that allows sorting, and may also have photographic information on her receiver. If the former, the dictator can take up to two possible actions. First, the sorting decision, that is the decision between allocating an amount  $w$  (sorting in) or receiving an amount  $w'$  (exiting out). Conditional on sorting in, she must now make the decision of how much to give, that is how to split the endowment  $w$  between herself,  $x$ , and the recipient,  $y$ , such that  $x + y = w$ .

I further hypothesize that individuals also sort based on who they are sharing with, and this sorting also manifests itself as animus. As such, I also allow the dictator to consider the race of the recipient  $r$ . This consideration only occurs if the dictator has photographic information. Thus, the dictator has preferences over her environment  $D$ , her payoff,  $x$ , the payoff to the recipient  $y$  and the similarity of the race of the recipient  $r$ . It is these preferences that determine sorting or not sorting, and potentially the giving decision:

$$U_i = U_i(D, x, y, r) \quad (3)$$

where  $D$  is an indicator variable such that  $D = 1$  if the environment has sorting and 0 otherwise; and  $r$  is an indicator variable such that  $r = 1$  if the dictator has photographic information and is the same race as the individual, and 0 otherwise.

Within subjects, the theory of animus dictates that not only is an individ-

---

<sup>8</sup>Formal definitions for these types can be found in Appendix B.

<sup>9</sup>This assumption that the test is registering bias as opposed to cultural knowledge of stereotypes is consistent with accepted interpretations of the IAT (Nosek and Hansen, 2008).

<sup>10</sup>The actual IAT D-score is truncated at -2 and 2, but this does not affect model predictions.

ual’s utility greater for an equal amount given to the preferred race<sup>11</sup>:

$$i \geq 0.15 \Rightarrow U_i(D, \bar{x}, w - \bar{x}, 1) \geq U_i(D, \bar{x}, w - \bar{x}, 0) \quad (4)$$

but also that a person is willing to take a utility hit to express his or her distaste. Here, that means a willingness to sort out (even if the sort is costly) for the sole purpose of *not sharing*:

$$i \geq 0.15 \wedge (w < w') \Rightarrow U_i(1, w', 0, 0) > U_i(0, w, 0, 0) \quad (5)$$

This unwillingness to interact is a core concept of animus. As such, across subjects, the model of animus predictions that greater bias should have more costly sorting, in addition to less sharing across races. That is:

$$U_i(D, \bar{x}, w - \bar{x}, 0) > U_j(D, \bar{x}, w - \bar{x}, 0) \Rightarrow i < j, \forall \bar{x} < w \quad (6)$$

$$w < w' \Rightarrow U_i(1, w', 0, 0) > U_j(1, w', 0, 0), \forall i > j \quad (7)$$

In this experiment, I restrict my focus to the across subject design. Broadly speaking, I ask two initial empirical questions based on this model. If the answer to either of these first two questions is yes, it suggests that there is a clear pathway from the the *hot-phase* IAT task to some of the *cold-phase* decisions it has been used to explain. Absent evidence of this pathway, I ask a final question concerning meta-awareness of bias:

1. Does the IAT predict giving behavior?
2. Does the IAT predict sorting out of giving environments?
3. Do biased givers attempt to mitigate their bias with small gifts?

## IV Experiment

### IV.1 Procedures

Given that previous lab experiments have demonstrated that these different types of individuals exist, I ask what are the IAT’s implications for both laboratory and naturally occurring behavior. I use the toolbox of experimental economics to see if IAT performance is related to differential treatment of receivers and if so, to what extent. In doing so, I examine the IAT as a

---

<sup>11</sup>I have written this model as biased against people of color, but it is trivial to generalize to all racial bias.

predictor of pro-social behavior in an experimental market. This behavior includes giving as well as sorting out of potential giving environments.

To properly ask (and answer) these questions this experiment necessarily progresses in two stages: first the dictator game (potentially with a sorting option), and second with the IAT. Upon arriving at the lab subjects are randomly split into receivers and dictators. I will now explain the two roles in turn.

In a standard dictator game, a first mover is given \$10 and asked how much she would like to give to a paired (and passive) player; her choice ends the game. Thus, giving in this game is non-strategic. I begin with this standard (no information) treatment to gauge dictator giving without information on the race of the recipient.

From here, I differ from a standard game in that some treatments employ a sorting environment. Specifically, I offer some dictators an exit option as in LMW. In other words, dictators are given a chance to leave the game in such a way that the passive player never knows he or she was playing a dictator game. In doing so, I aim to disentangle social pressure as a motive for giving. This opportunity (choice) can be either costly or free. The costly option is necessarily payoff dominated by at least one dictator game choice.

Finally, these treatments are run in two types of sessions: Ones with no information (anonymous), and pictures sessions, where dictators can see who they are passing to, and use that picture as a proxy for race. For the most part, we are concerned with outcomes in the “Pictures” sessions. However, the anonymous treatments serve as an interesting comparison and are necessary for commenting on the social closeness afforded by a picture. Further, the cross between pictures sessions and sorting treatments allows us to see whether implicit bias is affecting behavior on either the extensive or intensive margin. That is, the decision *to engage* in giving as well as *how much* to give.

Table 2: Dictators by Treatment

	Baseline	Sorting	
		Costly	Free
No Information	20	13	20
Pictures	48	68	59
Total	68	81	79

Source: Author’s calculation

As such, this experiment necessitates a 2x3 design. The treatment cells are as follows: A standard (baseline) dictator game, and two dictator games with

sorting: costless and costly. In costless sorting, the dictator receives the same amount in entry and exit (\$10). In costly sorting, the dictator receives \$9 upon exit. These dictator games are all played across both anonymous and pictured sessions. The treatment cells and number of dictators that participated in each treatment are described further in table 2 as well as in the data section below.

After roles are assigned, the dictators are randomly paired with a receiver, and in the *Pictures* treatments shown a picture of that receiver’s face. The photos serve as a proxy for race. In the *No Information* treatments dictators are not informed about their receivers. In both versions, dictators are then explained the rules of the dictator game. In all but the baseline treatment, they are asked whether or not they choose to participate. In the event that a dictator elects to not participate (takes the exit option), their receiver is not given any information about allocation task, and the dictators are given their exit fee (\$9 or \$10, depending on treatment). Otherwise, dictators decide how to allocate a sum of \$10 between themselves and their receiver.

Meanwhile, the receivers are passive in their role. They have their pictures taken, are guaranteed a show-up fee, and asked to participate in a different task. In this case, that task is a real-money,  $1x$  risk-preference elicitation (Holt and Laury, 2002), the results of which I discuss in a companion paper (Lee, *in-progress*). The receiver task is constant across treatments.

The next task in the experiment is a race IAT (as described above) on all subjects. I run this task second because an IAT can possibly influence amounts passed. However, knowing they have just participated in a dictator game should not influence IAT score, as evidence shows it is difficult to fake or otherwise manipulate (Fiedler and Bluemke, 2005). I then close by collecting demographic data in the form of a survey, and pay subjects privately. Complete subject instructions and survey questions can be found in Appendices C and D, respectively.

## IV.2 Data

These experiments were conducted during the summer and fall of 2015 at the Center for Experimental Economics at Georgia State University (*ExCEN*). Subjects were recruited via email using the center’s recruiter. While I strove for sessions to be racially balanced, this was not possible given the makeup of the subject pool. However, I believe this to be non-problematic given the experimental design, as well as the evidence cited above on implicit attitudes and minority groups.

Overall, I ran 17 experimental sessions across the 6 treatments, with a



Table 3: Dictator Summary Statistics

Variable	Mean	Std. Dev.	N
Male	0.399	0.491	228
Black	0.724	0.448	228
Catholic	0.092	0.29	228
Previous Experience	0.794	0.405	228
Business or Econ Major	0.268	0.444	228
Age	21.775	4.838	227
Year in School	3.149	1.07	221
GPA	3.302	0.448	189

Source: Author's calculation

Table 4: Baseline Comparison of Roles in the Experiment

Panel A: $\chi^2$ Tests			
Covariate	Dictator %	Receiver %	p-Value
Male	40.08	51.54	0.014*
Black	72.24	67.84	0.306
Catholic	9.257	9.69	0.873
Previous Experience	79.29	76.21	0.430
Business or Economics Major	26.87	21.59	0.189
Panel B: <i>Rank-Sum</i> Tests			
Covariate	Dictator Mean	Receiver Mean	p-Value
Age	21.78	21.15	0.216
Year in School	3.15	3.10	0.835
GPA	3.30	3.26	0.563

Notes: \* Significant at the 5% level

Source: Author's calculation

roughly equal balance of subjects across treatment rows<sup>12</sup>. In total, 227 dictators (i.e. 454 subjects) participated in the experiment. Table 3 describes the demographic breakdown of the dictators. Dictators in this experiment are (on average) 22, with a 3.3 GPA. Roughly 72% are Black and 40% are Male. Most have previous experience in economics experiments, and the modal year

<sup>12</sup>Given my power analysis and the fact that receiving in the *No-Information* treatments is anonymous, I didn't require as many subjects.

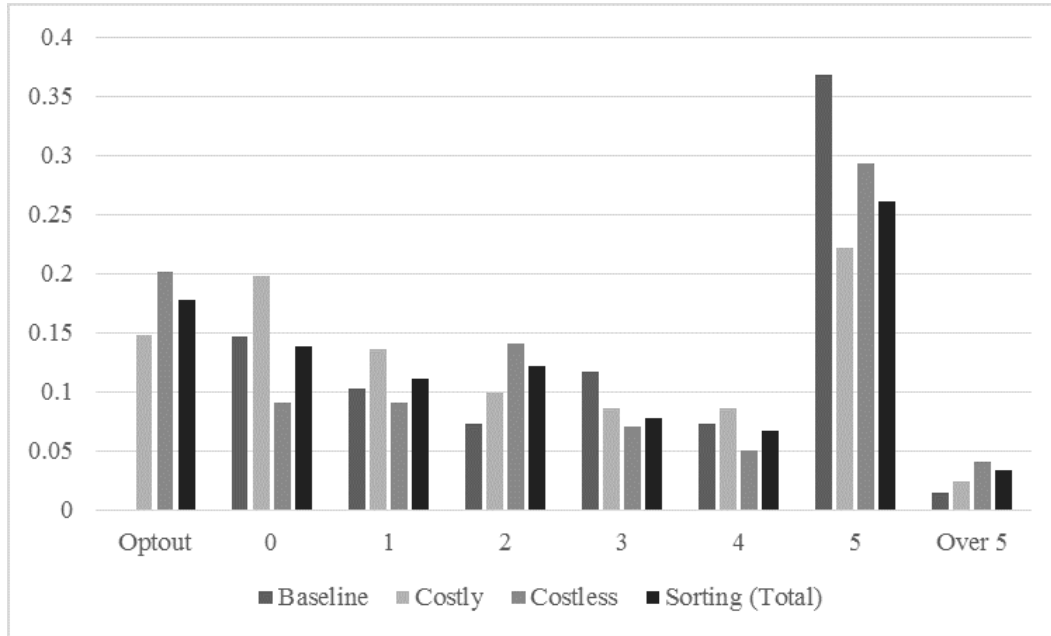
Table 5: Experimental Summary Statistics

Variable	Mean	Std. Dev.	N
Passed to Male	0.513	0.501	228
Passed to Black	0.675	0.469	228
Amount Passed	2.692	2.238	228
Opted Out (Total)	0.186	0.389	167
Opted Out (Costly)	.148	0.357	81
Opted Out (Costless)	0.202	0.404	99
IAT D-score	0.054	0.495	225

Source: Author's calculation

in school is senior<sup>13</sup>.

Figure 1: Distribution of Amounts Passed



Source: Author's illustration

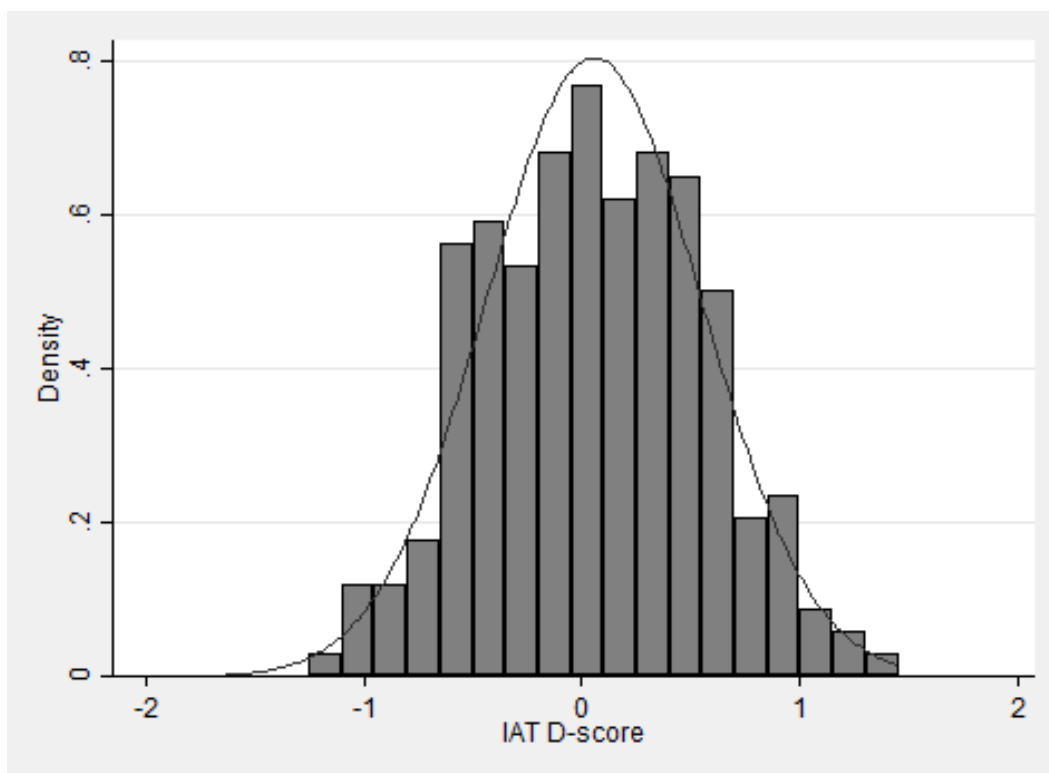
Non-parametric analyses in the form of  $\chi^2$  and Rank-Sum tests examine covariate balance between roles. For the most part, I find no significant differ-

<sup>13</sup>This is perhaps an artifact of running a summer experiment, where both former juniors and recent grads identify as seniors.

ence across them, and conclude that the sample is balanced<sup>14</sup>. These results are reported in full in table 4.

Finally, table 5 provides a brief description of dictator choices and performance in the experiment. On average, 27% of the endowment was passed, and a little more than 18% of those offered an exit option opted out, with more people exiting when it is costless. Rank Sum tests show sorting significantly decreases sharing, even when sorting is costly (Sorting:  $z = 2.146$ ,  $p < 0.05$ ; Costly Sorting:  $z = 2.370$ ,  $p < 0.05$ ). These numbers are roughly similar to previous findings. Full distributions of amounts passed are illustrated further in figure 1.

Figure 2: Distribution of IAT D-scores



Source: Author's illustration

Regarding the IAT, the average D-score was 0.05, suggesting little to no automatic bias. I depict these scores in figure 2 for further exploration. The

<sup>14</sup>There were significantly more males in the receiver role (117 as opposed to 91), but this is likely a byproduct of hypothesis testing across several covariates.

scores follow a fairly normal distribution, consistent with both model assumptions and extant results across a variety of subject pools. The modal score is in the bin 0-0.15 (no automatic bias). However, there is significant implicit bias in the sample. Over 44% of dictators have an IAT D-score greater than or equal to 0.15, indicating a pro-white implicit bias. Consonant with the above evidence, this bias is present and perhaps stronger in subjects identifying as black, with a mean IAT score of 0.162.

## V Discrete Results

We have seen descriptively that sorting environments affect giving behaviors. However, given the empirical questions asked above, I now turn my focus to the role of the IAT in making these economic decisions. I first explore this role by simply looking at average amounts passed, broken up by the dictator’s bias. Specifically, table 6 shows the mean pass broken down by both the strength of the association, and the recipient. First of all, these differences are not significant. Secondly, if implicit bias had a one-to-one mapping into giving behaviors, we would expect passes to black subjects would get smaller as we move down the table (strengthen the bias towards whites), and the opposite pattern for whites. However, these directional patterns do not emerge, particularly in the black recipient column. Here, those who have dictators biased against them end up earning *more* on average.

Table 6: Average Amount Passed by IAT score and Race of Receiver

<b>Strength of Implicit Bias</b>	<b>Passed to:</b>		
	<b>Black</b>	<b>White</b>	<b>Anonymous</b>
Strong for Blacks	2.07	1.67	2.5
Moderate for Blacks	2.28	2.33	2.75
Slight for Blacks	3.43	1.67	2.83
Little to None	2.18	2.55	2.33
Slight for Whites	3.13	3.5	3.33
Moderate for Whites	2.47	2.33	2.54
Strong for Whites	2.68	3.33	2.4

Source: Author’s calculation

Next, I take what we learned in table 6 and discretize IAT score into the blunt question of “do I (implicitly) like or dislike my recipient?”

I express this question in equation 8:

$$Outcome_i = \alpha_0 + \beta_1(LikeReceiver_j) + \beta_2(DislikeReceiver_j) \quad (8)$$

Here, I regress an outcome variable on two variables *Like* and *Dislike*. The outcome takes the form of either a continuous variable representing the percent of endowment shared, or a binary variable indicating whether a dictator took an exit option. The two variables *Like* and *Dislike* are essentially binary interaction terms defined formally as follows in equation 9:

$$\begin{aligned} Like &= \begin{cases} 1 & \text{when } IAT \geq 0.15 \text{ and Receiver is White} \\ 1 & \text{when } IAT \leq -0.15 \text{ and Receiver is Black} \\ 0 & \text{otherwise} \end{cases} \\ Dislike &= \begin{cases} 1 & \text{when } IAT \leq -0.15 \text{ and Receiver is White} \\ 1 & \text{when } IAT \geq 0.15 \text{ and Receiver is Black} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (9)$$

That is to “like” your receiver means to either hold a pro-white bias and pass to a white receiver, or hold a pro-black bias and pass to a black receiver. I later decompose the variable into these two components (pro white, white receiver and pro-black, black receiver). Similarly, to “dislike” means to have the one of same IAT scores as above, but with the race of your receiver flipped. Accordingly, the intercept term,  $\alpha$ , represents those dictators who hold little to no implicit bias ( $-0.15 < IAT < 0.15$ ).

Results from these discrete estimations are presented in table 7. In the first column we find that unbiased givers share about 23% of their endowment, and being biased against (or in favor of) your receiver does not significantly alter this giving pattern. Furthermore, both directions of bias remain insignificant when decomposing the *Like* and *Dislike* variables into their respective components in column 2.

Similarly, table 7, columns 3 and 4 look at how bias influences the probability of opting out. In both the blunt (column 3) and decomposed (column 4) measures, neither liking nor disliking one’s receiver has any significant impact on giving.

These results indicate that bias does not affect the decision giving *on average*. However, a relevant question is do dictators that are biased against black (white) receivers behave differently than the average dictator with a black (white) receiver. To answer this question, I conduct an exercise similar to the one outlined in equation 8, but restrict the sample based on race of receiver and only regress on the *Dislike* variable.

Table 7: Discrete IAT Estimations

VARIABLES	OLS–Percent Shared (1)	(2)	Probit–Opted Out (3)	(4)
Like Receiver	0.0696 (0.0469)		0.0783 (0.336)	
Pro-White, White Receiver		0.0529 (0.0593)		-0.578 (0.568)
Pro-Black, Black Receiver		0.0774 (0.0507)		0.292 (0.358)
Dislike Receiver	0.0291 (0.0492)		0.166 (0.322)	
Pro-White, Black Receiver		0.0423 (0.0498)		0.0955 (0.337)
Pro-Black, White Receiver		-0.0289 (0.101)		0.456 (0.504)
Constant	0.229*** (0.0391)	0.229*** (0.0391)	-0.887*** (0.257)	-0.887*** (0.257)
Observations	172	172	126	126

Robust standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ 

Source: Author's calculation

Even when we isolate the sample by race of receiver, biased dictators are not behaving in ways that are not statistically different than the average dictator, nor is this difference *economically* significant. While the above results are indicative that implicit bias fails to overcome selfish concerns, they have mostly examined the effect of IAT score on economic behaviors. Another way of looking at the question is to treat the data as observational, and ask (with some abuse of notation) what is the treatment effect of being paired with someone you hold a bias against?

To answer this question I exploit the random assignment of roles and partners and implement propensity score matching. Here, I treat each person as having a particular bias strength and direction, ranging from strongly pro-black to strongly pro-white (e.g. see table 6). I match on the strength and direction of this bias as well as covariates describing the dictator's age, race, and sex, and specify the treatment as passing to someone you hold a bias

Table 8: Discrete Estimations, Conditional on Race of Receiver

VARIABLES	Black Receiver		White Receiver	
	(OLS)	(Probit)	(OLS)	(Probit)
Dislike Receiver	-0.00310 (0.0407)	-0.132 (0.290)	-0.0727 (0.100)	0.952* (0.576)
Constant	0.274*** (0.0267)	-0.659*** (0.191)	0.273*** (0.0381)	-1.383*** (0.374)
Observations	127	93	45	33

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Source: Author's calculation

against. That is, passing to a black person if you hold a pro-white bias and vice-versa. I find no significant treatment effect of passing to someone you are biased against (ATT=0.14,  $p=0.625$ ).

**Result 1** *Existence of bias towards receiver does not predict dictator giving*

## VI Continuous Results

However, we measure IAT score as a continuous variable, and are able to comment not only on the *existence* of implicit bias, but also the *strength* of that bias. As such, one would think that more severe biases would exert more influence on the giving and sorting decisions. To address this dosage question, I standardize the IAT score and outline the following reduced form empirical specification:

$$Outcome_i = \beta_0 + \beta_1 IAT_i + \beta_2 (IAT_i * Race_j) + \beta' X + \varepsilon_i \quad (10)$$

This standardization allows me to interpret coefficients as the effect of a one standard deviation increase in IAT score. In this specification I again regress an outcome variable on two variables of interest: that dictator's IAT score and an interaction term of dictator's IAT score with the race of her recipient, as well as a vector of demographic controls for both dictators and receivers. The interaction term allows us to examine this giving conditional on being paired with the object of one's bias. This interaction is also consistent with the model assumption that the IAT manifests as animus. The controls are

necessary because observed differences in the outcome variable may be driven by factors unrelated to a dictator’s implicit bias. Different specifications below may highlight different sets of these parameters in my analysis.

## VI.1 Dictator Giving

I continue with a graphical exploration of the IAT’s relationship to giving. Figure 3 shows the amount passed given a dictator’s IAT score. Despite the IAT’s popularity in academic work, there is no clear linear relationship between IAT score and amount passed ( $\rho = -0.01$ ). Further, in each “column” of IAT score there appears to be a similar bimodal distribution of amount passed. This suggests that levels of implicit bias do not necessarily map into the behaviors of interest.

To confirm these findings econometrically, we turn to table 9 which presents this paper’s main estimates. In these models I restrict the sample to only the dictators in sessions with photographs, although the results hold when expanded to the full sample. Additionally, I have used both dictator and receiver dummies for African-American, rather than what race a subject is biased against. While this may be a coarse measure, this modeling technique makes more sense in terms of coefficient interpretation since IAT score is increasing in the level of anti-black bias. Further, these results are consistent with the discrete estimations from section V and robust to the alternate specification of “biased against receiver”<sup>15</sup>.

In panel A of table 9 I start with a simple OLS and regress percent shared on the parameters of interest. We see that neither implicit bias nor its interaction with a black receiver yields a significant predictor of giving. These results hold true in specifications that control for race and gender of dictator, the receiver, and both. Further, these controls also have no significant effect on giving.

However, the presence of a sorting option consistently and significantly decreases the amount shared by around 10%. This result suggests that in terms of giving behaviors, people aren’t acting on their implicit biases, and perhaps are able to control any bias they may hold. Instead, social preferences unrelated to the IAT, especially pressure to give, appear to be strongly influencing these pro-social behaviors (or lack thereof).

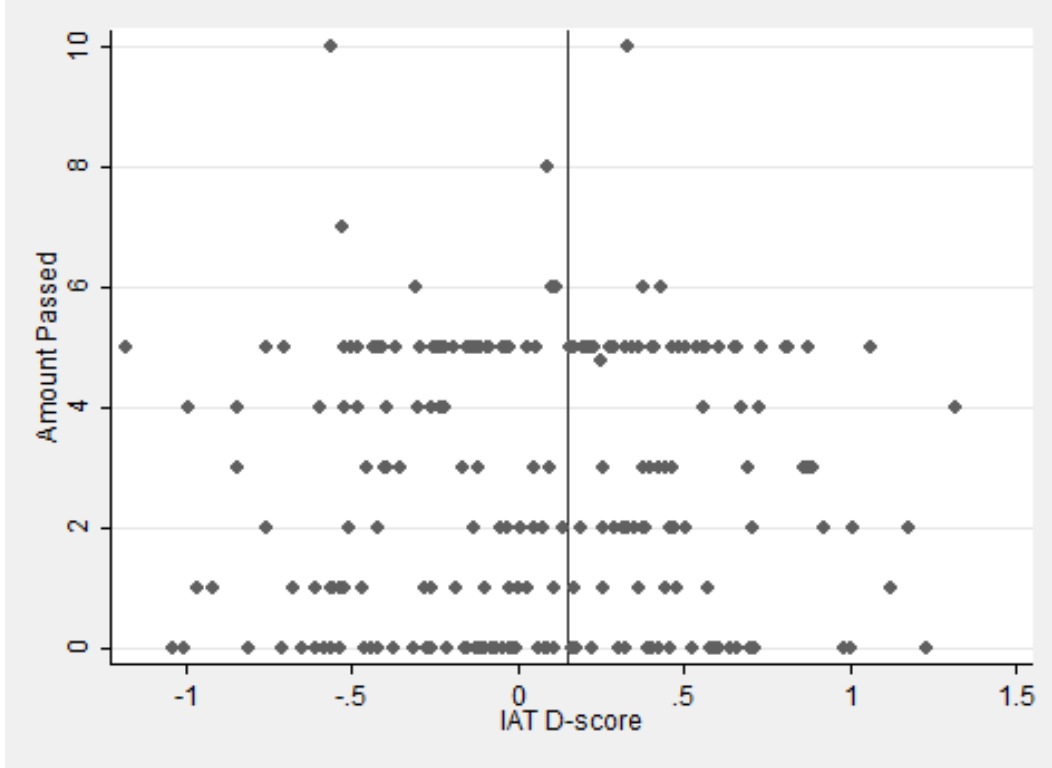
Next, to account for the 27% of dictators who either gave nothing or opted out, I replicate the OLS results with a left-censored Tobit model<sup>16</sup>. These

<sup>15</sup>See, for instance, tables 7 and 13.

<sup>16</sup>Robust standard errors are calculated using jackknife estimation. A double-hurdle model (Cragg, 1971) would be inappropriate here because to account separately for the opt-out process requires restricting the sample to only those in sessions with sorting. Results



Figure 3: Scatter Plot of IAT Score and Amount Passed



Notes: The solid line indicates the IAT-D "bias" threshold of 0.15  
Source: Author's illustration

results are shown in table 9, panel B, and are not categorically different than the OLS results. That is, IAT score is positive but insignificant, the interaction term is negative but not significant, controls lack significance, and the presence of a sorting option is strongly and negatively significant.

Following LMW, I assess the determinants of sharing in table 10. Specifically, I compare the relative importance of implicit bias (in column 1) to the presence of the sorting option, as well as self-reported demographics that could potentially affect sharing (in column two). Again, one's amount of implicit bias does not significantly determine sharing. Magnitudes of these results are similar when I run the full model, including IAT score with demographic controls (column 3). Additionally, I calculate coefficients of partial determination<sup>17</sup>. This measure shows that not only does implicit bias lack statistical

from this model are presented in Appendix E.

<sup>17</sup> $(R^2 - R_i^2)/(1 - R_i^2)$  where  $R_i^2$  is the  $R^2$  with predictor  $i$  removed from the equation.

Table 9: The IAT's Effect on Percent Shared

Panel A: OLS					
Variable	(1)	(2)	(3)	(4)	(5)
IAT D-score	0.0197 (0.0353)	0.0169 (0.0351)	0.0286 (0.0450)	0.0262 (0.0385)	0.0389 (0.0366)
IATxPassedBlack	-0.0617 (0.0805)	-0.0439 (0.0795)	-0.0494 (0.0797)	-0.0595 (0.0867)	-0.0665 (0.0843)
Sorting Option		-0.101*** (0.0355)	-0.0960*** (0.0368)	-0.0944*** (0.0361)	-0.0867** (0.0377)
Dictator Controls			X		X
Receiver Controls				X	X
Constant	0.269*** (0.0176)	0.343*** (0.0278)	0.388*** (0.0509)	0.339*** (0.0460)	0.382*** (0.0641)
Panel B: Tobit					
Variable	(1)	(2)	(3)	(4)	(5)
IAT D-Score	0.0298 (0.0485)	0.0251 (0.0481)	0.0361 (0.0483)	0.0426 (0.0535)	0.0549 (0.0516)
IATxPassedBlack	-0.0854 (0.110)	-0.0574 (0.109)	-0.0632 (0.109)	-0.0883 (0.120)	-0.0959 (0.117)
Sorting Option		-0.145*** (0.0448)	-0.138*** (0.0462)	-0.132*** (0.0452)	-0.122** (0.0469)
Dictator Controls			X		X
Receiver Controls				X (0.0478)	X (0.0495)
Constant	0.222*** (0.0254)	0.327*** (0.0336)	0.366*** (0.0688)	0.323*** (0.0576)	0.359*** (0.0849)
Observations	172	172	172	172	172

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Source: Author's calculation

significance, but one's IAT score accounts for less than 4% of the unexplained variance, and lacks economic significance as well.

The above exercises hold true when instead of looking at the coarse measure of race of receiver, I look at the finer measure of being biased against one's receiver. In figure 4, I graph box plots for a further analysis of what happens

Table 10: Determinants of Sharing

Variable	(1)	(2)	(3)	Partial R <sup>2</sup> 's
IAT D-score	-0.00247 (0.0169)		0.00850 (0.0195)	0.036
Sorting Option		-0.0847** (0.0423)	-0.0830* (0.0448)	0.158
Age		0.00642*** (0.00245)	0.00614** (0.00253)	0.148
Male		0.00345 (0.0407)	-0.00223 (0.0426)	0.005
Black		-0.0992* (0.0502)	-0.107** (0.0519)	0.185
Catholic		-0.000280 (0.0708)	0.00947 (0.0761)	0.012
Previous Experience		-0.0258 (0.0466)	-0.0277 (0.0468)	0.047
Major: Business or Econ		0.00407 (0.0420)	0.00614 (0.0429)	0.012
GPA		-0.112*** (0.0409)	-0.110*** (0.0411)	0.219
Constant	0.268*** (0.0176)	0.657*** (0.164)	0.659*** (0.164)	
Observations	172	146	143	
R-squared	0.000	0.138	0.136	

Robust standard errors in parentheses

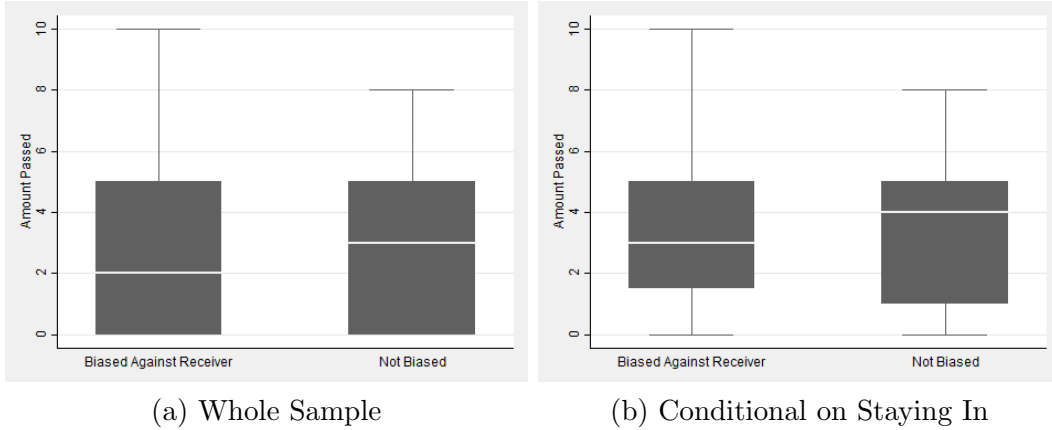
\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Source: Author's calculation

when a dictator is biased against the race of his or her receiver. For these figures that means both passing to a black receiver when biased against blacks ( $Receiver = Black | IAT \geq 0.15$ ) as well as passing to a white receiver when biased against whites ( $Receiver = Black | IAT \leq -0.15$ ).

Clearly there is no difference in giving when I consider the whole sample in figure 4a. But, this result also holds when I consider only those dictators who did not take an exit option in figure 4b. We will see a similar result regarding dictators choosing to opt out in the following subsection on Dictator Sorting.

Figure 4: Sharing When Dictator is Biased



Source: Author's illustration

Finally, I compare the picture treatments to the anonymous ones. Using rank-sum tests, amounts given by the dictator do not appear to be different across these two treatment rows ( $z = -0.039, p = 0.969$ ). This holds when we ignore baseline treatments and consider only those with a sorting option ( $z = -1.268, p = 0.205$ ), or restrict the sample to dictators paired with the object of their bias ( $z = -0.863, p = 0.388$ ). Since a dictator cannot see his or her receiver in the anonymous treatments, it is unlikely that implicit racial bias comes into play in this sharing decision. The lack of difference between the two treatment rows here is further indicative of the null results above.

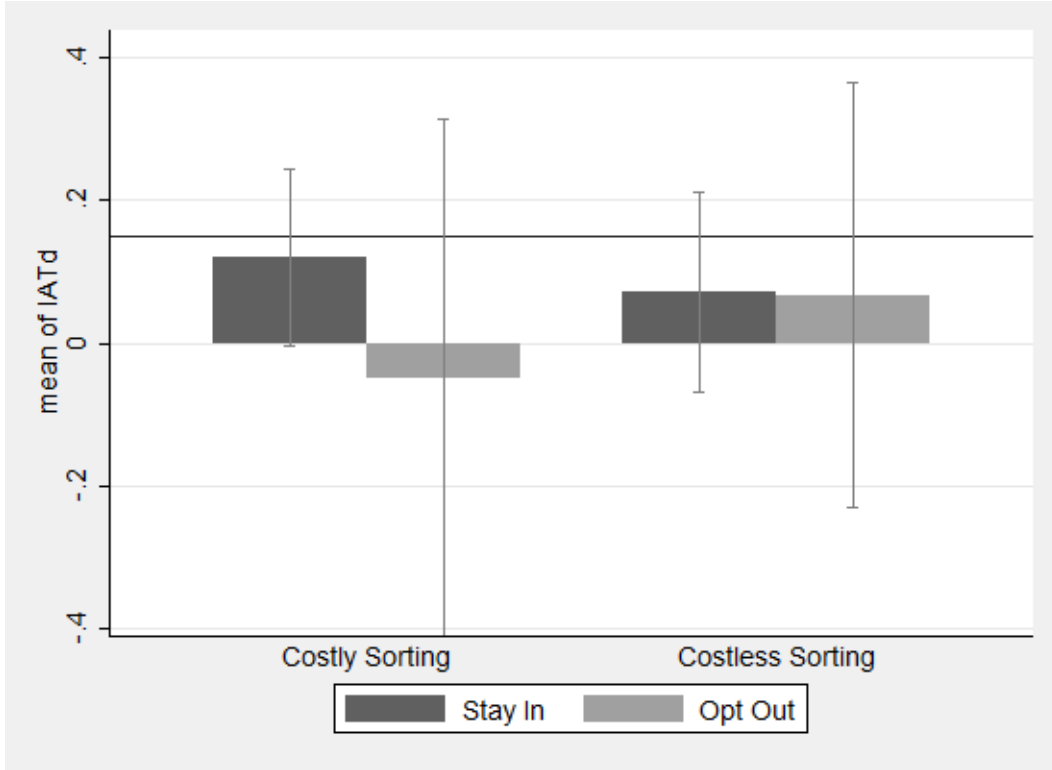
Given the overwhelming evidence above, I now declare the first result, regarding implicit bias and dictator giving:

**Result 2** *Amount of Implicit Bias (as indicated by IAT D-score) does not predict dictator giving*

## VI.2 Dictator Sorting

Perhaps the above results are indicative that biased dictators are forward thinking with regard to these biases or otherwise self-aware enough to recognize their biases. If so, they may be simply choosing not to enter sharing environments where they can express this distaste, or similarly choosing to express this distaste through their opt-out. However, we see in figure 5 the average IAT score for dictators in treatments with an exit option. Under costly and costless sorting schemes, the mean IAT score is descriptively smaller amongst

Figure 5: Bar Graph of IAT Scores and Sorting



Notes: The solid line indicates the IAT-D "bias" threshold of 0.15  
Source: Author's illustration

those who stay in (as compared to those who opt out), whereas in costless sorting the mean IAT score is essentially the same. However, in both cases, this difference is not significant (Costly:  $t = 1.04$ ,  $p = 0.30$ ; Costless:  $t = 0.03$ ,  $p = 0.98$ ).

Accordingly, I estimate the probability of opting out in table 11. This model uses a probit regression and necessarily restricts the sample to only those dictators with an exit option (that is, those in sorting treatments,  $n=159$ ). The variable structure is intended to mimic the experimental design, using dummy variables for treatment and a measurement variable to indicate IAT score. In this model, there are no significant coefficients, suggesting that overall, one's IAT score does not seem to influence the decision to sort out, with this result holding even when controlling for both the financial and social costs of sorting.

Nonetheless, this exploration again calls for a deeper analysis. Following

Table 11: The Probability of Opting Out

Probit Regression	
Variable	Coefficient
IAT D-score	-0.095 (0.114)
Costless Sorting (Pictures)	0.369 (0.253)
Costly Sorting (Anonymous)	-0.463 (0.547)
Costless Sorting (Anonymous)	0.116 (0.370)
Constant	-0.976*** (0.184)
Observations	159
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	
Source: Author's calculation	

equation 10 I look at the econometric results to confirm. In this case I ignore the anonymous treatments ( $n = 127$ ) and run Probit estimations to determine what effect (if any) IAT score has on the probability of opting out. Table 12 shows the marginal effects of these estimations. Consistent with the results above, the IAT has no significant effect on sorting. This holds when I control for whether the sorting is costly, and race and gender of the dictator, receiver, and both. Similar to the analysis under dictator giving, the signs of these coefficients are also unexpected. We see that more biased dictators opt out less. Specifically, an increase in IAT score by 1 standard deviation leads to roughly an 11% smaller chance of opting out.

As a check, I examine what happens to sorting when a dictator is biased against the race of his or her receiver ( $n = 75$ ). In this case I draw a bar graph in figure 6. Confirming the results above, there is no evidence that bias has an effect on sorting, even when the dictator holds an implicit bias against the receiver's race.

Finally, we extend the cross-treatment exercise from above and compare anonymous sorting to sorting when photo information is present, by way of Pearson's test. Again, there is no statistical difference between opting out in the two treatment rows ( $\chi^2 = 0.611, p = 0.434$ ). This holds in costly sorting

Table 12: The IAT's Effect on Sorting

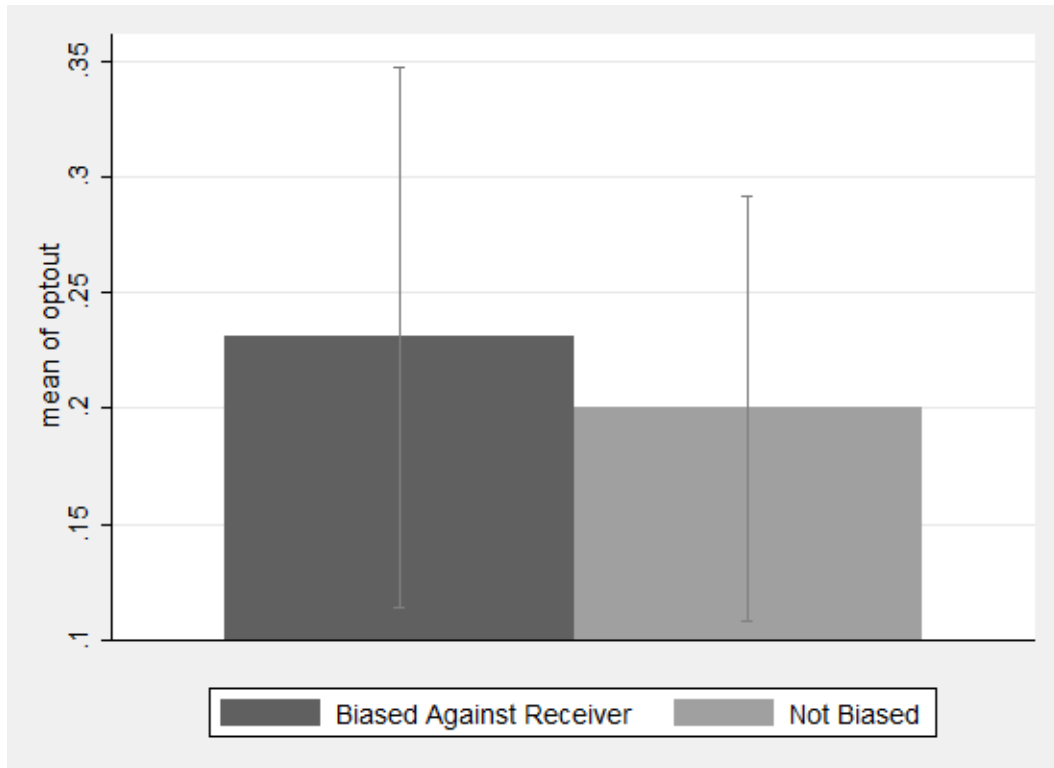
Probit Marginal Effects					
Variable	(1)	(2)	(3)	(4)	(5)
IAT D-score	-0.114 (0.0727)	-0.105 (0.0744)	-0.108 (0.0814)	-0.111 (0.0871)	-0.115 (0.0933)
IATxPassedBlack	0.234 (0.175)	0.206 (0.177)	0.214 (0.185)	0.201 (0.194)	0.207 (0.202)
Costly Sorting		-0.0949 (0.0731)	-0.0977 (0.0730)	-0.0960 (0.0721)	-0.102 (0.0722)
Dicator Controls			X		X
Receiver Controls				X	X
Observations	126	126	126	126	126

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Source: Author's calculation

Figure 6: Sorting When Dictator is Biased



Source: Author's illustration

( $\chi^2 = 0.623, p = 0.430$ ), and when passing to someone who's race you are biased against ( $\chi^2 = 0.707, p = 0.400$ ), As such it is also unlikely that implicit bias is influencing giving on the extensive margin, inclusive of sorting decisions.

**Result 3** *Amount of Implicit Bias (as indicated by IAT D-score) does not predict sorting in or out of the dictator game*

## VII Small Gifts, a Robustness Check

Thus far, I have suggested that the IAT does not predict giving or sorting behaviors. However, I have also left the door open for dictators to have awareness of their biases, meta-cognitive abilities with respect to it, or both. This may suggest that differences in giving are more subtle than the ones suggested above. For instance, what if biased dictators are giving, but their giving is concentrated in small(er) gifts?

To test for this concentration, I utilize the *Dislike* variable from equation 9 above, noting that this variable highlights cases of both pro-white and anti-white bias. I also generate dummy variables for various small gift amounts. I then run Pearson's  $\chi^2$  tests to see if giving in those small amounts is different for biased and non-biased dictators in each of the pictures treatments. Full results from these tests are depicted in table 13.

Table 13: The IAT and Small Gifts

	p-value for Gift Size:			
	0	$\leq 1$	$\leq 2$	N
No Sorting	0.738	0.209	0.369	48
Sorting	0.646	0.319	0.968	127
Receiver is Black	0.367	0.256	0.647	130
Sorting & Receiver is Black	0.310	0.181	0.753	94
Whole Sample	0.927	0.893	0.496	175

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Source: Author's calculation

Small giving is not different between biased and unbiased dictators in every specification. This result suggests that biased giving is not concentrated in small giving, and lends further credence to the above discussion of dictator giving as a whole.

**Result 4** *Biased givers are no more likely to give small ( $\leq \$2$ ) gifts*



## VIII Conclusion

Racial bias is a persistent concern in the social sciences. In the past two decades, a proposed method of detecting it, the Implicit Association Test, has caught fire amongst the academics who study bias. At the time of this writing, the IAT's original paper has over 6600 citations, with researchers claiming it has implications on all sorts of economic outcomes, from workplace discrimination and managerial behavior, to egalitarian ideals and general social welfare. Yet, economists have only recently started to explore these claims in detail.

In this paper I have undertaken an in-depth examination of one of those claims in particular—that implicit (racial) bias is a predictor of pro-social behavior. I focused on these behaviors due to a growing literature suggesting the importance of the relationship between bias and pro-sociality. In doing so, I critique the extant literature stemming from the IAT. I then write a model of giving under conditions of implicit bias and conduct a laboratory experiment to test those model predictions.

Specifically, I test biased giving using a dictator game where acts of giving are both non-strategic and non-spontaneous, and therefore easily controlled (by the subject). Additionally, in some treatments I include a sorting (exit) option to see if biased givers simply choose to avoid the potential giving transactions altogether.

I find that, contrary to model predictions and previous literature, implicit bias fails to predict giving on both the extensive and intensive margins. That is, not only does implicit bias not predict amounts shared in the dictator game, it also does not predict examples of zero sharing, or the choice to exit a giving environments. Furthermore, these results hold not only in fine bins of analysis, but also wider and more powerful ones, such as when I restrict my sample to small gifts, or dictators paired with receivers of the race they hold implicit biases against.

To the best of my knowledge, this is the first paper to explore the implications of a Race IAT in an economics experiment. As such, the analysis in this paper represents a necessary step forward in this line of research that previously consisted of compelling, but unsubstantiated claims.

The dictator game is a compelling example in that it consists of a very simple economic decision. If the IAT fails to map into this class of *cold-phase* decisions, what are the implications for decisions which may be more complex but also require more deliberation, such as hiring?

However, more research is needed as the dictator game is also a very clear-cut decision, and perhaps the IAT could be better used to predict maps into

so-called *fuzzier* or multi-level economic decisions, decisions made in groups or ones where the use of heuristics have been shown to play a prominent role.

In this vein, we might think of implicit bias as mapping into a spectrum of pro-social activities with dictator giving at one end of the spectrum and a potentially different result at the other. If this is true, then future field experiments could prove to be a fruitful area of research.

Finally, as the popularity of the IAT grows in academia, so does its use in the public domain. As such, this paper also speaks to policy in general, and jurisprudence in particular. The typical anti-discrimination statute requires proof that harmful actions were "because of" discrimination. More and more, implicit bias is being recognized as a source of this liability. For instance, in a recent Supreme Court case regarding the Fair Housing Act, Chief Justice Roberts wrote for the majority that:

Recognition of disparate impact liability under the FHA also plays a role in uncovering discriminatory intent: It permits plaintiffs to counteract *unconscious prejudices and disguised animus* that escape easy classification as disparate treatment. In this way disparate-impact liability may prevent segregated housing patterns that might otherwise result from covert and illicit stereotyping.  
(*Texas DoH v. ICP Inc.*, 2015)

Italics are my own. What this means is that bias can be classified under the law as resulting in differential treatment even if one is not aware of the held bias, as in an implicit bias. And hence, we need further explorations of implicit bias and its potential to map into this sort of decision making, else we could be establishing ineffective policies.

## References

- Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc. 576 U.S. Supreme Court, 2015.
- Joseph G Altonji and Rebecca M Blank. Race and gender in the labor market. *Handbook of labor economics*, 3:3143–3259, 1999.
- Kenneth Arrow et al. The theory of discrimination. *Discrimination in labor markets*, 3(10):3–33, 1973.
- Gary S Becker. *The economics of discrimination*. University of Chicago press, 1957.

- Marianne Bertrand and Sendhil Mullainathan. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review*, 94(4):991–1013, 2004.
- Marianne Bertrand, Dolly Chugh, and Sendhil Mullainathan. Implicit discrimination. *American Economic Review*, pages 94–98, 2005.
- Lex Borghans, Angela Lee Duckworth, James J Heckman, and Bas Ter Weel. The economics and psychology of personality traits. *Journal of Human Resources*, 43(4):972–1059, 2008.
- Tomas Broberg, Tore Ellingsen, and Magnus Johannesson. Is generosity involuntary? *Economics Letters*, 94(1):32–37, 2007.
- Jamie Brown-Kruse and David Hummels. Gender effects in laboratory public goods contribution: Do individuals put their money where their mouth is? *Journal of Economic Behavior & Organization*, 22(3):255–267, 1993.
- Crosby Burns. The costly business of discrimination. *Washington: Center for American Progress*, 2012.
- C Bram Cadsby and Elizabeth Maynes. Gender and free riding in a threshold public goods game: experimental evidence. *Journal of economic behavior & organization*, 34(4):603–620, 1998.
- Colin Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2003.
- Marco Castillo and Ragan Petrie. Discrimination in the lab: Does information trump appearance? *Games and Economic Behavior*, 68(1):50–59, 2010.
- Kerwin Kofi Charles and Jonathan Guryan. Studying discrimination: Fundamental challenges and recent progress. *Annu. Rev. Econ.*, 3(1):479–511, 2011.
- Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.
- John G Cragg. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, pages 829–844, 1971.

- Jason Dana, Daylian M Cain, and Robyn M Dawes. What you dont know wont hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2):193–201, 2006.
- Catherine C Eckel and Ragan Petrie. Face value. *The American Economic Review*, pages 1497–1513, 2011.
- Paul J Ferraro and Ronald G Cummings. Cultural diversity, discrimination, and economic outcomes: an experimental analysis. *Economic Inquiry*, 45(2):217–232, 2007.
- Chaim Fershtman and Uri Gneezy. Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, pages 351–377, 2001.
- Klaus Fiedler and Matthias Bluemke. Faking the iat: Aided and unaided response control on the implicit association tests. *Basic and Applied Social Psychology*, 27(4):307–316, 2005.
- Uri Gneezy, John List, and Michael K Price. Toward an understanding of why people discriminate: Evidence from a series of natural field experiments. Technical report, National Bureau of Economic Research, 2012.
- Anthony G Greenwald, Debbie E McGhee, and Jordan LK Schwartz. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6):1464, 1998.
- Anthony G Greenwald, Brian A Nosek, and Mahzarin R Banaji. Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of personality and social psychology*, 85(2):197, 2003.
- Anthony G Greenwald, T Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R Banaji. Understanding and using the implicit association test: Iii. meta-analysis of predictive validity. *Journal of personality and social psychology*, 97(1):17, 2009.
- David M Grether and Charles R Plott. Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, pages 623–638, 1979.
- Andrew Hanson and Zackary Hawley. Do landlords discriminate in the rental housing market? evidence from an internet field experiment in us cities. *Journal of Urban Economics*, 70(2):99–114, 2011.

- Andrew Hanson, Zackary Hawley, and Aryn Taylor. Subtle discrimination in the rental housing market: Evidence from e-mail correspondence with landlords. *Journal of Housing Economics*, 20(4):276–284, 2011.
- James J Heckman. Detecting discrimination. *The Journal of Economic Perspectives*, pages 101–116, 1998.
- Charles A Holt and Susan K Laury. Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655, 2002.
- John T Jost, Laurie A Rudman, Irene V Blair, Dana R Carney, Nilanjana Dasgupta, Jack Glaser, and Curtis D Hardin. The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in organizational behavior*, 29:39–69, 2009.
- Jerry Kang. *Implicit bias: A primer for courts*. National Center for State Courts, 2009.
- Edward P Lazear, Ulrike Malmendier, and Roberto A Weber. Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, 4(1):136–163, 2012.
- Daniel J. Lee. Risk aversion and implicit bias. Mimeo.
- John A List. The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics*, pages 49–89, 2004.
- Sara Lowes, Nathan Nunn, James A Robinson, and Jonathan Weigel. Understanding ethnic identity in africa: Evidence from the implicit association test (iat). *American Economic Review*, 105(5):340–45, 2015.
- Adam W Meade. Freeiat: An open-source program to administer the implicit association test. *Applied psychological measurement*, 33(8):643–643, 2009.
- James Mill. *Analysis of the Phenomena of the Human Mind*, volume 1. London: Longmans, Green, Reader and Dyer, 1869.
- David Neumark. Detecting discrimination in audit and correspondence studies. *Journal of Human Resources*, 47(4):1128–1157, 2012.

- Michael I Norton, Malia F Mason, Joseph A Vandello, Andrew Biga, and Rebecca Dyer. An fmri investigation of racial paralysis. *Social cognitive and affective neuroscience*, page nss010, 2012.
- Brian A Nosek and Jeffrey J Hansen. The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition & Emotion*, 22(4):553–594, 2008.
- Brian A Nosek, Mahzarin Banaji, and Anthony G Greenwald. Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, 6(1):101, 2002.
- Edmund S Phelps. The statistical theory of racism and sexism. *The american economic review*, pages 659–661, 1972.
- Joseph Price and Justin Wolfers. Racial discrimination among nba referees. *The Quarterly journal of economics*, 125(4):1859–1887, 2010.
- Ernesto Reuben, Paola Sapienza, and Luigi Zingales. How stereotypes impair womens careers in science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408, 2014.
- Dan-Olof Rooth. Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3):523–534, 2010.
- Bradley J Ruffle and Richard Sosis. Cooperation and the in-group-out-group bias: A field test on israeli kibbutz members and city residents. *Journal of Economic Behavior & Organization*, 60(2):147–163, 2006.
- Robert Slonim and Pablo Guillen. Gender selection discrimination: Evidence from a trust game. *Journal of Economic Behavior & Organization*, 76(2):385–405, 2010.
- John L Solow and Nicole Kirkwood. Group identity and gender in public goods experiments. *Journal of Economic Behavior & Organization*, 48(4):403–412, 2002.
- Jennifer Triplett. Racial bias and prosocial behavior. *Sociology Compass*, 6(1):86–96, 2012.
- John Yinger. Evidence on discrimination in consumer markets. *The Journal of Economic Perspectives*, pages 23–40, 1998.

## Appendix A) IAT Screenshots

### *On Screen Instructions*

**African American****European American**


Put your middle or index fingers on the **E** and **I** keys of your keyboard. Words or images representing the categories at the top will appear one -by-one in the middle of the screen. When the item belongs to a category on the left, press the **E** key; when the item belongs to a category on the right, press the **I** key. Items belong to only one category. If you make an error, an **X** will appear - fix the error by hitting the other key.

This is a timed sorting task. **GO AS FAST AS YOU CAN** while making as few mistakes as possible. Going too slow or making too many errors will result in an uninterpretable score. This task will take about 5 minutes to complete.

Press the **space bar** to begin.

### *Concept (Facial) Sorting*

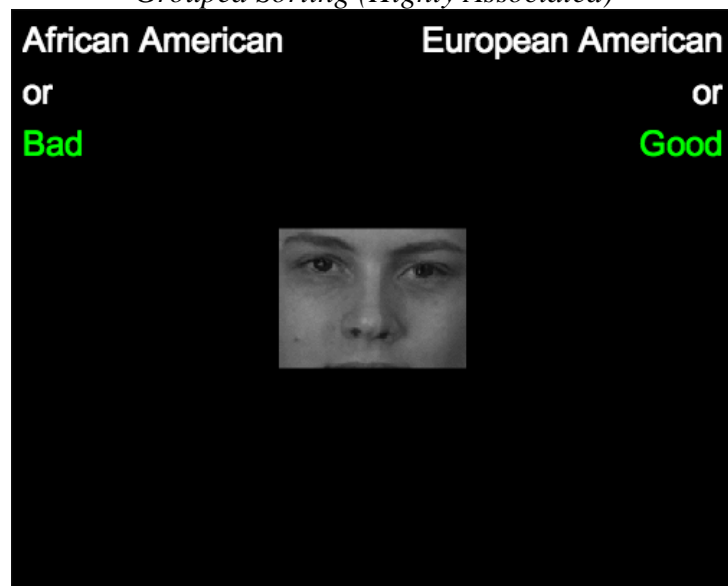
**African American****European American**



*Attribute (Word) Sorting*

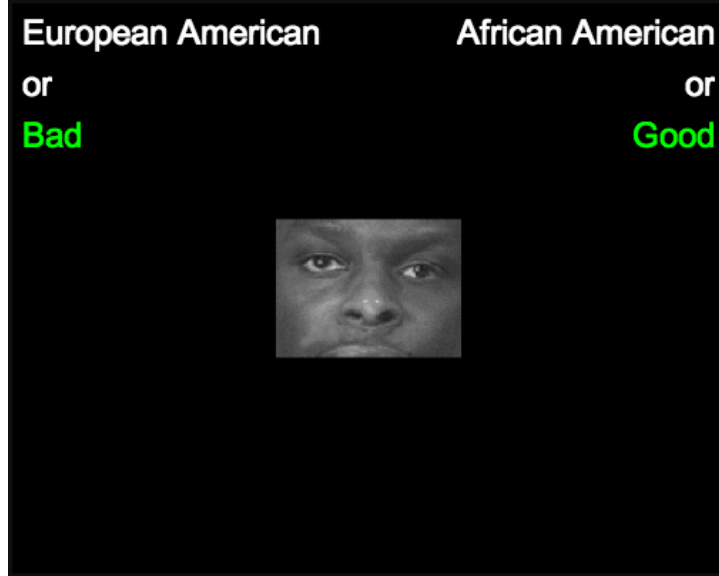


*Grouped Sorting (Highly Associated)*





*Grouped Sorting (Less Associated)*



## Appendix B

In this appendix, I adapt the language of Lazear et al. (2012) and include formal definitions of Reluctant, Willing, and Non-sharers

**Definition 1.** A **Willing Sharer** (i) shares a positive amount in a sharing environment and (ii) prefers to be in such an environment when  $w = w'$ .

$$(i) \arg \max_{x \in [0, w]} U(1, x, w - x) < w$$

$$(ii) \max_{x \in [0, w]} U(1, x, w - x) > U(0, w, 0)$$

**Definition 2.** A **Reluctant Sharer** (i) shares a positive amount when in a sharing environment but (ii) prefers to not have the option when there is no financial reward to sharing.

$$(i) \arg \max_{x \in [0, w]} U(1, x, w - x) < w$$

$$(ii) \max_{x \in [0, w]} U(1, x, w - x) < U(0, w, 0)$$

**Definition 3.** A **Non-Sharer** (i) does not share, even if the environment allows for it.

$$(i) \arg \max_{x \in [0, w]} U(1, x, w - x) = w$$

Appendix C) Subject Instructions  
*Treatment: Pictures, Costly Sorting*

Instructions

Thank you for agreeing to participate. This is an experiment in two parts. We are interested in how people make decisions in social situations. Please read the instructions carefully, as your task may not be the same as those around you. During the session please do not talk or communicate with the other participants. If you have a question, please raise your hand and a research assistant will come answer it.

Everyone has already earned \$5 for showing up. Additionally, you may have an opportunity to earn more. We will pay you privately in cash at the end of the session. None of the other participants will know the amount you have earned.

Group A Instructions

In the first part of this experiment, you have been given the choice of whether or not to participate in the following activity. That is, participating in this activity is optional.

You have been randomly paired with the participant displayed on your screen. This person is completing a different task that may include different payments, and does not know that he or she is participating with you. If you choose to participate, you will be given \$10. It is your task to decide how much to distribute between yourself and the person with whom you are paired. In other words, you must decide how much money, between \$0 and \$10 to give to the other person and how much to keep for yourself. You may select any amount between \$0 and \$10. For example, you may decide to give \$9 to the other person and keep \$1 for yourself, or you may instead decide to give \$1 to the other person and keep \$9 for yourself. If you choose to participate, I will explain the activity to the other person. That is, the other person will learn the rules of the allocation task and the assigned amounts you assigned. He or she will not see your picture. The assigned amounts will then be paid to both you, in addition to your show-up fees.

Moreover, you may decide to not participate in the above activity. If you choose this option, you will receive a fixed amount of \$9 (plus the \$5 for participation). The other person will receive \$5 for participation. He or she will not receive any information about this activity. Please indicate your choice on the sheet below.

Decision Sheet

I wish to (circle one) Participate/Not Participate

If you are participating, please indicate

Amount of money to give to the other person: \_\_\_\_\_

Amount of money to keep for yourself : \_\_\_\_\_

*(these two quantities must sum up to \$10.00)*

## *Treatment: Pictures, Costless Sorting*

### Instructions

Thank you for agreeing to participate. This is an experiment in two parts. We are interested in how people make decisions in social situations. Please read the instructions carefully, as your task may not be the same as those around you. During the session please do not talk or communicate with the other participants. If you have a question, please raise your hand and a research assistant will come answer it.

Everyone has already earned \$5 for showing up. Additionally, you may have an opportunity to earn more. We will pay you privately in cash at the end of the session. None of the other participants will know the amount you have earned.

### Group A Instructions

In the first part of this experiment, you have been given the choice of whether or not to participate in the following activity. That is, participating in this activity is optional.

You have been randomly paired with the participant displayed on your screen. This person is completing a different task that may include different payments, and does not know that he or she is participating with you. If you choose to participate, you will be given \$10. It is your task to decide how much to distribute between yourself and the person with whom you are paired. In other words, you must decide how much money, between \$0 and \$10 to give to the other person and how much to keep for yourself. You may select any amount between \$0 and \$10. For example, you may decide to give \$9 to the other person and keep \$1 for yourself, or you may instead decide to give \$1 to the other person and keep \$9 for yourself. If you choose to participate, I will explain the activity to the other person. That is, the other person will learn the rules of the allocation task and the assigned amounts you assigned. He or she will not see your picture. The assigned amounts will then be paid to both you, in addition to your show-up fees.

Moreover, you may decide to not participate in the above activity. If you choose this option, you will receive a fixed amount of \$10 (plus the \$5 for participation). The other person will receive \$5 for participation. He or she will not receive any information about this activity. Please indicate your choice on the sheet below.

### Decision Sheet

I wish to (circle one) Participate/Not Participate

If you are participating, please indicate

Amount of money to give to the other person: \_\_\_\_\_

Amount of money to keep for yourself : \_\_\_\_\_

*(these two quantities must sum up to \$10.00)*

### *Treatment Pictures, No Sorting*

#### Instructions

Thank you for agreeing to participate. This is an experiment in two parts. We are interested in how people make decisions in social situations. Please read the instructions carefully, as your task may not be the same as those around you. During the session please do not talk or communicate with the other participants. If you have a question, please raise your hand and a research assistant will come answer it.

Everyone has already earned \$5 for showing up. Additionally, you may have an opportunity to earn more. We will pay you privately in cash at the end of the session. None of the other participants will know the amount you have earned.

#### Group A Instructions

You have been randomly paired with the participant displayed on your screen. In this part of the experiment, you will be given \$10. It is your task to decide how much to distribute between yourself and the person with whom you are paired. In other words, you must decide how much money, between \$0 and \$10 to give to the other person and how much to keep for yourself. You may select any amount between \$0 and \$10. For example, you may decide to give \$9 to the other person and keep \$1 for yourself, or you may instead decide to give \$1 to the other person and keep \$9 for yourself. After you make your decision, I will explain the activity to the other person, that is, the other person will learn the rules of the allocation task and the assigned amounts you assigned. He or she will not see your picture. The assigned amounts will then be paid to both you, in addition to your show-up fees.

#### Decision Sheet

Amount of money to give to the other person: \_\_\_\_\_

Amount of money to keep for yourself : \_\_\_\_\_

*(these two quantities must sum up to \$10.00)*

*Treatment: No Information, Costly Sorting*

Instructions

Thank you for agreeing to participate. This is an experiment in two parts. We are interested in how people make decisions in social situations. Please read the instructions carefully, as your task may not be the same as those around you. During the session please do not talk or communicate with the other participants. If you have a question, please raise your hand and a research assistant will come answer it.

Everyone has already earned \$5 for showing up. Additionally, you may have an opportunity to earn more. We will pay you privately in cash at the end of the session. None of the other participants will know the amount you have earned.

Group A Instructions

In the first part of this experiment, you have been given the choice of whether or not to participate in the following activity. That is, participating in this activity is optional.

You have been randomly paired with a participant in this room. This person is completing a different task that may include different payments, and does not know that he or she is participating with you. If you choose to participate, you will be given \$10. It is your task to decide how much to distribute between yourself and the person with whom you are paired. In other words, you must decide how much money, between \$0 and \$10 to give to the other person and how much to keep for yourself. You may select any amount between \$0 and \$10. For example, you may decide to give \$9 to the other person and keep \$1 for yourself, or you may instead decide to give \$1 to the other person and keep \$9 for yourself. If you choose to participate, I will explain the activity to the other person. That is, the other person will learn the rules of the allocation task and the assigned amounts you assigned. He or she will not learn who you are. The assigned amounts will then be paid to both you, in addition to your show-up fees.

Moreover, you may decide to not participate in the above activity. If you choose this option, you will receive a fixed amount of \$9 (plus the \$5 for participation). The other person will receive \$5 for participation. He or she will not receive any information about this activity. Please indicate your choice on the sheet below.

Decision Sheet

I wish to (circle one) Participate/Not Participate

If you are participating, please indicate

Amount of money to give to the other person: \_\_\_\_\_

Amount of money to keep for yourself : \_\_\_\_\_

*(these two quantities must sum up to \$10.00)*

*Treatment: No Information, Costless Sorting*

Instructions

Thank you for agreeing to participate. This is an experiment in two parts. We are interested in how people make decisions in social situations. Please read the instructions carefully, as your task may not be the same as those around you. During the session please do not talk or communicate with the other participants. If you have a question, please raise your hand and a research assistant will come answer it.

Everyone has already earned \$5 for showing up. Additionally, you may have an opportunity to earn more. We will pay you privately in cash at the end of the session. None of the other participants will know the amount you have earned.

Group A Instructions

In the first part of this experiment, you have been given the choice of whether or not to participate in the following activity. That is, participating in this activity is optional.

You have been randomly paired with a participant in this room. This person is completing a different task that may include different payments, and does not know that he or she is participating with you. If you choose to participate, you will be given \$10. It is your task to decide how much to distribute between yourself and the person with whom you are paired. In other words, you must decide how much money, between \$0 and \$10 to give to the other person and how much to keep for yourself. You may select any amount between \$0 and \$10. For example, you may decide to give \$9 to the other person and keep \$1 for yourself, or you may instead decide to give \$1 to the other person and keep \$9 for yourself. If you choose to participate, I will explain the activity to the other person. That is, the other person will learn the rules of the allocation task and the assigned amounts you assigned. He or she will not learn who you are. The assigned amounts will then be paid to both you, in addition to your show-up fees.

Moreover, you may decide to not participate in the above activity. If you choose this option, you will receive a fixed amount of \$10 (plus the \$5 for participation). The other person will receive \$5 for participation. He or she will not receive any information about this activity. Please indicate your choice on the sheet below.

Decision Sheet

I wish to (circle one) Participate/Not Participate

If you are participating, please indicate

Amount of money to give to the other person: \_\_\_\_\_

Amount of money to keep for yourself : \_\_\_\_\_

*(these two quantities must sum up to \$10.00)*

*Treatment: No Information, No Sorting*

Instructions

Thank you for agreeing to participate. This is an experiment in two parts. We are interested in how people make decisions in social situations. Please read the instructions carefully, as your task may not be the same as those around you. During the session please do not talk or communicate with the other participants. If you have a question, please raise your hand and a research assistant will come answer it.

Everyone has already earned \$5 for showing up. Additionally, you may have an opportunity to earn more. We will pay you privately in cash at the end of the session. None of the other participants will know the amount you have earned.

Group A Instructions

You have been randomly paired with a participant in this room. In this part of the experiment, you will be given \$10. It is your task to decide how much to distribute between yourself and the person with whom you are paired. In other words, you must decide how much money, between \$0 and \$10 to give to the other person and how much to keep for yourself. You may select any amount between \$0 and \$10. For example, you may decide to give \$9 to the other person and keep \$1 for yourself, or you may instead decide to give \$1 to the other person and keep \$9 for yourself. After you make your decision, I will explain the activity to the other person, that is, the other person will learn the rules of the allocation task and the assigned amounts you assigned. He or she will not learn who you are. The assigned amounts will then be paid to both you, in addition to your show-up fees.

Decision Sheet

Amount of money to give to the other person: \_\_\_\_\_

Amount of money to keep for yourself : \_\_\_\_\_

*(these two quantities must sum up to \$10.00)*



### *Receiver Instructions (Constant Across Treatments)*

#### Instructions

Thank you for agreeing to participate. This is an experiment in two parts. We are interested in how people make decisions in social situations. Please read the instructions carefully, as your task may not be the same as those around you. During the session please do not talk or communicate with the other participants. If you have a question, please raise your hand and a research assistant will come answer it.

Everyone has already earned \$5 for showing up. Additionally, you may have an opportunity to earn more. We will pay you privately in cash at the end of the session. None of the other participants will know the amount you have earned.

#### Group B Instructions

In the first part of this experiment, you are asked to complete the attached questionnaire. You will earn money based on how you answer these questions. After finishing, you will be asked to participate in an additional activity. The additional activity will not affect your payment in this part of the experiment.

In this questionnaire, you will be presented with a table that contains information on 10 different decisions that you must make. For each of the 10 decisions you must select either option 1 or option 2. The outcome of each option depends on the role of a 10-sided die. You will be paid based on your decisions in this questionnaire and partly on chance. Below is an example of the first three decisions you will make:

Decision			Option 1			Option 2		
1	Roll	1 for \$2	or	2-10 for \$1.60	Roll	1 for \$3.85	or	2-10 for \$0.10
2	Roll	1,2 for \$2	or	3-10 for \$1.60	Roll	1,2 for \$3.85	or	3-10 for \$0.10
3	Roll	1-3 for \$2	or	4-10 for \$1.60	Roll	1-3 for \$3.85	or	4-10 for \$0.10

Here is how I will pay you for this activity: I will first roll the 10-sided die to determine which decisions will receive payment and then re-roll the 10-sided die to determine your final earnings based on whether or not you selected option 1 or 2. All die rolls will be conducted after you have completed the experiment.

In the example above, suppose that I roll the 10-sided die and it lands on 1. Then the first row will be selected for payment. Now supposed I reroll the die and it lands on 6. If this is true you will receive \$1.60 if you had selected Option 1 and \$0.10 if you had selected Option 2. However, if the 10-sided die lands on 1 you will receive \$2.00 if you had selected Option 1 and \$3.85 if you had selected Option 2.

**Please indicate your decision for each of the 10 rows on the opposite side of this sheet:**

Please indicate your choice by circling either option 1 or 2 in the far right column. Only choose one option for each decision:

Decision		Option 1				Option 2				My Choice	
<b>1</b>	Roll	1 for \$2	or	2-10 for \$1.60	Roll	1 for \$3.85	or	2-10 for \$0.10	Option 1	Option 2	
<b>2</b>	Roll	1,2 for \$2	or	3-10 for \$1.60	Roll	1,2 for \$3.85	or	3-10 for \$0.10	Option 1	Option 2	
<b>3</b>	Roll	1-3 for \$2	or	4-10 for \$1.60	Roll	1-3 for \$3.85	or	4-10 for \$0.10	Option 1	Option 2	
<b>4</b>	Roll	1-4 for \$2	or	5-10 for \$1.60	Roll	1-4 for \$3.85	or	5-10 for \$0.10	Option 1	Option 2	
<b>5</b>	Roll	1-5 for \$2	or	6-10 for \$1.60	Roll	1-5 for \$3.85	or	6-10 for \$0.10	Option 1	Option 2	
<b>6</b>	Roll	1-6 for \$2	or	7-10 for \$1.60	Roll	1-6 for \$3.85	or	7-10 for \$0.10	Option 1	Option 2	
<b>7</b>	Roll	1-7 for \$2	or	8-10for \$1.60	Roll	1-7 for \$3.85	or	8-10for \$0.10	Option 1	Option 2	
<b>8</b>	Roll	1-8 for \$2	or	9,10 for \$1.60	Roll	1-8 for \$3.85	or	9,10 for \$0.10	Option 1	Option 2	
<b>9</b>	Roll	1-9 for \$2	or	10 for \$1.60	Roll	1-9 for \$3.85	or	10 for \$0.10	Option 1	Option 2	
<b>10</b>	Roll	1-10 for \$2	or	- for \$1.60	Roll	1-10 for \$3.85	or	- for \$0.10	Option 1	Option 2	

## Appendix D) Demographic Survey

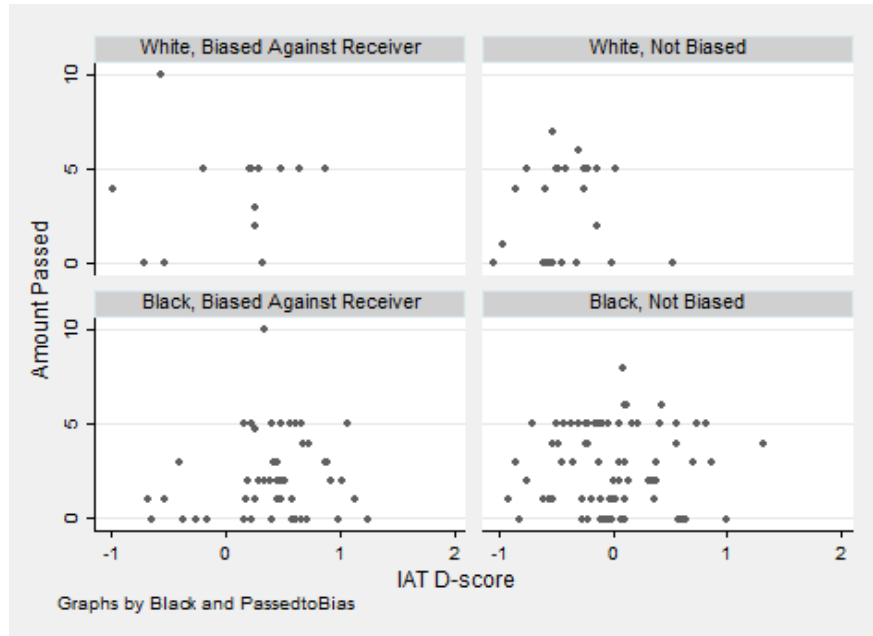
### Demographic Survey

Below are several questions relating to your background. Your answers here will help us in conducting statistical analysis. Your name will not be matched with your responses and all information will be kept confidential. Please indicate if you prefer not to answer a particular question or if you would like to leave the study at any time. Please answer the questions honestly and to the best of your ability.

- 1) What is your age? \_\_\_\_\_
- 2) What gender do you identify with:
  - ☐ Male
  - ☐ Female
  - ☐ Prefer Not to Answer
- 3) Which of these groups best describes you?
  - ☐ White
  - ☐ Black or African-American
  - ☐ Hispanic
  - ☐ American Indian or Alaska Native
  - ☐ Asian
  - ☐ Native Hawaiian or Other Pacific Islander
  - ☐ Other
  - ☐ Prefer Not to Answer
- 4) What religion do you currently identify with?
  - ☐ Catholic
  - ☐ Protestant
  - ☐ Muslim
  - ☐ Jewish
  - ☐ Agnostic
  - ☐ No Religion
  - ☐ Don't Know
  - ☐ Prefer Not to Answer
  - ☐ Other
- 5) Have you participated in an economics experiment previously?
  - ☐ Yes
  - ☐ No
  - ☐ Don't Know
  - ☐ Prefer Not to Answer
- 6) What is your current year in school?
  - ☐ Freshman
  - ☐ Sophomore
  - ☐ Junior
  - ☐ Senior
  - ☐ Graduate Student
  - ☐ I am not currently enrolled in school
  - ☐ Prefer Not to Answer
- 7) What is your GPA?
  - ☐ \_\_\_\_\_
  - ☐ Prefer Not to Answer
- 8) What is your Major?  
\_\_\_\_\_

# Appendix E

Scatter plots of IAT score and Amount Passed by Race and Bias of Dictator



In this appendix we start by looking at giving in finer bins in the photo treatments. Specifically the bias of the dictator. Here our definition of bias is IAT scores beyond  $\pm 0.15$ . The greatest difference in means exists between passing to the same and other for those holding an Anti-Black bias. This difference is not significant.

Table 1: Average Amounts Passed By Bias and Equivalence of Race

		IAT Threshlod		
		Anti-White	None	Anti-Black
Same Race		2.964	2.179	2.516
		(1.971, n=28)	(2.342, n=28)	(2.206, n=48)
Other Race		2.742	2.600	3.233
		(2.756, n=31)	(2.591, n=10)	(1.960, n=30)
Avg. Pass		2.847	2.289	2.792
Total Obs		59	38	78

Std. Deviations & Observations in Parentheses

IAT cutoffs at bias thresholds of  $\leq -0.15$  and  $\geq 0.15$

Next we look at giving in the context of a double-hurdle model (Cragg, 1971). This more flexible model, displayed in table 2 allows two separate processes for choosing to stay in, and how much one participates conditional on staying in. However, in order to specify these

processes, it is necessary that I restrict the sample to those sessions with a sorting option (n=126). Even with this increased flexibility, IAT is still neither a significant predictor of dictator giving nor sorting.

Table 2: The IAT's Effect on Percent Shared Hurdle Model

Panel A: Hurdle				
Variable	(1)	(2)	(3)	(4)
IAT D-score	0.0584 (0.116)	0.107 (0.118)	0.108 (0.119)	0.154 (0.226)
Receiver is Black		0.340 (0.271)	0.338 (0.271)	0.350 (0.276)
Receiver is Same Gender		-0.559** (0.253)	-0.558** (0.253)	-0.561** (0.253)
Costly Sorting			-0.0440 (0.236)	-0.0535 (0.239)
IATxPassedBlack				-0.129 (0.533)
Constant	0.428*** (0.116)	0.525** (0.257)	0.550* (0.290)	0.552* (0.290)
Panel B: Above				
Variable	(1)	(2)	(3)	(4)
IAT D-score	-0.363 (0.311)	-0.383 (0.303)	-0.381 (0.300)	-0.976 (0.611)
Receiver is Black		-0.487 (0.696)	-0.554 (0.692)	-0.735 (0.702)
Receiver is Same Gender		0.878 (0.606)	0.922 (0.600)	1.029* (0.604)
Costly Sorting			-0.589 (0.570)	-0.609 (0.564)
IATxPassedBlack				1.600 (1.412)
Constant	3.251*** (0.345)	3.167*** (0.636)	3.510*** (0.696)	3.555*** (0.688)
Sigma	2.306*** (0.264)	2.265*** (0.256)	2.246*** (0.252)	2.223*** (0.247)
Observations	126	126	126	126

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1