

Article

# Does Implicit Bias Predict Dictator Giving?

Daniel J. Lee 

Jones Graduate School of Business, Rice University, 6100 Main Street, Houston, TX 77005-1827, USA; dlee55@rice.edu; Tel.: +1-713-348-2466

Received: 1 August 2018; Accepted: 19 September 2018; Published: 21 September 2018

**Abstract:** Implicit associations and biases are carried without awareness or conscious direction, yet there is reason to believe they may be influenced by social pressures. In this paper, I study social pressure as a motive to give, as well as giving itself under conditions of implicit bias. In doing so, I pair the Implicit Association Test (IAT), commonplace in other social sciences, with a laboratory dictator game with sorting. I find that despite its popularity, the IAT does not predict dictator giving and social pressure does not explain acts of giving from biased dictators. These results are indicative of the meaningful difference between having an implicit bias and acting on one. As such, results can be thought of as a bound on the external validity of the IAT.

**Keywords:** IAT; implicit bias; race; prosocial behavior; social pressures

**JEL Classification:** C91; D64; J15

---

## 1. Introduction

Varied organizations, such as police forces, universities and even Starbucks Coffee, have recently started implementing implicit bias trainings. This trend speaks to recent advances in the social sciences suggesting that racial animus stems from subtle cognitive processes, which we carry without awareness or conscious direction [1,2]. These processes are called implicit biases and associations. The Implicit Association Test (IAT, discussed further below) introduced a tractable measure of implicit biases without having to rely on unreliable self-reporting mechanisms, thus lending support to these suggestions. However, the extent to which IAT scores, and implicit biases more generally, are able to predict behavior is unknown.

In this paper, I use a laboratory experiment to examine implicit bias as a predictor of pro-social behaviors. These behaviors are economically important, and a growing body of work suggests a relationship between bias and giving [3,4]. Pro-social behaviors are both non-strategic and non-spontaneous, and therefore easily controlled by laboratory subjects. Accordingly, we can think of these behaviors as bounding the set of actions into which implicit bias can map. Additionally, I highlight social pressure as the potential reason this relationship may not be so straightforward, as social pressure can weaken the extent to which implicit bias maps into more costly economic behavior.

Specifically, I extend the design of Lazear et al. [5] (LMW) and implement a dictator game with sorting, paired with an IAT to study the effect implicit bias has on both the decision to give and social pressure as a motive for giving. I find that one's implicit bias predicts neither the decision of whether to give, nor the decision of how much to give.

I frame these questions in the context of motives of giving argument, which states that concerns other than altruism or fairness may impact giving [6]. A natural implication of implicit bias in this framework is that if social pressures are strong enough, spiteful dictators may still give. Additionally, I explore alternative avenues through which one's implicit bias can correlate to giving, such as a sorting option. This option allows (potentially biased) dictators to avoid these pressures by opting out of the task entirely. This approach allows me to better identify the biases of different sharers and how

those biases manifest in the market. The results remain strikingly consistent, and implicit bias does not predict sorting out of the giving environment, meaning that individuals with higher levels of implicit bias are no more likely to sort than those with lower levels of bias. These results help to rule out the attenuating effects social pressure may have on dictator game decisions, given a certain level of implicit bias.

This paper represents a necessary first step in this line of research. Through the lab environment, I am able to show how the provision of racial information affects the decision to act on a bias. Though not used to examine bias prior to this study, the motives for the giving argument and the particulars of a sorting design apply nicely to this line of inquiry. A notable feature of sorting designs is they allow us to identify the marginal giver [7]. Recent work in other domains (e.g., college admissions) has demonstrated that detecting bias requires comparing marginal rather than average entrants [8]. Further, people are unwilling to make judgments about individuals of different races [9]. It follows that implicit bias should influence the entry decision.

This paper contributes to the laboratory literature in several novel ways. First, it extends the research on sorting in dictator games by examining bias as a potential pathway of reluctance to share and social pressure as a potential moderator of that bias. This paper is unique in allowing different sorting options in a dictator game while also providing racial information of the recipients. In doing so, this paper also adds to the discussion on bias in the lab (notable examples of which are Fershtman and Gneezy [10], Ferraro and Cummings [11], Ben-Ner et al. [12] and Slonim and Guillen [13]). In comparing IAT scores to both the observed rates of giving and differential exits, I am able to provide convincing economic magnitudes on both the nature and direction of implicit bias. To the best of my knowledge, this is the first laboratory paper to examine the psychological pathways of racial bias by using the IAT. This represents an important step forward as different pathways may lead to different economic behaviors and therefore have different policy prescriptions.

However, the contribution of this paper is not intended to be limited to laboratory experiments. In fact, it dovetails with the nascent field work on discrimination and implicit bias, which mostly occurs outside the U.S. and has yet to find significant consensus. Price and Wolfers [14] argue implicit bias explains discriminatory behavior amongst NBA referees. In Africa, Lowes et al. [15] find evidence of ethnic homophily in the DRC capital, though Berge et al. [16] find little evidence of ethnic biases in Nairobi. In Europe, implicit bias has been predictive of negative hiring conditions [17] and job performance [18]. The alleged interaction between implicit bias and labor market decisions suggests a role for further economic analysis in other areas of decision-making, including the open questions of pro-social behavior.

## 2. Background

Since bias cannot be randomly assigned, the issue of how it can be measured remains open, particularly when subjects may be unaware of the biases they hold. Social psychologists Greenwald, McGhee and Schwartz first described implicit bias as automatic and analogized the mechanics of it to those of a reflex. As such, they claimed to be able to test for it using their “Implicit Association Test”, 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).<sup>2</sup> When instructions oblige highly associated categories (e.g., flower + pleasant) to share

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

<sup>2</sup> examples of pleasant words: joy, love, peace, wonderful, pleasure, glorious, laughter, happy; examples of unpleasant words: agony, terrible, horrible, nasty, evil, awful, failure, hurt.

a response key, performance is faster than when less associated categories (e.g., insect + pleasant) share a key. Greenwald et al. [19]

To sum, the IAT consists of four timed sorting tasks. In it, subjects match features (such as faces) to more and less 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 the word “furniture” than the word “food”, and hence more likely to be sorted faster as such. It is through this primitive of differential timing that one reveals her/her implicit biases.<sup>3</sup> Table 1 shows the the progression of IAT tasks; screen captures of which are presented in Supplementary Materials S1.

**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 [21].

The standard scoring metric for the IAT, the *D-Score*[22], is similar to Cohen’s measure for effect size, *d*. It is calculated in Equation (1) as the difference in mean latencies within test blocks divided by the standard deviation of latencies across test blocks. For the purposes of IAT scoring, only paired trials are considered test blocks.<sup>4</sup>

$$D = \frac{\bar{x}_{\text{stage 3}} - \bar{x}_{\text{stage 5}}}{SD_{\text{Stages 3 and 5}}}. \quad (1)$$

Accordingly, the *D-Score* can be either positive or negative. In the case of a black-white IAT, a positive score indicates a positive preference for whites, and vice versa for a negative score.<sup>5</sup> A score of or around zero indicates little or no preference. The authors further classify and interpret the IAT using Cohen’s [23] conventional measures for effect size, by binning *D-scores* at  $\pm 0.15, 0.35, 0.65$  for ‘slight’, ‘moderate’ and ‘strong’ associations, respectively. For the remainder of the paper, I will continue to use these bins to categorize IAT scores.

<sup>3</sup> It is important to note a key distinction here: 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 [20].

<sup>4</sup> See Stages 3 and 5 in Table 1.

<sup>5</sup> In this paper, I will continue to use the terminology of preferences so as to remain consistent with the literature.

Implicit bias testing has worthwhile aims. If efficacious, the IAT allows us to observe both the magnitude and direction of a personal bias, which people may not know or may be unwilling to divulge. Implicit bias appears to be profound, with some studies even showing extant anti-black biases among minority groups [24]. Oft cited examples of these biases in decision-making include decisions regarding managerial hiring or where to direct budget cuts [25,26].<sup>6</sup>

Meta analyses seem to illustrate that implicit biases are persistent [1,28], yet these claims are also contested.<sup>7</sup> For instance, Norton et al. [9] hypothesize that not wanting to appear biased (or wanting to appear race neutral) can cause a “race-paralysis” in IAT tasks, while Arkes and Tetlock [30] highlight non-prejudicial reasons for ‘failing’ IAT scores. Further, Forscher et al. [31] find that changing one’s IAT score does not bring about a behavioral change. A potential reason for these conflicting findings is that external forces, such as social pressures, may mitigate this bias. This reason has important implications for both lab and field results.

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. In other words, the question of interest is not whether implicit bias exists, but whether this bias predicts discrimination in marketplace behaviors, and if so, what are the dosage implications? That is, how much more do severe (as opposed to moderate or slight) levels of implicit bias impact economically relevant decisions, and how are they mitigated by outside forces?

Giving behaviors offer a simple first pass at these questions. Despite their apparent simplicity, the underlying motives for giving behaviors remain elusive [32]. For instance, Camerer [33] notes the relatively high rates (10–30%) of dictator giving in the lab. Reluctance to give in general [34] and social pressures specifically have been posited to explain not only this disparity between field and laboratory giving [6], but also discretionary behavior more generally [35]. Extant lab results support social pressure as a motive to give [36,37]. To this end, sorting environments paired with a dictator game allow for much cleaner identification of bias in pro-social behavior. Here, since the second player is passive, it is much more likely that differences in giving or sorting are due to bias rather than to strategic concerns.

Given the pervasiveness of implicit bias, what are its implications for both laboratory and naturally occurring behavior? I use the toolbox of experimental economics to see if (and to what extent) IAT performance is related to differential treatment of receivers in a dictator game. In doing so, I examine the IAT as a 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 a race (black-white) IAT. The race IAT has yet to be used in the economics literature, despite the fact that racial relations remain one of American society’s most divisive issues.

### 3. Experiment

#### 3.1. Procedures

In a standard dictator game, subjects are randomly split into receivers and dictators upon arriving at the lab.<sup>8</sup> A first mover is given \$10 and asked how much she/he would like to give to a paired (and passive) receiver. Her/his choice ends the game. Thus, giving in this game is non-strategic. I use this standard (no information) treatment to gauge dictator giving without information on the race of the recipient.

<sup>6</sup> For a more thorough review of the recent psychology and management literature regarding the IAT, see Jost et al. [27].

<sup>7</sup> The state of this debate was recently summarized in more detail by Singal [29].

<sup>8</sup> In untabulated results, I confirm demographic balance between dictator and receiver roles.

This experiment differs from a standard game in that some treatments employ a sorting environment. Specifically, in those treatments, I offer dictators an exit option as in LMW. In other words, dictators in those treatments 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.<sup>9</sup> This option can be either costly or free, where the costly option is necessarily payoff-dominated by at least one dictator game choice.

Finally, I run these treatments in two types of sessions: sessions with no information (anonymous) and sessions with pictures, where dictators can see with whom they are paired and use that picture as a proxy for race.<sup>10</sup> In later estimations, race is controlled for using self-reported survey data. 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 picture sessions and sorting treatments allows us to see whether implicit bias affects behavior on either the extensive or intensive margin; that is, the decision to engage in giving, as well as how much to give.

As such, this experiment necessitates a  $2 \times 3$  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. There is no significant difference in the dispersion of participant decisions between treatments (Bartlett’s test,  $p = 0.12$ ).

**Table 2.** Dictators by treatment.

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

After roles are assigned, the dictators are randomly paired with a receiver, and in the pictures treatments are first shown a picture of that receiver’s face. The pictures serve as a proxy for race. In the no information treatments, dictators are not informed about their receivers in any way. In all treatments, dictators are then explained the rules of the dictator game. In the sorting treatments, dictators are asked whether or not they choose to participate. In the event that a dictator elects to not participate (takes the exit option), that dictator’s receiver is not given any information about the allocation task, and the dictator is given the exit fee (\$9 or \$10, depending on treatment). Otherwise, the dictator decides how to allocate a sum of \$10 between herself/himself and her/his paired receiver.

Meanwhile, the receivers are passive in their role. They have their pictures taken (pictures treatment only), are guaranteed a show-up fee and asked to participate in a different task. In this case, that task is a real-money, 1xrisk-preference elicitation [39]. The receiver task is constant across treatments. The next task in the experiment is a race IAT (as described above), assessed on all subjects. I run this task second because the act of taking an IAT can prime someone to behave differently in a dictator task. However, knowing they have just participated in a dictator game should not influence the IAT score, which is difficult to fake or otherwise manipulate [40]. I then close by collecting demographic data in the form

<sup>9</sup> LMW term a subject who prefers to opt out, but gives in the absence of an exit option a “reluctant giver”. DellaVigna et al. [6] note that this pattern of behavior is consistent with a model of social pressure (Akerlof and Kranton [38]) where giving is utility-reducing for the subject.

<sup>10</sup> In picture treatments, all subjects (both dictator and receiver) are photographed. Upon consenting to participate, subjects are told they will take a digital photograph, which may be seen by others in the room and will take part in a decision-making task. They are also told not everyone has the same task.

of a survey and pay subjects privately. Complete subject instructions and survey questions can be found in Supplementary Materials S2 and S3, respectively.

### 3.2. 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. Overall, I ran 17 experimental sessions across the 6 treatments, with a roughly equal balance of subjects across treatment rows.<sup>11</sup>

In total, 228 dictators (i.e., 456 subjects) participated in the experiment. Table 3, Panel A, 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 in school is senior.<sup>12</sup> I strove for sessions to be racially balanced; however, this was not possible given the makeup of the subject pool. While at first blush, this racial imbalance may appear to be problematic, I believe it is not for several reasons, including past research on similar subject pools [41], the previous evidence on extant biases among minority groups [24], the one-on-one nature of the experimental design and the fact that ExCEN was selected particularly for its population of minority students. Robustness checks examining differences in cross-racial giving support these results.

**Table 3.** Experimental summary statistics (dictators only).

<b>Panel A: Demographics</b>				
<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	
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 EconMajor	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	
<b>Panel B: Tasks</b>				
<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	
Male Receivers	0.513	0.501	228	
Black Receivers	0.675	0.469	228	
Amount Passed	2.692	2.238	228	
Opted Out (Total)	0.186	0.389	167	
Opted Out (Costly)	0.148	0.357	81	
Opted Out (Costless)	0.202	0.404	99	
IAT D-score	0.054	0.495	225	

Similarly, Table 3, Panel B, 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 was costless. Rank sum tests showed that sorting significantly decreases sharing, even when sorting is costly (sorting:  $z = 2.146$ ,  $p < 0.05$ ; costly sorting:  $z = 2.370$ ,  $p < 0.05$ ). Full distributions of amounts passed are illustrated further in Figure 1. There is meaningful implicit bias in the sample. Over 44% of dictators have an IAT

<sup>11</sup> Given my power analysis and the fact that receiving in the no-information treatments is anonymous, I did not require as many subjects.

<sup>12</sup> This is perhaps an artifact of running a summer experiment, where both former juniors and recent grads identify as seniors.

score greater than or equal to 0.15, indicating a pro-white implicit bias. Consonant with past work, this bias is present and perhaps even stronger in subjects who identify as black, with a mean score of 0.162. I depict the distribution of these scores in Figure 2 for further exploration. Both the amounts of endowment passed and the distribution of IAT scores are consistent with extant results across a variety of subject pools.

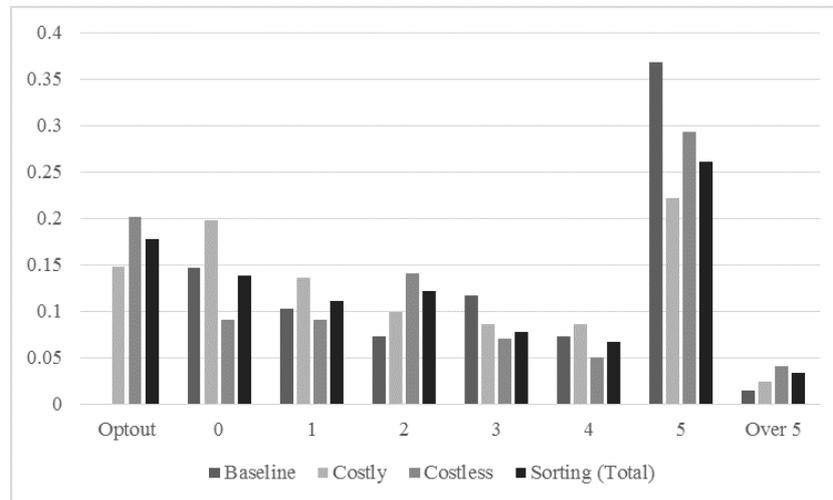


Figure 1. Distribution of amounts passed.

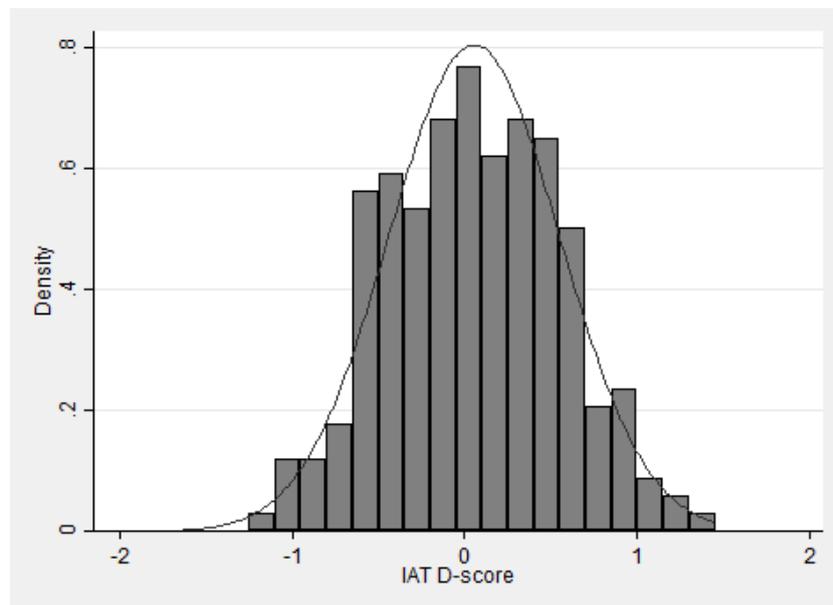


Figure 2. Distribution of IAT D-scores.

The current study reports the effects of implicit bias (as scored by the IAT) on behaviors in a modified dictator game. However, the main findings are that implicit bias is not predictive of behavior, and as such, many of the results that follow are null results.

Accordingly, I report the results of power analyses for differences in giving for different effect sizes. In the pictures treatments, 68 out of 175 dictators scored as biased against their receivers. At a 5% significance level, the power of finding a difference in giving of 1, 5 and 10 percent of one’s endowment is respectively 0.087, 0.41 and 0.88. A power of 0.8 is generally considered acceptable. Thus, this sample size is roughly large enough to detect differences in giving of 8% of one’s endowment. This is consistent with effect sizes found in Lane’s [42] meta-analysis of biased behavior in the economics lab. Further,

the findings in this paper are robust to a variety of specifications and controls. Regardless, one should always be cautious in interpreting non-significant results, and implications for future research are discussed further in the Conclusion.

#### 4. Discrete Results

We have seen descriptively that sorting environments affect giving behaviors, but what is the role of the implicit bias (as ostensibly measured by the IAT) in making these economic decisions? Table 4 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 endowments passed to black subjects to get smaller as favorable bias towards white subjects increases. Additionally, we would expect to see the opposite pattern for white recipients. However, these directional patterns did not emerge, particularly for black recipients. Here, those who had dictators biased against them ended up earning more on average.

**Table 4.** Average amount passed by IAT score and race of receiver.

Strength of Implicit Bias	Passed to:		
	Black	White	Anonymous
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

Given Table 4, I discretize the IAT score into the blunt question of “does one (implicitly) like or dislike the recipient?” and regress an outcome variable on two variables *Like* and *Dislike*. I express this question in Equation (2):

$$Outcome_i = \alpha_0 + \beta_1(LikeReceiver_j) + \beta_2(DislikeReceiver_j). \quad (2)$$

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 (3):

$$Like = \begin{cases} 1 & \text{when } IAT \geq 0.15 \text{ and receiver is white.} \\ 1 & \text{when } IAT \leq -0.15 \text{ and receiver is black.} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

$$Dislike = \begin{cases} 1 & \text{when } IAT \leq -0.15 \text{ and receiver is white.} \\ 1 & \text{when } IAT \geq 0.15 \text{ and receiver is black.} \\ 0 & \text{otherwise.} \end{cases}$$

That is, to “like” one’s receiver means either to hold a pro-white bias and pass to a white receiver or to hold a pro-black bias and pass to a black receiver, with “dislike” similarly defined. 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 5. In the first column, we find that unbiased givers shared about 23% of their endowment, and a bias against (or in favor of) one’s receiver did not significantly alter this giving pattern. Furthermore, both directions of bias remained insignificant when decomposing the *Like* and *Dislike* variables into component parts in Column 2.

**Table 5.** Discrete IAT estimations.

Variables	OLS, Percent Shared		Probit, Opted Out	
	(1)	(2)	(3)	(4)
Like Receiver	0.07 (0.047)		0.078 (0.336)	
Pro-White, White Receiver		0.053 (0.059)		−0.578 (0.568)
Pro-Black, Black Receiver		0.077 (0.051)		0.292 (0.358)
Dislike Receiver	0.029 (0.049)		0.166 (0.322)	
Pro-White, Black Receiver		0.042 (0.05)		0.096 (0.337)
Pro-Black, White Receiver		−0.029 (0.101)		0.456 (0.504)
Constant	0.229 *** (0.039)	0.229 *** (0.039)	−0.887 *** (0.257)	−0.887 *** (0.257)
Observations	172	172	126	126

Robust standard errors in parentheses. \*\*\*  $p < 0.01$ .

Table 5, Columns 3 and 4 repeat this exercise to 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 had any significant impact on remaining in the game.

While these results indicate that bias did not affect the decision giving on average, a relevant question is whether dictators who are biased against black (white) receivers behaved differently than the average dictator. Here, I restrict the sample based on the race of receiver and only regress on the *Dislike* variable. These results are shown in Table 6.

**Table 6.** Discrete estimations, conditional on the race of the receiver.

Variables	Black Receiver		White Receiver	
	(OLS)	(Probit)	(OLS)	(Probit)
Dislike Receiver	−0.003 (0.041)	−0.132 (0.29)	−0.073 (0.1)	0.952 * (0.576)
Constant	0.274 *** (0.027)	−0.659 *** (0.191)	0.273 *** (0.038)	−1.383 *** (0.374)
Observations	127	93	45	33

Robust standard errors in parentheses. \*\*\*  $p < 0.01$ , \*  $p < 0.1$ .

Even when we isolate the sample by race of receiver, biased dictators were not behaving in ways that were statistically different than the average dictator, nor was this difference economically significant in that the differences were less than pennies on the dollar. While the above results are indicative that subjects were not acting on their implicit bias in this context, another way of looking at the question is to ask what is the treatment effect of being paired with someone against whom you hold a bias?

Here, I exploit the random assignment of both roles and partners and implement propensity score matching. Since each person had a particular bias strength and direction, I matched on the strength and direction of this bias, as well as covariates describing the dictator's age, race and sex and specified the treatment as passing to a receiver against whom the dictator holds a bias; that is, passing to a black

receiver if the dictator holds a pro-white bias and vice versa. I found no significant treatment effect of passing to a receiver against whom the dictator is biased ( $ATT = 0.14, p = 0.625$ ).

**Result 1.** *The sign of IAT score did not predict dictator giving.*

## 5. Continuous Results

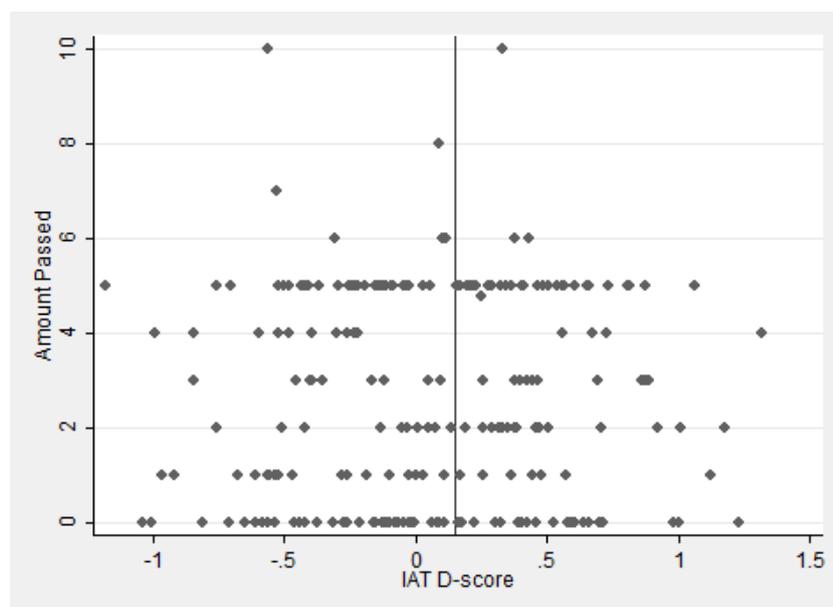
However, since the IAT is measured as a continuous variable, we can comment not only on the existence of implicit bias, but also the strength of that bias. As such, one would think that more severe biases would exert more influence on the giving and sorting decisions. To address this dosage question,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

$$Outcome_i = \beta_0 + \beta_1 IAT_i + \beta_2 (IAT_i * Race_j) + \beta' X + \varepsilon_i. \quad (4)$$

This standardization allowed me to interpret coefficients as the effect of a one standard deviation increase in IAT score. In this specification, I again regressed 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 the 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 assumption that implicit bias manifests as animus. The controls were necessary because observed differences in the outcome variable may have been driven by factors unrelated to a dictator's implicit bias. Different specifications below highlight different sets of these parameters in my analysis.

### 5.1. Dictator Giving

Figure 3 shows the amount passed given a dictator's IAT score. Despite the IAT's popularity in academic work, there was no clear linear relationship between IAT score and amount passed ( $\rho = -0.01$ ). Further, throughout the distribution of IAT scores, there appeared to be a similar bimodal pattern of amounts passed, suggesting that levels of implicit bias do not necessarily correlate with the behaviors of interest.



**Figure 3.** Scatter plot of IAT score and amount passed. Notes: The solid line indicates the IAT-D “bias” threshold of 0.15.

To confirm these findings econometrically, we turn to Table 7 which presents this paper's main estimates. In these models, I restricted the sample to only the pictures sessions. Additionally, I have used indicator variables for black dictators and black receivers, rather than a covariate for the race against which a subject is biased. While this may be a coarse measure, this modeling technique makes more sense in terms of coefficient interpretation since the IAT score is increasing in the level of anti-black bias. Further, these results are consistent with the discrete estimations from Section 4 and robust to the alternate specification of "biased against receiver".<sup>13</sup>

In Panel A of Table 7, I start in Model (1) with a simple OLS and regress percent shared on one's IAT score and the interaction term of  $IAT * passing$  to a black receiver, neither of which yielded a significant predictor of giving. These results held true in additional models, including specifications that control for race and gender of the dictator, the receiver and both. Further, these controls also had no significant effect on giving.

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

**Table 7.** The IAT's effect on percent shared.

<b>Panel A: OLS</b>					
<b>Variable</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>	<b>(5)</b>
IAT D-score	0.02 (0.035)	0.017 (0.035)	0.029 (0.045)	0.026 (0.039)	0.039 (0.037)
IATxPassedBlack	-0.062 (0.081)	-0.044 (0.08)	-0.05 (0.08)	-0.06 (0.087)	-0.067 (0.084)
Sorting Option		-0.101 *** (0.036)	-0.096 *** (0.037)	-0.094 *** (0.036)	-0.087 ** (0.038)
Dictator Controls			X		X
Receiver Controls				X	X
Constant	0.269 *** (0.018)	0.343 *** (0.028)	0.388 *** (0.051)	0.339 *** (0.046)	0.382 *** (0.064)
<b>Panel B: Tobit</b>					
<b>Variable</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>	<b>(5)</b>
IAT D-Score	0.03 (0.049)	0.025 (0.048)	0.036 (0.048)	0.043 (0.054)	0.055 (0.052)
IATxPassedBlack	-0.085 (0.110)	-0.057 (0.109)	-0.063 (0.109)	-0.088 (0.120)	-0.096 (0.117)
Sorting Option		-0.145 *** (0.045)	-0.138 *** (0.046)	-0.132 *** (0.045)	-0.122 ** (0.047)
Dictator Controls			X		X
Receiver Controls				X	X
Constant	0.222 *** (0.025)	0.327 *** (0.034)	0.366 *** (0.069)	0.323 *** (0.058)	0.359 *** (0.085)
Observations	172	172	172	172	172

Robust standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ .

<sup>13</sup> See, for instance, Tables 5 and 11.

Next, to account for the 27% of dictators who either gave nothing or opted out, I replicated the OLS results with a left-censored Tobit model.<sup>14</sup> These results are shown in Table 7, Panel B, and are not categorically different than the OLS results. That is, the coefficient on IAT score was positive, but insignificant; the interaction term was negative and insignificant; and all controls were insignificant. However, the presence of a sorting option was strongly and negatively significant.

Following LMW, I assess the determinants of sharing in Table 8. Specifically, I compare the relative importance of implicit bias (in Column 1) to the presence of the sorting option, as well as to self-reported demographics that could potentially affect sharing (in Column 2). Again, one's amount of implicit bias did not significantly determine sharing. Magnitudes of these results were similar when I ran the full model, including the IAT score with demographic controls (Column 3). Additionally, I calculated coefficients of partial determination.<sup>15</sup> This measure shows that not only did implicit bias lack statistical significance, but one's IAT score accounted for less than 4% of the unexplained variance and lacked economic significance, as well.

**Table 8.** Determinants of sharing.

Variable	Partial			R <sup>2</sup> 's
	(1)	(2)	(3)	
IAT D-score	−0.003 (0.017)		0.009 (0.02)	0.036
Sorting Option		−0.085 ** (0.042)	−0.083 * (0.045)	0.158
Age		0.006 *** (0.003)	0.006 ** (0.003)	0.148
Male		0.004 (0.041)	−0.002 (0.043)	0.005
Black		−0.099 * (0.05)	−0.107 ** (0.052)	0.185
Catholic		0.00 (0.071)	0.01 (0.076)	0.012
Previous Experience		−0.026 (0.047)	−0.028 (0.047)	0.047
Major: Business or Econ		0.004 (0.042)	0.006 (0.043)	0.012
GPA		−0.112 *** (0.041)	−0.110 *** (0.041)	0.219
Constant	0.268 *** (0.018)	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. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

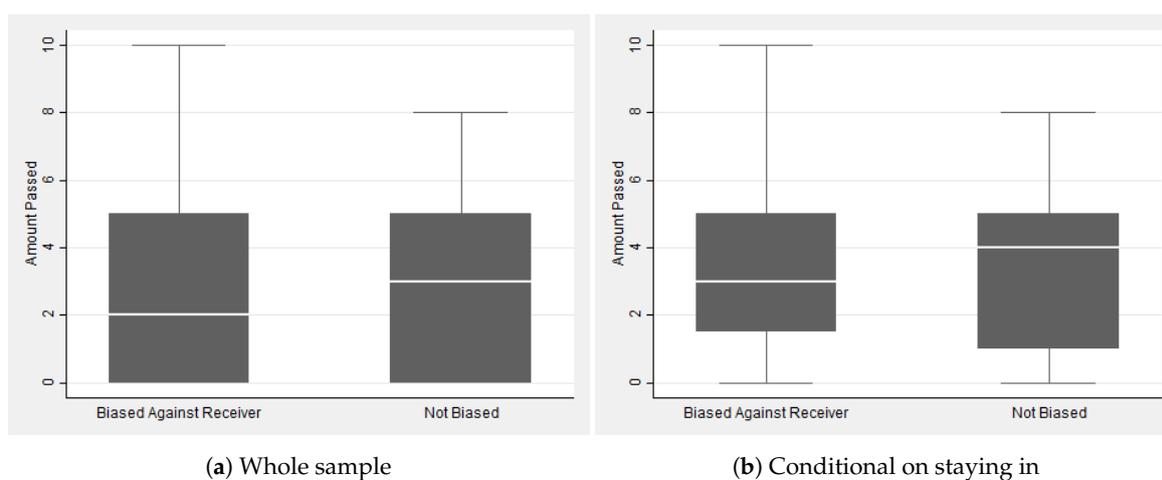
The above exercises held true when, instead of looking at the coarse measure of race of receiver, I looked at the finer measure of bias against one's receiver. In Figure 4, I graph box plots for a further analysis of what happens when a dictator is biased against the race of the receiver. For these figures, 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 ( $Receiver = Black | IAT \leq -0.15$ ).

<sup>14</sup> Robust standard errors were calculated using jackknife estimation. A double-hurdle model [43] 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 from this model are presented in Supplementary Materials S4.

<sup>15</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.

Clearly, there is no difference in giving when I considered the whole sample in Figure 4a. However, this result also held when I considered 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.

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



**Figure 4.** Sharing when the dictator is biased.

**Result 2.** *The magnitude of IAT score did not predict dictator giving.*

### 5.2. Dictator Sorting

Perhaps the above results indicate 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 opting 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 was descriptively smaller amongst those who stayed in (as compared to those who opted out), whereas in costless sorting, the mean IAT score was essentially the same. However, in both cases, this difference was 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 9. This model uses a probit regression and necessarily restricts the sample only to 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 the IAT score. In this model, there were no significant coefficients, suggesting that overall, one's IAT score did 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.

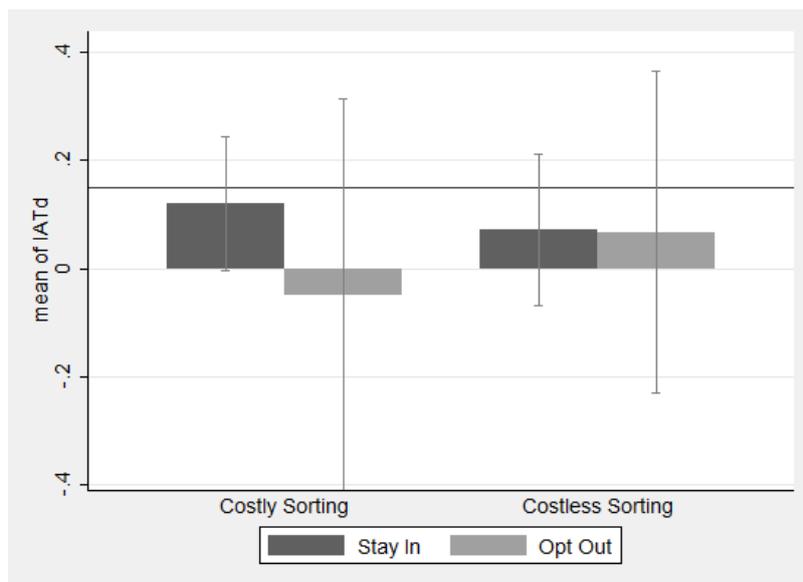


Figure 5. Bar graph of IAT scores and sorting.

Table 9. 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. \*\*\*  $p < 0.01$ .

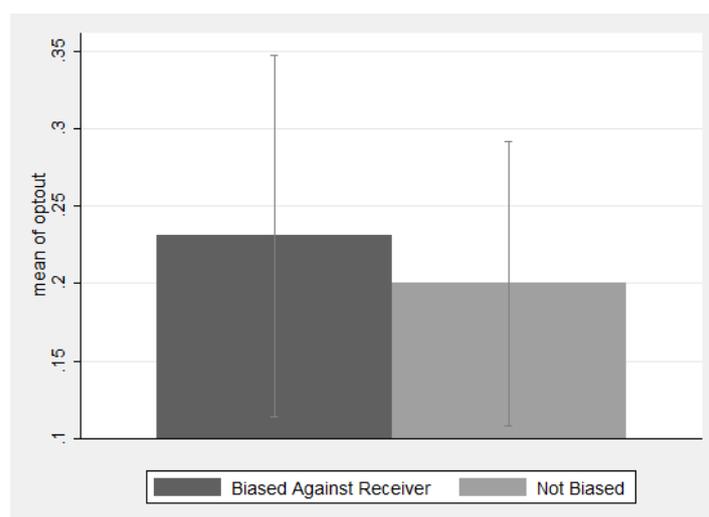
Nonetheless, this exploration again calls for a deeper analysis. Following Equation (4), I looked at the econometric results to confirm. In this case, I ignored the anonymous treatments ( $n = 127$ ) and ran probit estimations to determine what effect (if any) IAT score had on the probability of opting out. Table 10 shows the marginal effects of these estimations. Consistent with the results above, the IAT had no significant effect on sorting. This held when I controlled for whether the sorting was costly and the race and gender of the dictator, receiver and both. Similar to the analysis under dictator giving, the signs of these coefficients were also unexpected. We see that more biased dictators opted out less frequently. Specifically, an increase in IAT score by one standard deviation led to roughly an 11% smaller chance of opting out.

**Table 10.** The IAT's effect on sorting.

Probit Marginal Effects					
Variable	(1)	(2)	(3)	(4)	(5)
IAT D-score	−0.114 (0.073)	−0.105 (0.074)	−0.108 (0.081)	−0.111 (0.087)	−0.115 (0.093)
IATxPassedBlack	0.234 (0.175)	0.206 (0.177)	0.214 (0.185)	0.201 (0.194)	0.207 (0.202)
Costly Sorting		−0.095 (0.073)	−0.098 (0.073)	−0.096 (0.072)	−0.102 (0.072)
Dictator Controls			X		X
Receiver Controls				X	X
Observations	126	126	126	126	126

Robust standard errors in parentheses.

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

**Figure 6.** Sorting when the dictator is biased.

Finally, we extended the cross-treatment exercise from above and compared anonymous sorting to sorting with pictures, by way of Pearson's test. Again, there was no statistical difference between opting out in the two treatment rows ( $\chi^2 = 0.611, p = 0.434$ ). This held in costly sorting ( $\chi^2 = 0.623, p = 0.430$ ) and when passing to someone against whose race the participant holds a bias ( $\chi^2 = 0.707, p = 0.400$ ). As such, it is also unlikely that implicit bias was influencing giving on the extensive margin, inclusive of sorting decisions.

**Result 3.** *The magnitude of IAT score did not predict sorting in or out of the dictator game.*

## 6. Small Giving: A Robustness Check

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

To test for this concentration, I utilize the *Dislike* variable from Equation (3) above, noting that this variable highlights cases of both pro-white and anti-white bias. I also generated dummy variables for various small gift amounts. I then ran Pearson's  $\chi^2$  tests to see if giving in those small amounts was different for biased and non-biased dictators in each of the pictures treatments. Full results from these tests are depicted in Table 11.

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.

**Table 11.** The IAT and small giving.

	<i>p</i> -Value for Pass Size:			
	0	≤ 1	≤ 2	<i>N</i>
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 and Receiver is Black	0.310	0.181	0.753	94
Whole Sample	0.927	0.893	0.496	175

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

Taken together, Results 1–4 indicate that social pressures are unlikely to explain biased acts of giving. While the extent to which implicit bias affects behavior remains an open area of debate, these results can similarly be thought of as bounding the sort of behaviors they affect.

## 7. Conclusions

For millions of Americans, racial bias is a persistent concern. In the past two decades, a proposed method of detecting it, the Implicit Association Test, has caught fire among the academics who study bias. At the time of this writing, the IAT's original paper has over 10,000 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. I then conducted a laboratory experiment to test how implicit bias maps into pro-social behaviors.

Specifically, I tested 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 included a sorting (exit) option to examine social pressure as a motive for biased giving, when biased givers may simply prefer to avoid the environment altogether.

I find that, contrary to much of the previous literature where behavior has not been incentivized, 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 giving environments. Furthermore, these results hold not only in fine bins of analysis, but also in wider and more powerful ones, such as when I restrict my sample to small gifts or dictators paired with receivers against whom they hold an implicit bias.

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 fascinating, but unsubstantiated claims.

The dictator game is a compelling example in that it consists of a very simple economic decision. That the IAT fails to map into this class of decisions may be indicative of when decisions are more complex or 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 so-called fuzzier or multi-level economic decisions, such as decisions made in groups or ones where the use of heuristics has been shown to play a prominent role.

Similarly, the present study is limited in that it used photographs to provide racial information.<sup>16</sup> Of course, pictures convey much more information than recipient's race, including presented gender [44], prior friendship [45] and subtle facial cues [46]. Indeed, disentangling the effects of this information is an active area of research [47].

As stated above, however, in order to account for the potential effects of receiver gender, controls were included in the estimations, and receiver gender had no significant impact on dictator giving. Further, of the 175 dictators in sessions with pictures, only four (2.29%) had the same school year and major as their recipient, suggesting a lower likelihood of dictators being paired with those in their social circle. Last, while IRB protocols required pictures to be deleted immediately after the sessions, making coding facial cues impossible, the randomization process should divorce this confounder from the final result. Given that other common ways of making receiver race more salient, such as by replacing photographs with demographic descriptions, could expose the study to experimenter demand effects, the use of photographs in this study is justifiable. However, future research should investigate the isolated impact of receiver race.

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 more broadly. Consider the example of jurisprudence, where 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, Associate Justice Kennedy 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.<sup>17</sup>  
(*Texas DoH v. ICP Inc.* [48])

What this means is that bias is classified under the law as resulting in differential treatment even if one is not aware of the held bias, as in an implicit bias. Hence, we need further explorations of implicit bias, its potential to map into this sort of decision-making and potential mitigating factors, else we could be establishing ineffective policies, or worse, actively harmful ones.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4336/9/4/73/s1>: The Supplementary Materials contain screen captures of the IAT task (S1), experimental instructions (S2), the post-experiment questionnaire (S3) and additional estimations outside the scope of the present paper (S4).

**Acknowledgments:** This paper was made possible with the generous support of the United Nations University World Institute for Development Economics Research (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. A previous version of this paper was circulated as "Racial bias and the validity of the Implicit Association Test".

**Conflicts of Interest:** The author declares no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

<sup>16</sup> I thank an anonymous referee for suggesting this argument.

<sup>17</sup> Italics are my own.

## Abbreviations

The following abbreviations are used in this manuscript:

IAT     Implicit Association Test  
 LMW    Lazear et al. [5]

## References

- Bertrand, M.; Chugh, D.; Mullainathan, S. Implicit discrimination. *Am. Econ. Rev.* **2005**, *95*, 94–98. [[CrossRef](#)]
- Kang, J. *Implicit Bias: A Primer for Courts*; National Center for State Courts: Williamsburg, VA, USA, 2009.
- Triplet, J. Racial Bias and Prosocial Behavior. *Sociol. Compass* **2012**, *6*, 86–96. [[CrossRef](#)]
- Stepanikova, I.; Triplet, J.; Simpson, B. Implicit racial bias and prosocial behavior. *Soc. Sci. Res.* **2011**, *40*, 1186–1195. [[CrossRef](#)]
- Lazear, E.P.; Malmendier, U.; Weber, R.A. Sorting in experiments with application to social preferences. *Am. Econ. J. Appl. Econ.* **2012**, *4*, 136–163. [[CrossRef](#)]
- DellaVigna, S.; List, J.A.; Malmendier, U. Testing for altruism and social pressure in charitable giving. *Quart. J. Econ.* **2012**, *127*, 1–56. [[CrossRef](#)]
- DellaVigna, S.; List, J.A.; Malmendier, U.; Rao, G. The importance of being marginal: Gender differences in generosity. *Am. Econ. Rev.* **2013**, *103*, 586–590. [[CrossRef](#)]
- Bhattacharya, D.; Kanaya, S.; Stevens, M. Are University Admissions Academically Fair? *Rev. Econ. Stat.* **2017**, *99*, 449–464. [[CrossRef](#)]
- Norton, M.I.; Mason, M.F.; Vandello, J.A.; Biga, A.; Dyer, R. An fMRI investigation of racial paralysis. *Soc. Cogn. Affect. Neurosci.* **2012**, *8*, 387–393. [[CrossRef](#)] [[PubMed](#)]
- Fershtman, C.; Gneezy, U. Discrimination in a segmented society: An experimental approach. *Quart. J. Econ.* **2001**, *116*, 351–377. [[CrossRef](#)]
- Ferraro, P.J.; Cummings, R.G. Cultural diversity, discrimination, and economic outcomes: An experimental analysis. *Econ. Inq.* **2007**, *45*, 217–232. [[CrossRef](#)]
- Ben-Ner, A.; McCall, B.P.; Stephane, M.; Wang, H. Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. *J. Econ. Behav. Organ.* **2009**, *72*, 153–170. [[CrossRef](#)]
- Slonim, R.; Guillen, P. Gender selection discrimination: Evidence from a trust game. *J. Econ. Behav. Organ.* **2010**, *76*, 385–405. [[CrossRef](#)]
- Price, J.; Wolfers, J. Racial Discrimination Among NBA Referees. *Quart. J. Econ.* **2010**, *125*, 1859–1887. [[CrossRef](#)]
- Lowes, S.; Nunn, N.; Robinson, J.A.; Weigel, J. Understanding Ethnic Identity in Africa: Evidence from the Implicit Association Test (IAT). *Am. Econ. Rev.* **2015**, *105*, 340–345. [[CrossRef](#)]
- Berge, L.I.O.; Bjorvatn, K.; Galle, S.; Miguel, E.; Posner, D.N.; Tungodden, B.; Zhang, K. How Strong are Ethnic Preferences? *Natl. Bur. Econ. Res.* **2015**. [[CrossRef](#)]
- Rooth, D.O. Automatic associations and discrimination in hiring: Real world evidence. *Labour Econ.* **2010**, *17*, 523–534. [[CrossRef](#)]
- Glover, D.; Pallais, A.; Pariente, W. Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. *Quart. J. Econ.* **2017**, *132*, 1219–1260. [[CrossRef](#)]
- Greenwald, A.G.; McGhee, D.E.; Schwartz, J.L. Measuring individual differences in implicit cognition: The implicit association test. *J. Personal. Soc. Psychol.* **1998**, *74*, 1464. [[CrossRef](#)]
- Mill, J. *Analysis of the Phenomena of the Human Mind*; Longmans, Green, Reader and Dyer: London, UK, 1869.
- Meade, A.W. FreeIAT: An open-source program to administer the implicit association test. *Appl. Psychol. Meas.* **2009**, *33*, 643–643. [[CrossRef](#)]
- Greenwald, A.G.; Nosek, B.A.; Banaji, M.R. Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Personal. Soc. Psychol.* **2003**, *85*, 197. [[CrossRef](#)]
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Academic Press: Cambridge, MA, USA, 2013.
- Nosek, B.A.; Banaji, M.; Greenwald, A.G. Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dyn. Theory Res. Pract.* **2002**, *6*, 101. [[CrossRef](#)]
- Rudman, L.A.; Glick, P. Prescriptive gender stereotypes and backlash toward agentic women. *J. Soc. Issues* **2001**, *57*, 743–762. [[CrossRef](#)]

26. Rudman, L.A.; Ashmore, R.D. Discrimination and the implicit association test. *Group Process. Intergr. Relat.* **2007**, *10*, 359–372. [[CrossRef](#)]
27. Jost, J.T.; Rudman, L.A.; Blair, I.V.; Carney, D.R.; Dasgupta, N.; Glaser, J.; Hardin, C.D. 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. *Res. Organ. Behav.* **2009**, *29*, 39–69. [[CrossRef](#)]
28. Greenwald, A.G.; Poehlman, T.A.; Uhlmann, E.L.; Banaji, M.R. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *J. Personal. Soc. Psychol.* **2009**, *97*, 17. [[CrossRef](#)] [[PubMed](#)]
29. Singal, J. Psychology’s Favorite Tool for Measuring Racism Isn’t Up to the Job. *New York Magazine*, 11 January 2017.
30. Arkes, H.R.; Tetlock, P.E. Attributions of implicit prejudice, or “Would Jesse Jackson ‘fail’ the Implicit Association Test?” *Psychol. Inq.* **2004**, *15*, 257–278. [[CrossRef](#)]
31. Forscher, P.S.; Lai, C.; Axt, J.; Ebersole, C.R.; Herman, M.; Nosek, B.A.; Devine, P.G. A meta-analysis of change in implicit bias. *arXiv* **2016**. [[CrossRef](#)]
32. Brañas-Garza, P.; Durán, M.A.; Paz Espinosa, M. The role of personal involvement and responsibility in unfair outcomes: A classroom investigation. *Ration. Soc.* **2009**, *21*, 225–248. [[CrossRef](#)]
33. Camerer, C. *Behavioral Game Theory: Experiments in Strategic Interaction*; Princeton University Press: Princeton, NJ, USA, 2003.
34. Dana, J.; Weber, R.A.; Kuang, J.X. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Econ. Theory* **2007**, *33*, 67–80. [[CrossRef](#)]
35. Garicano, L.; Palacios-Huerta, I.; Prendergast, C. Favoritism under social pressure. *Rev. Econ. Stat.* **2005**, *87*, 208–216. [[CrossRef](#)]
36. Dana, J.; Cain, D.M.; Dawes, R.M. What you don’t know won’t hurt me: Costly (but quiet) exit in dictator games. *Organ. Behav. Hum. Decis. Process.* **2006**, *100*, 193–201. [[CrossRef](#)]
37. Broberg, T.; Ellingsen, T.; Johannesson, M. Is generosity involuntary? *Econ. Lett.* **2007**, *94*, 32–37. [[CrossRef](#)]
38. Akerlof, G.A.; Kranton, R.E. Economics and identity. *Quart. J. Econ.* **2000**, *115*, 715–753. [[CrossRef](#)]
39. Holt, C.A.; Laury, S.K. Risk aversion and incentive effects. *Am. Econ. Rev.* **2002**, *92*, 1644–1655. [[CrossRef](#)]
40. Fiedler, K.; Bluemke, M. Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic Appl. Soc. Psychol.* **2005**, *27*, 307–316. [[CrossRef](#)]
41. Exadaktylos, F.; Espín, A.M.; Branäs-Garza, P. Experimental subjects are not different. *Sci. Rep.* **2013**, *3*, 1213. [[CrossRef](#)] [[PubMed](#)]
42. Lane, T. Discrimination in the laboratory: A meta-analysis of economics experiments. *Eur. Econ. Rev.* **2016**, *90*, 375–402. [[CrossRef](#)]
43. Cragg, J.G. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econ. J. Econ. Soc.* **1971**, *39*, 829–844. [[CrossRef](#)]
44. Dufwenberg, M.; Muren, A. Generosity, anonymity, gender. *J. Econ. Behav. Organ.* **2006**, *61*, 42–49. [[CrossRef](#)]
45. Brañas-Garza, P.; Durán, M.A.; Espinosa, M.P. Favouring friends. *Bull. Econ. Res.* **2012**, *64*, 172–178. [[CrossRef](#)]
46. Nettle, D.; Harper, Z.; Kidson, A.; Stone, R.; Penton-Voak, I.S.; Bateson, M. The watching eyes effect in the Dictator Game: it’s not how much you give, it’s being seen to give something. *Evolut. Hum. Behav.* **2013**, *34*, 35–40. [[CrossRef](#)]
47. Brañas-Garza, P.; Bucheli, M.; Espinosa, M.P. Altruism and information. In Proceedings of the 10th Nordic Conference on Behavioral and Experimental Economics, Tampere, Finland, 25–26 September 2015.
48. Kennedy, A. *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.*; 576; Opin. U. S. Supreme Court: Washington, DC, USA, 2015.

