Columbia Law School Scholarship Archive

Faculty Scholarship

Faculty Publications

2019

Price and Prejudice: An Empirical Test of Financial Incentives, Altruism, and Racial Bias

Kristen Underhill Columbia Law School

Follow this and additional works at: https://scholarship.law.columbia.edu/faculty_scholarship

Part of the Banking and Finance Law Commons, Jurisprudence Commons, Law and Economics Commons, and the Law and Society Commons

Recommended Citation

Kristen Underhill, *Price and Prejudice: An Empirical Test of Financial Incentives, Altruism, and Racial Bias*, 48 J. LEGAL STUD. 245 (2019). Available at: https://scholarship.law.columbia.edu/faculty_scholarship/2797

This Article is brought to you for free and open access by the Faculty Publications at Scholarship Archive. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Scholarship Archive. For more information, please contact scholarshiparchive@law.columbia.edu, rwitt@law.columbia.edu.

Price and Prejudice: An Empirical Test of Financial Incentives, Altruism, and Racial Bias

Kristen Underhill

ABSTRACT

Many argue that paying people for good behavior can crowd out beneficial motivations like altruism. But little is known about how financial incentives interact with harmful motivations like racial bias. Two randomized vignette studies test how financial incentives affect bias. The first experiment varies the race of a hypothetical patient in need of a kidney transplant (black or white), an incentive (\$18,500 or none), and addition of a message appealing to altruism. Incentives encouraged donation but introduced a significant bias favoring white patients. The second experiment assesses willingness to donate to a patient (black or white) without an incentive and then introduces incentives varying in size (\$3,000, \$18,800, or \$50,000) and source (charity, government, or patient's own funds). Incentives encouraged donation but were significantly more effective in encouraging donation to white patients. Biasing effects are most pronounced for medium-sized incentives. Incentives may have an inadvertent biasing effect for altruistic behavior.

1. INTRODUCTION

Money is persuasive. Numerous private and public policies draw on the assumption that financial incentives effectively motivate good behavior. These efforts span the law, ranging from premium surcharges for smokers, cash transfer initiatives for good students, penalties for Med-

KRISTEN UNDERHILL is Associate Professor of Law and Associate Professor of Population and Family Health at Columbia University. I am grateful for feedback and formative conversations on this research with Ian Ayres, Sarah Calabrese, John Dovidio, Jeff Fagan, Chris Kahler, Amy Kapczynski, Daniel Markovits, Ed Morrison, Don Operario, Brian Richardson, Bill Sage, Eric Talley, the anonymous referees, and many other scholars who have engaged with my work. I am grateful to Juliana Bennington for research assistance. Funds for this work were provided by the Oscar M. Ruebhausen Fund at Yale Law School. All errors herein are my own.

[Journal of Legal Studies, vol. 48 (June 2019)] © 2019 by The University of Chicago. All rights reserved. 0047-2530/2019/4802-0008\$10.00 icaid beneficiaries who skip their checkups, tax breaks for charitable donations, incentives for criminal informants, and payments for corporate whistle-blowers (Underhill 2018). These and other uses of incentives, however, have drawn criticism from commentators who question the effectiveness and moral impacts of carrots and sticks (Grant 2012; Sandel 2012). A central concern about incentive-based policies is that greed will displace or erode more laudable motivations such as altruism, professionalism, and moral conviction (Bowles 2008). In some areas of the law, such as encouraging the donation of human organs (42 U.S.C. sec. 274e [2015]) or according Good Samaritans a lower standard of care in tort (Sutton 2010), the concern for preserving moral motivations is so strong that statutes may proscribe financial rewards entirely.

Anxieties about incentive-based policies reflect long-held concerns about motivational crowding out, a process by which incentives may interact with other motivations for decisions (Bowles and Polanía-Reves 2012; Frey and Jegen 2001; Underhill 2016). Researchers in law, economics, and psychology have identified distinct ways that incentivesoften but not always financial rewards-can diminish or overpower intrinsic motivations (Benkler 2006). Strong crowding-out effects may cause incentive-based policies to backfire entirely (Gneezy and Rustichini 2000); weaker crowding-out effects simply reduce incentives' anticipated effectiveness (Frey and Jegen 2001). In the decades since Titmuss (1972) and Arrow (1972) debated crowding-out concerns regarding blood donation, modern concerns about motivational crowding out have spanned almost every area of the law, and commentators warn that incentives may cause disengagement, poor performance, coercion, and the erosion of moral values. But while legal scholarship has examined how incentives may infringe on noble motivations, it has overlooked how incentives may also affect less honorable impulses, such as bias, spite, and revenge.

This paper identifies how incentives may interact with invidious motivations, focusing particularly on racial bias. If money crowds out both good and bad morals, incentives might be expected to reduce harmful motivations like racial preferences. But instead this paper provides original evidence to show that in some settings, incentives may unexpectedly amplify racial discrimination. When behavior is susceptible to bias, such as unconscious racial preferences, financial incentives may inadvertently encourage more biased activity. Indeed, for actions that are highly susceptible to bias, or where biases are especially damaging, it may be unwise to offer incentives—or, perhaps, wise to offer incentives that are just large enough to overwhelm their biasing effects.

To support these claims, this paper presents two experimental vignette studies testing the interaction between incentives and racial bias. Vignettes used the context of participants' willingness to be a living kidney donor to a patient who was either white or black. The setting of organ donation has been fertile ground for fears about crowding out of altruism, but there has been little rigorous testing of this concern. Organ donation in the United States is governed by the National Organ Transplant Act of 1984 (NOTA; Pub. L. No. 98-507, 98 Stat. 2339), which imposes criminal penalties for transferring solid organs in exchange for valuable consideration. Federal courts have upheld this provision, in part on the basis of the argument that permitting financial incentives would displace the altruistic motivations that support our current donation system. Even living donors, who frequently bear financial costs associated with donating (for example, missed work and travel; Reese, Boudville, and Garg 2015), must go without compensation. The decision to forgo incentives for organ donation is of pressing public significance, with a national waiting list of over 113,000 patients and 20 deaths per day among those listed (Health Resources and Services Administration 2019). Organ shortages particularly affect patients of color, and racial disparities are pronounced in the national organ allocation system (Goodwin and Gewertz 2009; Kulkarni et al. 2019; Rodrigue et al. 2015). In this context, focusing on the crowding out of invidious motivations raises new questions about using incentives, particularly for living donors who direct their donations to particular recipients. Many scholars have proposed strategies for incentivizing donations, and prior work has investigated the impacts of tax incentives on donation (Bilgel and Galle 2014), but no research has considered how incentives may interact with racial bias in organ donation or elsewhere. It is important to identify how paying for organs may influence these inequalities.

The studies herein find that incentives increase willingness to donate to both white and black recipients; crowding out of altruism was minimal. But surprisingly, incentives were significantly more helpful for motivating altruistic responses toward white rather than black recipients. These findings suggest that incentives may disinhibit invidious racial preferences, contrary to crowding-out theories. This paper explores several explanations for these effects and the potential remedies of increasing incentive size and bundling incentives with messages emphasizing altruism. According to our findings, incentives may encourage altruistic behavior with little impact on motivation, although any NOTA reforms would need to consider additional concerns beyond altruism crowding out, such as undue inducement. Efforts to introduce incentives in organ donation and other fields, however, should be attentive to the possibility that rewards may amplify invidious biases if incentive structures are not carefully designed. The biasing impact of incentives may be strongest with medium-sized incentives, and it may be alleviated by explicitly invoking prosocial motivations, increasing incentive size, or limiting the uses of incentives to activities that are not intended to benefit a known recipient.

2. BACKGROUND

2.1. Motivational Crowding Out

Motivational crowding out inverts traditional economic assumptions about human behavior: namely, that self-interest drives us and that rewards supplement our other motivations. But incentives can interact with our preexisting motivations in unpredictable ways. Crowding-out scholarship identifies two rough categories of motivation: extrinsic motivations refer to the desire for a separable outcome, such as praise, money, or avoiding punishment; intrinsic motivations, in contrast, are our varied reasons for acting on a task's "inherent satisfactions" (Ryan and Deci 2000, p. 71). In research on motivational crowding out, the classification of extrinsic and intrinsic corresponds roughly to the difference between economic self-interest and all other motivations (Feldman 2011).

Motivational crowding out refers to any process by which the offer of an extrinsic incentive reduces individuals' intrinsic motivation to behave as desired (Frey and Jegen 2001). Because the phenomenon is counterintuitive, it is useful to specify three broad explanations for crowding-out effects (Bowles and Polanía-Reyes 2012).

First, signaling mechanisms stipulate that agents interpret incentives as messages (Bénabou and Tirole 2003). When a principal offers an incentive for a particular task, the size, type, and presence of that incentive can convey information to the agent, which can interfere with motivations. Incentives may signal that a task is dangerous, unpopular (Gordon 2010; Vermeule 2012), or unpleasant (Bolle and Otto 2010); that the principal disrespects or distrusts the agent (Kamenica 2012); that the principal seeks to control the agent (Feld and Frey 2007) or has bad moral values (Beretti, Figuières, and Grolleau 2013); that the principal is incompetent or extravagant (Bénabou and Tirole 2006); or that the agent's decision

should be based on self-interest (Meier 2006; Gneezy and Rustichini 2000). Incentives can also interrupt the signals that agents want to send to others, which can interfere with reputational motivations (Bénabou and Tirole 2006). When observers can attribute an agent's good behavior to an incentive, the agent can no longer signal her good character through her choices. Incentives can also disrupt the signals that the agent sends herself, which undermines "warm-glow" or self-image motivations (Andreoni 1990; Imas 2014).

Second, incentives can diminish self-determination. We derive value from acting on our intrinsic motivations, including the value of controlling our behavior (Bowles 2008). Incentives render a measure of our motivations unnecessary (Bowles 2008; Frey and Jegen 2001). Incentives may also displace some control from the agent to the principal; an incentivized agent can either conform to her principal's desires or reject the incentive, but either choice is now reactive to the principal's offer (Bénabou and Tirole 2003). Incentives thus undermine self-determination, which can lead agents to disengage (Galle 2014).

A third explanation for crowding-out effects posits that preferences including preferences for incentives and other types of benefits—are endogenous to the policy environments to which we are exposed (Bar-Gill and Fershtman 2005). When a change in policy occurs, such as the offer of a new incentive, this change may produce a corresponding shift in the extent to which we value incentives compared with other types of benefits. These preference changes may spread over time and across society via processes such as cultural transmission or imitation (Bar-Gill and Fershtman 2005).

2.2. Racial Bias as a Motivation

This study considers racial bias as a form of intrinsic motivation—namely, the motivation to favor one race over another. As applied to tasks (like organ donation in our experiments), this motivation may be not so much a single motivation as a disparity in motivations to take actions that affect people of different races. Under this view, we may consider bias to be a disproportionate motivation to help people of a favored race compared with others or a disproportionate motivation to harm people of a disfavored race. Explicit and implicit racial biases correspond to conscious and unconscious racial preferences that influence behavior (Jolls and Sunstein 2006).

In the United States, implicit antiblack biases are common, including among those who consciously disagree with racist beliefs. Research on aversive racism finds that many people who genuinely wish to behave without prejudice—and who may indeed suppress their bias when race is made salient—nonetheless behave according to discriminatory implicit biases when race is more subtle (Dovidio and Gaertner 2004; Pearson, Dovidio, and Gaertner 2009). Antiblack implicit bias is most commonly assessed through the Implicit Association Test (IAT), which measures the speed at which respondents can associate pleasant or unpleasant concepts with white versus black racial categories (Greenwald, McGhee, and Schwartz 1998). Research using the IAT suggests that implicit biases are strong predictors of behavior (Stepanikova, Triplett, and Simpson 2011), which indicates that implicit bias is indeed a motivation (or a race-based discrepancy in motivations) affecting individual choices.

Although no study has yet tested how implicit biases respond to financial incentives, prior studies have found that implicit biases can limit altruism toward people of color (Rudman and Ashmore 2007; Stepanikova, Triplett, and Simpson 2011). If racial bias may indeed influence altruistic decisions, there are outstanding questions about how incentives may interact with those biases.

2.3. Organ Donation as a Case Study

Organ donation is an attractive case study for examining altruism, racial bias, and crowding out. Racial disparities in organ allocation are pervasive (Rodrigue et al. 2015), even after recent modifications to the kidney allocation system that mitigated some inequalities (Kulkarni et al. 2019). Although white individuals account for 41 percent of people on current waiting lists, they receive 55 percent of organ transplants; corresponding figures are 29 percent and 21 percent for African American patients, 20 percent and 16 percent for Hispanic patients, and 8 percent and 6 percent for Asian patients (Health Resources and Services Administration 2019). Patients of color experience longer waiting times and higher mortality rates while waiting for organs (Rodrigue et al. 2015), are less likely to be moved to active listing for kidney transplants (Kulkarni et al. 2019), and experience greater need for organ transplants due to health inequalities (Bratton, Chavin, and Baliga 2011). Most of the disparities are attributed to structural racism (Goodwin 2009). But recently social media has provided new means of soliciting donations directly, which raises concerns about interpersonal biases in directed donations (Neidich et al. 2012).

Crowding-out fears are also strong in this area: the federal ban on incentives for solid organ donors in NOTA in part reflects the fear that incentives will displace altruism (Gross 2008; *Flynn v. Holder*, 684 F.3d 852, 855–56 [9th Cir. 2012]). There are other reasons for maintaining an altruism-based donation system; principal among these is the fear of undue inducement for poor donors, and some also express concerns about commodifying organs (Goodwin 2004; Choi, Gulati, and Posner 2014; Satz 2010; Grant 2012). But legislative debate and continuing commentary on NOTA has also invoked the theory advanced by Titmuss (1972) that paying blood donors would drive away altruists (Gross 2008). The lack of a randomized test of financial incentives has impoverished this discussion.

3. EVIDENCE

This paper presents two randomized vignette studies. The first considers how incentives and appeals to altruism affect bias; the second principally considers how the size and source of incentives affect bias. Participants in both experiments read vignettes about a patient in need of a living kidney donation and then answered questions about willingness to donate, beliefs about the patient, beliefs about donation, demographic characteristics, explicit racial bias (McConahay 1986), implicit racial bias using a survey-based variant of the IAT (modeled on Lemm et al. [2008] and Greenwald, McGhee, and Schwartz [1998] and adapted here for the survey tool Qualtrics), altruism using a scale predictive of organ donation behavior (Morgan and Miller 2002), and materialism (collected only in experiment 1; Richins 2004). The primary outcome was willingness to donate, assessed on a continuous 1-10 scale. Participants were drawn from Amazon Mechanical Turk (MTurk), an online platform that makes Web-based tasks available to compensated workers. All were US adults who had completed at least 100 prior tasks on MTurk with a 95 percent completion rate. Data collection took place in small batches over different times and days of the week, using the Qualtrics survey platform. Participants were not eligible to reenroll, and participants who completed the first experiment were excluded from the second experiment. Procedures were approved by the Yale Human Subjects Committee.

Vignettes communicated the race of the hypothetical patient by first name, randomly varying the name to have white or black associations. This facilitated the observation of implicit bias: describing the patient's race outright as white or black would make race salient, which can encourage aversive racist participants to consciously avoid prejudiced responses. Six first names were used for each race and randomly varied within each condition; names were drawn from a well-known field study that validated racial associations (Bertrand and Mullainathan 2004). The patient had the last name Johnson, which is the second most common last name among both white and black individuals nationally (US Census Bureau 2008), which makes it unlikely to signal race.

Although the vignette studies in this paper assess attitudinal outcomes, all of the mechanisms of crowding out are based on cognitive mechanisms (attitudes, beliefs, and changes in perceptions). Attitudinal measures are therefore a useful preliminary test of how incentives may drive crowding-out effects.

3.1. Experiment 1

Under the assumption that financial incentives crowd out intrinsic motivation, this experiment tests the hypothesis that financial incentives diminish differences in willingness to donate due to racial bias. That is, individuals without an incentive will be significantly more willing to donate to a white compared with a black patient, while incentivized individuals will not differ in their willingness to donate by a patient's race.

This study is a 2×3 randomized trial, varying race (white or black) and incentive (no incentive, incentive, or incentive accompanied by language appealing to altruistic values). Participants read a vignette about Mr. Johnson, whose first name was randomized to have either white or black associations. The patient needed a donation from a living kidney donor, and the vignette stipulated that the participant was a blood and tissue match and healthy enough to donate. The vignette also stated that all medical expenses would be covered by a local charity. The noincentive vignette ended there. In the incentive condition, the vignette specified that the charity "will also pay you \$18,500 to financially compensate you for donating your kidney." The amount was set on the basis of estimates in Becker and Elias (2007) and updated for inflation. In an initial round of data collection (n = 412), findings tend to show that financial incentives augmented racial bias. Recruitment was briefly paused to add a third incentive condition-the financial incentive plus a verbal appeal that attempted to amplify altruistic motivations in the presence of money ("The choice to be a living organ donor is a unique act of altruism, in which a healthy person willingly gives up an organ to help someone else survive. People who receive lifesaving transplants benefit enormously from the generosity of living donors."). Recruitment was then restarted and continued until reaching a prespecified power calculation for the second round (n = 977; according to the power calculation, 972 participants yielded 80 percent power to detect a small effect size at p <.05). Data from the two rounds were combined to maximize statistical power, and the round was included as a covariate in analyses (total n =1,389). Results did not differ significantly by round. Vignettes for both experiments are reproduced in the Online Appendix.

Participants were young and educated; 61 percent were between 25 and 45 years of age, and 89 percent reported at least some college coursework. Approximately 55 percent were female, and 62 percent were drawn from the South and Midwest; 76 percent were non-Hispanic white, and 99 percent had high English proficiency. About 45 percent identified with a denomination of Christianity; 29 percent were atheist or agnostic. Fifty-five percent of participants held a full-time job, and 40 percent reported an annual income between \$20,000 and \$50,000, with a median household size of three.

Analyses use ordinary least squares (OLS) regressions to test the main effects of incentive, the patient's race, altruism appeal, the interaction between race and incentive, and the interaction between race and altruism appeal. Incentive, race, and altruism appeal are dummy coded. Regression models begin with main effects and interactions and then progressively include covariates drawn from the vignette context and crowdingout theory. The covariates are willingness to be a deceased organ donor (to address the possibility that some individuals object to donating apart from incentive or race), materialism, trait altruism, implicit and explicit bias, and interactions between each type of bias and the patient's race. Analyses use Stata 14.

3.1.1. Results: Racial Bias Crowding In and Altruism Appeals. As Tables 1 and 2 demonstrate, the main effect of the incentives was to encourage rather than discourage donation. In every regression model, incentives significantly increased participants' willingness to donate regardless of the patient's race. Raw means for willingness to donate are displayed in Table 1, and regression models appear in Table 2.

Naked incentives moved willingness over the midpoint of the 1–10 continuous scale, with observed group means of 4.99 and 4.97 without incentives (for black and white patients, respectively) to 5.74 and 6.34 with incentives. The altruism appeal increased willingness to donate fur-

Group	Ν	Mean
No incentive:		
Black, no appeal to altruism	265	4.99 (2.75)
White, no appeal to altruism	266	4.97 (2.67)
Incentive:		
Black, no appeal to altruism	265	5.74 (3.06)
White, no appeal to altruism	266	6.34 (2.80)
Black, appeal to altruism	164	6.13 (2.93)
White, appeal to altruism	163	6.42 (2.88)

Table 1. Willingness to Donate: Group Means for Experiment 1

Note. Standard deviations are in parentheses.

ther, to respective means of 6.13 and 6.42 for black and white patients, but the marginal increase is not significant when compared with the incentive alone. Supplementary findings on altruism crowding out are presented in the Online Appendix. The main effect of the patient's race is not significant.

Overall, incentives increased willingness to donate by approximately .67 according to the best-fitting regression model (model 3 in Table 2). The size of this effect corresponds to a partial η^2 -value of .029 (95 percent confidence interval = [.014, .048]); using a rule of thumb defining a partial η^2 -value of .01 as a small effect and .06 as a medium-sized effect, this is a small-to-medium effect. A mean increase of .67 does not seem large on a 10-point scale, and interpretation is limited by the hypothetical nature of this vignette study and the generalizability of the subject population. But it is perhaps useful to note that compared with the no-incentive condition, the proportion of participants who reported willingness to donate above the scale midpoint of 5 increased by approximately 14 percent in the naked-incentive condition (from 46 to 60 percent). There were 6,831 organ donations by living donors in the United States in 2018 (Health and Research Services Administration 2019), and even a small increase in living kidney donations is meaningful.

The interaction between a patient's race and the financial incentive is the primary focus of this study. Analyses identify a significant interaction, which indicates that the effect of an incentive differed depending on whether the patient had a first name with white or black associations. Although incentives increased participants' willingness to donate to patients of both races, this increase was significantly larger for the white patient compared with the black patient. This results in a statistically significant racial bias favoring whites in the incentive-alone condition (t(529) = -2.38; p < .05).

The significant interaction between incentive and a patient's race emerged not only for the primary outcome—willingness to donate—but also for perceptions of the patient. Incentivized participants were significantly more likely to believe that the patient was at fault for his health problems when he was black, compared with when he was white (observed blame scores of 3.53 versus 3.17 on a 1–7 scale for black and white patients, respectively; t(529) = 3.48; p < .001). Using the same predictors as model 3, the interaction between race and incentive is significant for perceptions of patient fault ($\beta = -.45$; SE = .15; t(1,388) =-3.03; p < .01).

When the appeal to altruism is added to the incentive, the differences from the incentive-alone condition abate. Willingness to donate to the black patient rises such that the disparity is no longer significant (observed group means of 6.13 versus 6.42 for black and white patients, respectively; t(325) = -.90; p = .37). The interaction term between the altruism message and the patient's race is not statistically significant, however, for overall willingness to donate, as shown in Table 2. The tendency to blame black patients for their health problems also decreases such that the racial disparity is no longer significant (observed means of 3.22 versus 3.43 for black and white patients; t(325) = -1.47; p = .14). The interaction between the message and the patient's race is significant in analyses for patient blame (that is, using the same predictors as model 3, the interaction term is significant at $\beta = .61$; SE = .17; t(1,388) = 3.58; p < .001).

Taken together, these results suggest that without an incentive, people do not differ in their willingness to donate to a black versus a white patient. Incentives, however, crowd in bias—they introduce a significant preference for donating to a white patient compared with a black patient, and they prompt participants to believe that black patients are more to blame for their health problems. Adding an appeal to altruism results in highest mean willingness to donate and lessens the racial disparity, but the effects of altruism appeals do not reach statistical significance for willingness to donate. The finding that incentives amplify racial disparities is contrary to the study hypothesis, but the following sections consider possible explanations, using exploratory post hoc tests.

3.1.2. Explaining Racial Bias Crowding In. I consider four potential pathways by which incentives may augment the expression of racial

	Model 1	Model 2	Model 3	Model 4	Model 5
Incentive	.742	.713	.674	.684	.687
	(.247)	(.240)	(.232)	(.233)	(.232)
	[.003**]	[.003**]	$[.004^{**}]$	[.003**]	[.003**]
Patient's Race (White)	025	092	118	108	122
	(.247)	(.239)	(.232)	(.232)	(.232)
	[.920]	[.701]	[.612]	[.641]	[.599]
Altruism Message	.307	.322	.320	.314	.299
	(.291)	(.282)	(.273)	(.273)	(.273)
	[.290]	[.254]	[.241]	[.250]	[.273]
Patient's Race (White) × Incentive	.633	.801	.746	.729	.756
	(.349)	(.339)	(.328)	(.329)	(.328)
	[.070]	[.018*]	[.023*]	[.027*]	[.021*]
Patient's Race (White) × Altruism Message	319	465	478	476	512
	(.400)	(.388)	(.376)	(.376)	(.375)
	[.425]	[.231]	[.204]	[.206]	[.173]
Round	239	185	211	206	230
	(.179)	(.174)	(.168)	(.168)	(.168)
	[.182]	[.288]	[.211]	[.221]	[.172]
Deceased Donor Willingness		1.942	1.710	1.714	1.683
		(.208)	(.204)	(.204)	(.205)
		[***000]	[***000]	[***000]	[***000]
Altruism			060.	.087	.086
			(.012)	(.012)	(.012)
			[***000]	[***000]	[***000]

Table 2. Willingness to Donate: Regression Results for Experiment 1

Participant Altruism × Incentive			036	031	030
1			(.015)	(.015)	(.015)
			[.014*]	[.039*]	[.047*]
Materialism				018	019
				(.016)	(.016)
				[.242]	[.230]
Materialism × Incentive				.024	.027
				(.020)	(.020)
				[.232]	[.176]
Implicit Racism					035
					(.020)
					[.087]
Implicit Racism × Patient's Race (White)					.075
					(.026)
					$[.005^{**}]$
Explicit Racism					016
					(.016)
					[.322]
Explicit Racism × Patient's Race (White)					.022
					(.023)
					[.352]
Intercept	5.085***	3.430	3.697	3.688	3.734
	(.188)	(.255)	(.248)	(.248)	(.249)
Adjusted R ²	.040	.096	.154	.153	.157
Note. Scores for altruism, materialism, and i	mplicit and explic	cit racism are me	an centered. Standa	urd errors are in pai	rentheses; <i>p</i> -val-

ses; p-va¹ u par Index ues are in square brackets. * p < .05. ** p < .01. *** p < .001.

bias: distraction, amoral reasoning, in-group favoritism, and licensing. Subgroup analyses testing the first three of these pathways are reported in Table 3, and the results provide tentative support for all three, with little ability to disaggregate the explanations at this stage. Future studies should take on these questions on an a priori basis. The present analyses are speculative and post hoc, so they should be considered hypothesis generating rather than hypothesis testing. In sum, these results imply that incentives can disinhibit implicit biases derived from out-group animus or in-group favoritism or, alternately, that incentives may change the calculus of an altruistic decision (making people less attentive to social concerns like bias or fairness).

First, results provide some support for the idea that money is distracting, which can interfere with individuals' efforts to police themselves against behaving in a biased way. This dynamic is particularly visible among aversive racists: individuals who consciously disagree with racist beliefs but who nonetheless behave according to implicit racial bias when race is not made salient (Dovidio and Gaertner 2004). To test this effect, the analyses considers three subgroups of participants: aversive racists (who report an implicit antiblack bias on the IAT but below-average explicit racism), consistent racists (implicit antiblack bias, above-average explicit racism), and nonracists (no implicit antiblack bias, below-average explicit racism). Pairwise *t*-tests test disparities between willingness to donate to a white versus a black patient in each of the three incentive conditions (Table 3). Biased preferences that disfavor black patients reach statistical significance only among aversive racist participants and only in the naked-incentive condition.¹ This suggests that the biasing effect of incentives may be particular to aversive racists, who would otherwise pay more attention to race as part of an altruistic donation decision. Remind-

1. An alternative explanation for bias among aversive racists is that people who are morally opposed to racism may be more reluctant to profit from the health problems of a black patient compared with a white patient. But this explanation is contradicted by findings on patient blame. Aversive racists in the altruism condition were significantly less likely to blame the patient when he was black, compared with when he was white (t(181) = -2.60; p < .05). But when offered an incentive, aversive racists were significantly more likely to blame the black patient compared with the white patient (t(180) = 2.56; p < .05). Aversive racists may be hypervigilant about bias when they are encouraged to make decisions in purely altruistic terms. The lack of statistical significance in comparing willingness to donate to white versus black patients among consistent racists may reflect the fact that willingness to donate was lower overall in this group; this comparison approaches significance when consistent racists are exposed to both an incentive and an appeal to altruism.

			No Inc	entive		Naked In	centive	Inc	entive with Appeal
	N	Black Patient	White Patient	t-Statistic	Black Patient	White Patient	t-Statistic	Black Patient	White Patient <i>t</i> -Statistic
Aversive racist	482	5.15	5.44	t(181) =70 [.48]	5.74	6.71 t(1)	[80) = -2.37 [19*]	6.31	6.28 $t(115) = .06$ [.95]
Consistent racist	419	4.44	4.61	t(166) =42	5.47	6.00 t(1)	[49) =97	6.09	7.13 $t(98) = -1.85$
Nonracist	309	5.58	5.00	t(108) = 1.18	5.78	6.34 $t(1)$	(25) = -1.11	6.24	6.00 t(70) = .32
Respondent:	č						i		
Implicit problack bias	331	5.49	4.65	t(118) = 1.59 [.11]	6.25	6.58 <i>t</i> (1 [.4	[34) =71 [8]	6.42	6.03 t(73) = .56 [.58]
Non-Hispanic white	1,051	4.97	4.98	t(396) = .00 [1.00]	5.82	6.58 t(3)	(399) = -2.64 (09**]	6.06	6.51 t(250) = -1.24
Black	115	6.25	4.65	t(40) = 1.74 [.09]	5.70	6.48 t(4) = 1.3	(12) =88	5.94	$6.55 t(27) =58 \\ [.57]$
Concern about costs and benefits:				T		I	I		r I
Above average	746	4.47	4.67	t(257) =61 [.54]	5.76	6.72 t(3)	302) = -2.93 004^{**}	6.44	6.32 $t(181) = .27$ [.79]
Below average	643	5.48	5.25	t(270) = .71 [.48]	5.70	5.88 $t(2)$	225) =46 [5]	5.79	$6.57 t(142) = -1.53 \\ [.13]$

Table 3. Subgroup Analyses of Willingness to Donate: Pairwise Comparisons for Experiment 1

Note. *P*-values are in square brackets. * p < .05. ** p < .01.

ing these participants of altruistic motives in the incentive-plus-appeal condition may have made race salient once more, which encouraged more vigilance against implicit bias. This study supports this possibility, but confirmatory work is needed in other settings. A useful test may be to repeat this finding with other implicit biases or perhaps to test for a mirror-image effect among a sample showing problack/antiwhite bias in the IAT. Although 331 participants in this study reported a preference for black and against white on the IAT (compared with 901 participants with an antiblack bias and 157 with no preference), the study lacks data to identify people with aversive bias in the opposite direction.

A second, related explanation may be that incentives activate in-group favoritism. In-group favoritism refers to favorable associations with people in one's own group, and although it runs concurrently with implicit biases against others, it is based on personal identity rather than negative associations with a disfavored group (Smith, Levinson, and Robinson 2015). Incentives could augment in-group favoritism in several ways. First, they could make donation more like an exchange; this can increase interdependence, which is a predictor of in-group favoritism (Balliet, Wu, and De Dreu 2014). Alternately, incentives could make refusals easier for people who believe they have in effect paid for the privilege of refusing, but guilt avoidance and empathy motivations may make those refusals harder for people who share an in-group identity with the recipient (Smith, Levinson, and Robinson 2015; Everett, Faber, and Crockett 2015; Hein et al. 2010).

This study examines in-group favoritism by measuring preferences among white participants to donate to the white patient and preferences among black participants to donate to the black patient (Table 3). Analyses provide incomplete support for in-group favoritism. Non-Hispanic white participants were more willing to donate to the white patient, with findings reaching statistical significance in the naked-incentive condition. Black participants, however, did not show a mirror-image effect; for this group, willingness to donate does not vary significantly by the patient's race in any incentive condition. These findings are limited by sample size; with only 115, the number of black participants may be too small for statistical tests. Further work is needed to test this effect more fully.

This study also provides preliminary evidence for a third explanation: incentives may shift the decisional frame of a choice from an altruistic (social) decision to a self-interested, economic mode of reasoning. Participants reasoning this way may simply care less about the moral implications of their choices, including the possibility of making discriminatory decisions. This overlaps with the aversive-racism explanation above but leaves open the possibility that participants reasoning in a self-interested way may be conscious that they are discriminating. The present study explores this by identifing whether biased decisions differ between two groups of participants: those who say that financial costs and benefits are important when making a decision to donate and those who say that financial costs and benefits are unimportant. In the first group, racial bias is significant in the naked-incentive condition (see Table 3). Among participants reporting below-average concern about money, racial bias does not reach statistical significance in any of the three incentive conditions. A close look also shows that exposure to an incentive increased participants' beliefs that financial costs and benefits are important (see the Online Appendix). Together, these findings suggest that people exposed to financial incentives are more attentive to financial costs and benefits and simultaneously more likely to favor the white participant unless also exposed to an altruism appeal. But again, further work is needed to confirm these explorations.

A final potential explanation is licensing, which may be partially contradicted by the present study. This explanation proceeds as follows: when participants are biased against black patients, they always wish to refuse the donation, but they are reluctant to do so when it reflects poorly on their altruism. Incentives, however, can make refusals easier-offering money can in effect license participants' refusal because they can pay for the privilege of refusing, or they can decide (or represent) that they are offended by the offer. Prior work shows that individuals can exhibit racial bias in prosocial behavior when they can attribute their (biased) actions to virtuous motivations (Danilov and Saccardo 2016). People may also behave more unfairly under conditions in which ambiguity allows the formation of self-serving beliefs (Exley 2016; Haisley and Weber 2010), and self-persuasion can be motivated by incentives (Ambuehl 2016). This study partially contradicts the licensing hypothesis, however. It is true that participants who were offered incentives felt significantly more strongly that they were free to reject the request (see the Online Appendix). But overall, incentives increased the willingness to donate to both the white and the black patient. The significant bias in the nakedincentive condition arises from the larger margin of increased willingness when the patient is white. If incentives license racially biased refusals, they would be expected to reduce (rather than to increase) mean willingness to donate to black patients. But this study was not designed to test the licensing hypothesis, and additional work is needed.

3.1.3. Conclusions. The results of experiment 1 are contrary to the initial hypothesis. Instead of reducing invidious racial bias in willingness to donate, as crowding-out theories would predict, the financial incentive increased the disparity between participants' willingness to donate to a white rather than a black patient. This experiment suggests that although financial incentives may not crowd out altruistic motivations, they may amplify the expression of aversive racism or in-group preferences. Although it is not possible to disaggregate the explanations described above, disinhibition of implicit bias, in-group favoritism, and shifts in the decision frame are likely candidates.

3.2. Experiment 2

The second vignette study shifts the focus to incentive size and source. This study presents largely the same vignette as above, but it pauses midway through the story, assessing willingness after each of two parts (parts A and B). Questions after part A examine willingness to donate to a black or white patient without an incentive but with messages appealing to altruism or social norms around donation. Part B continues the story by adding an incentive that varies in its size and source. By reassessing willingness after part B, this experiment aims to identify how the size and source of an incentive motivates changes in willingness to donate and whether those incentives interact with a patient's race. This study also expands the research on messaging described above by providing an altruism message-only condition and by integrating messaging about social norms.

The two-part vignette manipulation in this experiment presented a continuous story about Mr. Johnson, the same patient as above, who again was in need of a living kidney donor. Part A uses a $2 \times 2 \times 2$ manipulation. All participants were presented with a patient who was black or white, and the vignette presented the donation decision in purely altruistic terms (no incentive in any condition). Participants were also randomly assigned to see a message appealing to altruism (the same altruism message as above versus no message) and/or a message appealing to social norms that describes the proportion of adults willing to be organ donors ("A recent survey found that approximately two-thirds of US voters are open to being a living kidney donor to anyone, including people they do not know"; social norms message versus no message). The social norms

message is drawn from a recent survey of US voters' organ donation preferences (Peters et al. 2016). Although messaging is not my central focus, this study fills a gap left by the prior study by testing the impact of an altruism appeal alone, and it expands the inquiry to include social norms messaging. After reading part A, participants reported their willingness to donate and answered brief questions about the patient.

All participants then proceeded to part B, which continued the story by introducing a newly offered incentive. Every participant was offered an incentive in the part B vignette, but the size and source varied on a 3 \times 3 matrix. The size of the incentive was randomly assigned to be small (\$3,000), medium (\$18,800), or large (\$50,000) in comparison to the burden of the task. The large amount reflects a recent study on attitudes toward donation that suggests that \$50,000 is likely the largest incentive feasible without triggering concerns about undue inducement (Peters et al. 2016). The source of the incentive was randomly assigned to be a charity, a federal government program, or payment from the patient. An example vignette read as follows: "As you recall, Greg Johnson is on a waiting list for a kidney transplant. ... You are healthy and a match for Mr. Johnson.... In addition, a local charity has offered to pay you \$3,000 to compensate you for donating your kidney." After part B, participants again reported their willingness to donate and answered questions about the patient.

The sample was generally similar to that for experiment 1 and enrolled 993 new participants. Approximately 77 percent were between 18 and 45 years old, and 89 percent reported some college coursework or more. As with the prior sample, about 55 percent were female, and 58 percent were from the South and Midwest; 73 percent were non-Hispanic white. Around 47 percent identified as Christian, while 29 percent were atheist or agnostic. About 56 percent had a full-time job; about 35 percent had an annual income between \$20,000 and \$50,000, and the median household size was three.

Results of the messaging manipulations in the first vignette are presented briefly below, relying on OLS regression analyses that include the patient's race, appeal to altruism, appeal to social norms, and interactions among these three terms. The main findings of interest in this study, however, are the extent to which each incentive changed the willingness to donate between the two vignettes in the sequence. Change scores are therefore computed for the difference in willingness to donate between the vignette's first part (no incentive) and second part (incentive, varied by size and source). For ease of interpreting the main effects, analysis of covariance (ANCOVA) analyses are reported here for tests of the overall main effects and interactions. The analyses tested the main effects of the patient's race, incentive size, incentive source, and interactions with change in willingness to donate. Analyses also control for whether participants saw the altruism and/or social norms messaging in the first vignette. Subsequent pairwise comparisons probe significant main and interaction effects.

3.2.1. Results: Social Norms and Incentives. Regression analyses show no main effects of the patient's race ($\beta = -.04$; SE = .35; t(992) = -.12; p = .90) or altruism appeals ($\beta = -.30$; SE = .61; t(992) = -.48; p = .63) but a significant main effect for social norms messaging ($\beta = .79$; SE = .36; t(992) = 2.21; p < .05). Participants who read the social norms appeals were significantly more willing to donate. No interaction effects are observed among any of the three independent variables or in the three-way interaction. Analyses using perceptions of patient fault as the dependent variable show no significant main or interaction effects.

Approximately 45 percent of participants reported no change in willingness to donate when incentives were introduced, while 49 percent of participants became more willing to donate, and 7 percent of the sample became less willing. Observed mean change in the willingness to donate after the offer of an incentive was 1.09 on the 1–10 scale overall, from 4.73 (SD = 2.80) to 5.82 (SD = 3.02). A paired *t*-test comparing mean willingness before and after the incentive is significant overall (t(992) =-17.04; p < .0001) and significant at each of the three levels of incentive. Mean change scores differ by incentive size, at 1.59 for large incentives, 1.18 for medium incentives, and .52 for small incentives. Group means by the patient's race, incentive size, and incentive source are shown in Table 4. As before, incentives increased rather than diminished willingness to donate (that is, mean change scores are positive regardless of incentive size or source), showing little evidence of crowding out even with incentives that are small relative to the burden of the task.

Analyses of change scores corroborate findings from experiment 1. The ANCOVA analyses of change scores show significant main effects for the patient's race and incentive size but not for incentive source, as shown in Table 5. Change scores are significantly larger when the patient is white, which means that although incentives increased the motivation to donate overall, they yield a larger increase for donations to white com-

Incentive Size	Source	Ν	Mean
Black patients:			
Large	Charity	54	1.50 (2.10)
	Government	54	1.24 (1.88)
	Patient	55	1.82 (2.52)
Medium	Charity	55	1.09 (1.97)
	Government	57	.70 (2.30)
	Patient	58	.90 (2.34)
Small	Charity	59	.64 (1.84)
Small	Government	54	.35 (1.49)
	Patient	54	.41 (1.17)
White patients:			
Large	Charity	55	1.62 (2.29)
	Government	52	1.90 (2.45)
	Patient	54	1.46 (2.03)
Medium	Charity	54	1.07 (1.90)
	Government	55	1.27 (1.97)
	Patient	56	2.04 (2.36)
Small	Charity	55	.82 (1.65)
	Government	57	.44 (1.07)
	Patient	55	.44 (1.03)

Table 4. Change in Willingness to Donate: Group Means forExperiment 2

Note. Standard deviations are in parentheses.

pared with black patients. Incentive size also exerts a significant effect on change scores, as may be expected—the large incentive was significantly more motivating than either the medium or small incentive, and the medium incentive was significantly more motivating than the small incentive. All results are unchanged when controlling for the variables used in experiment 1—namely, willingness to be a deceased donor, altruism, the interaction between altruism scores and incentive size, implicit bias, explicit bias, and the interactions between each type of bias and the patient's race.

As Figure 1 demonstrates, the incentive was consistently more motivating when the patient was white compared with when he was black. Post hoc analyses, however, reveal an interaction between the patient's race and incentive size. Although the interaction term is not significant in the ANCOVA model, pairwise comparisons suggest that the main effect of race stems largely from bias in the medium-sized incentive condition. Medium incentives—here, \$18,800, the inflation-adjusted amount from experiment 1—motivate a significantly larger increase in willingness to

Source	Sum of Squares	Degrees of Freedom	Mean Square	F-Statistic	<i>p</i> -Value
Patient's Race	17.66	1	17.66	4.58	.0325*
Incentive Size	192.87	2	96.44	25.03	.000***
Incentive Source	6.27	2	3.14	.81	.44
Patient's Race \times Incentive Size	11.03	2	5.51	1.43	.24
Patient's Race × Incentive Source	5.36	2	2.68	.70	.50
Incentive Size × Incentive Source	16.62	4	4.16	1.08	.37
Patient's Race \times Incentive Size \times					
Incentive Source	27.89	4	6.97	1.80	.12
Norm Appeal	.04	1	.04	.01	.92
Altruism Appeal	2.97	1	2.97	.77	.38
Error	3,748.49	973	3.85		

Table 5.	Change in	Willingness to	Donate:	Analysis of	Covariance	Results for	Experiment	2
----------	-----------	----------------	---------	-------------	------------	-------------	------------	---

Note. $R^2 = .070$; adjusted $R^2 = .052$.

donate for the white patient compared with the black patient, with mean change scores of 1.47 versus .89 for white versus black patients, respectively (t(333) = -2.42; p < .05). Differences in change scores are not significant in pairwise comparisons when incentives are small (.56 versus .47 for white versus black patients; t(322) = -.56; p = .58) or large (1.66 versus 1.52; t(332) = -.58; p = .56). This effect may fit with crowding-out theories suggesting that once incentives are large enough, the relative price effect may overwhelm any interactions between the incentive and intrinsic motivation (Frey and Jegen 2001).

3.2.2. Conclusions. This experiment builds on the prior study to probe how incentive size and source may mediate the biasing impacts of incentives. In this two-part manipulation, results suggest that the crowding in of racial bias may be limited to incentives of medium size. One explanation may be that there is a biasing zone: incentives that are small in comparison to the task burden may not attract enough attention to distract aversive racists, trigger amoral reasoning, or motivate in-group favoritism or licensing. In contrast, incentives that are very large may be so attractive that they override any inherent biasing effect. Further testing is needed to replicate these effects in other altruistic settings and other populations.

^{*} *p* < .05.

^{***} p < .001.



Figure 1. Change in willingness to donate by incentive

4. LIMITATIONS

This study has several limitations. The experimental methods assess attitudinal outcomes, rather than behavior, which raises questions about whether these hypothetical preferences will predict real-world choices (List and Gallet 2001). The goal of this study, however, was to provide a preliminary test of the cognitive mechanisms by which incentives may interfere with motivation and perceptions. The central outcomes of interest at this stage are therefore attitudinal. The use of randomized methods in this study allows causal inferences about the directional impact of incentives on attitudes; identifying whether individuals will be organ donors was not a primary study goal. At the time of study's design, this work also contemplated that effect sizes may be small, requiring sample sizes that would be highly resource intensive in a laboratory setting. Future work will be needed to identify whether these attitudinal impacts are reflected in behavior in real-world settings. The findings use a significance level of p < .05 but do not adjust for multiple statistical tests; p-values are reported here to aid in interpretation.

Some suggest that research using distinctive black and white names may conflate discrimination based on perceived race and perceived socioeconomic status (Freyer and Levitt 2004). For the purposes of this study, however, the precise character of the invidious bias is not essential. It is of interest to test the effect of incentives on any type of invidious discrimination, whether by race or socioeconomic status; either (or both) provides for a test of how incentives affect the expression of bias.

Others consider the generalizability of research findings using MTurk.

The MTurk participants tend to be more demographically diverse than other Internet-based samples and more diverse than college student samples (Buhrmester, Kwang, and Gosling 2011; Casler, Bickel, and Hackett 2013; Paolacci and Chandler 2014). The attitudes of MTurk participants toward money are similar to attitudes among college students, and studies comparing MTurk workers, college samples, and community samples find similar decision-making biases (for example, Goodman, Cryder, and Cheema 2013; Hoffman et al. 2017). The MTurk samples may differ in some respects, however, from the US general population (Krupnikov and Levine 2014). The MTurk participants may also be less naive than other samples regarding experimental measures and goals (Chandler, Mueller, and Paolacci 2014), and some non-US individuals may obtain fraudulent Internet protocol addresses to participate in US-only studies.

Finally, the experiments are limited to the organ donation setting, which may be unique. Other prosocial activities, such as charitable donations, allocation of pro bono professional work, and volunteering have not been tested. Future work is needed to corroborate these findings with other prosocial tasks.

5. DISCUSSION AND CONCLUSIONS

Although motivational crowding-out theorists have long debated how financial incentives may interact with altruism, prior studies have not tested how incentives may simultaneously interact with invidious motivations such as racial bias. The two vignette studies in this paper suggest that financial incentives for altruistic tasks may crowd in racial bias, introducing or amplifying disparities in agents' willingness to undertake tasks based on the race of the beneficiary. This may be particularly important when the incentives are of intermediate size in relation to the burden of the tast, as shown in the second experiment. Several distinct cognitive processes may explain these findings: incentives may distract aversive racists from their efforts to be unbiased, incentives may lead agents to make decisions on the basis of self-interested calculus and render moral and social implications less salient, incentives may augment in-group favoritism, and incentives may license a decision about donating that is made on the basis of the patient's race. The present study was designed to test the overall effect of financial incentives on racial bias rather than to

disaggregate different causal mechanisms, but subgroup analyses provide tentative support for several of these potential pathways.

Although further research is necessary to disentangle the causal explanation for these findings, the current analyses have several potential implications. The findings contradict the crowding-out theory that financial incentives will decrease the willingness to be an organ donor, which undermines that particular rationale for barring incentives in organ donation policy (although other rationales may remain, such as commodification concerns or the fear that incentives will exploit poorer donors). Incentives uniformly increased the willingness to donate regardless of the patient's race. But the present findings on racial bias also suggest that where principals use financial incentives to motivate altruistic activity, it may be wise to conceal the race of the beneficiary or to design donation systems to counteract potential biasing effects of incentives. One design possibility may be to reserve incentives for nondirected donations, although this raises additional questions of whether people are equally willing to donate to anonymous and identifiable individuals, and it would have the drawback of excluding incentives for people who donate to family or friends. But to enable any use of financial incentives for living kidney donors, Congress would need to modify NOTA or authorize waivers of the current prohibition on purchases of organs (42 U.S.C. sec. 274e) for incentive-based demonstration projects (see, for example, language proposed by Goodwin [2009]); modifications to many state statutes would also be needed.

These findings may have greater implications beyond the organ donation setting and indeed beyond altruistic behavior. Financial incentives are a common tool of governments, employers, insurers, schools, and other principals seeking to influence agents' behavior. But if financial incentives encourage bias—including but not limited to racial bias—the thoroughgoing reliance on incentives in these settings may be suspect. This result may have implications for the design of incentives in fields as diverse as health-care delivery (for example, pay-for-performance incentives), policing (for example, civil forfeitures), prosecution (for example, incentives for conviction rates), voting (for example, incentives for voting), jury duty (for example, incentives for summons responses or deliberations), and property law (for example, good-landlord programs). At an extreme, the biasing effects of incentives may lead to racially disparate impacts, even when the design of an incentive system is facially neutral. Further research should examine how incentives interact with biases in other settings and how incentives interact with biases due to characteristics other than race.

These findings, however, also gesture toward debiasing strategies for incentive-based policies, such as bundling incentives with altruism-based information in the first experiment or manipulating the size of the incentive in the second experiment. These are just two of many options for modifying incentive architecture (Underhill 2016), and future work may identify other ways to limit incentives' biasing effects. When money and morals provide conflicting sources of motivation, well-designed incentives may yet make room for the crowd.

REFERENCES

- Ambuehl, Sandro. 2016. An Offer You Can't Refuse? Incentives Change What We Believe. Working paper. University of Toronto, Rotman School of Management, Toronto.
- Andreoni, James. 1990. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *Economic Journal* 100:464–77.
- Arrow, Kenneth J. 1972. Gifts and Exchanges. Philosophy and Public Affairs 1:343–62.
- Balliet, Daniel, Junhui Wu, and Carsten K. W. De Dreu. 2014. Ingroup Favoritism in Cooperation: A Meta-Analysis. Psychology Bulletin 140:1556–81.
- Bar-Gill, Oren, and Chaim Fershtman. 2005. Public Policy with Endogenous Preferences. Journal of Public Economic Theory 7:841–57.
- Becker, Gary S., and Julio Jorge Elias. 2007. Introducing Incentives in the Market for Live and Cadaveric Organ Donations. *Journal of Economic Perspectives* 21(3):3–24.
- Bénabou, Roland, and Jean Tirole. 2003. Intrinsic and Extrinsic Motivation. Review of Economic Studies 70:489–520.
- ——. 2006. Incentives and Prosocial Behavior. American Economic Review 96:1652–78.
- Benkler, Yochai. 2006. The Wealth of Networks: How Social Production Transforms Markets and Freedom. New Haven, CT: Yale University Press.
- Beretti, Antoine, Charles Figuières, and Gilles Grolleau. 2013. Using Money to Motivate Both "Saints" and "Sinners": A Field Experiment on Motivational Crowding-Out. Kyklos 66:63–77.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* 94:991–1013.
- Bilgel, Firat, and Brian D. Galle. 2014. Paying for Altruism: The Case of Organ

Donation Revisited. Working paper. Okan University, Faculty of Business and Administrative Sciences, Istanbul.

- Bolle, Friedel, and Philipp E. Otto. 2010. A Price Is a Signal: On Intrinsic Motivation, Crowding-Out, and Crowding-In. *Kyklos* 63:9–22.
- Bowles, Samuel. 2008. Policies Designed for Self-Interested Citizens May Undermine "The Moral Sentiments": Evidence from Economic Experiments. *Sci*ence, June 19, pp. 1605–9.
- Bowles, Samuel, and Sandra Polanía-Reyes. 2012. Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature* 50:368–425.
- Bratton, Charles, Kenneth Chavin, and Prabhakar Baliga. 2011. Racial Disparities in Organ Donation and Why. Current Opinion in Organ Transplantation 16:243–49.
- Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspectives on Psychological Science 6:3–5.
- Casler, Krista, Lydia Bickel, and Elizabeth Hackett. 2013. Separate but Equal? A Comparison of Participants and Data Gathered via Amazon's MTurk, Social Media, and Face-to-Face Behavioral Testing. Computers in Human Behavior 29:2156–60.
- Chandler, Jesse, Pam Mueller, and Gabriele Paolacci. 2014. Nonnaïveté among Amazon Mechanical Turk Workers: Consequences and Solutions for Behavioral Researchers. *Behavior Research Methods* 46:112–30.
- Choi, Stephen J., Mitu Gulati, and Eric A. Posner. 2014. Altruism Exchanges and the Kidney Shortage. *Law and Contemporary Problems* 77:289–322.
- Danilov, Anastasia, and Silvia Saccardo. 2016. Discrimination in Disguise. Working paper. University of Cologne, Graduate School in Management, Economics and Social Sciences, Cologne.
- Dovidio, John F., and Samuel L. Gaertner. 2004. Aversive Racism. Advances in Experimental Social Psychology 36:1–52.
- Everett, Jim A. C., Nadira S. Faber, and Molly Crockett. 2015. Preferences and Beliefs in Ingroup Favoritism. *Frontiers in Behavioral Neuroscience* 9, art. 15, pp. 1–21.
- Exley, Christine L. 2016. Excusing Selfishness in Charitable Giving: The Role of Risk. *Review of Economic Studies* 83:587–628.
- Feld, Lars P., and Bruno S. Frey. 2007. Tax Compliance as the Result of a Psychological Tax Contract: The Role of Incentives and Responsive Regulation. *Law* and Policy 29:102–20.
- Feldman, Yuval. 2011. The Complexity of Disentangling Intrinsic and Extrinsic Compliance Motivations: Theoretical and Empirical Insights from the Behavioral Analysis of Law. Washington University Journal of Law and Policy 35:11–51.

- Frey, Bruno S., and Reto Jegen. 2001. Motivation Crowding Theory. Journal of Economic Surveys 15:589–611.
- Freyer, Roland G., Jr., and Steven D. Levitt. 2004. The Causes and Consequences of Distinctively Black Names. *Quarterly Journal of Economics* 119:767–805.
- Galle, Brian. 2014. Tax, Command, . . . or Nudge? Evaluating the New Regulation. *Texas Law Review* 92:837–94.
- Gneezy, Uri, and Aldo Rustichini. 2000. A Fine Is a Price. *Journal of Legal Studies* 29:1–17.
- Goodman, Joseph K., Cynthia E. Cryder, and Amar Cheema. 2013. Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making* 26:213–24.
- Goodwin, Michele. 2004. Altruism's Limits: Law, Capacity, and Organ Commodification. *Rutgers Law Review* 56:305–407.
- 2009. Rethinking Federal Organ Transplantation Policy: Incentives Best Implemented by State Governments. Pp. 111–21 in When Altruism Isn't Enough, edited by Sally Satel. Washington, DC: AEI Press.
- Goodwin, Michele, and Nevin Gewertz. 2009. Rethinking Colorblind State Action: A Thought Experiment on Racial Preferences. Law and Contemporary Problems 72:251–81.
- Gordon, Wendy J. 2010. Discipline and Nourish: On Constructing Commons. Cornell Law Review 95:733–55.
- Grant, Ruth W. 2012. Strings Attached: Untangling the Ethics of Incentives. Princeton, NJ: Princeton University Press.
- Greenwald, Anthony G., Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology* 74:1464–80.
- Gross, Jed Adam. 2008. *E Pluribus* UNOS: The National Organ Transplant Act and Its Postoperative Complications. *Yale Journal of Health Policy, Law, and Ethics* 8:145–252.
- Haisley, Emily C., and Roberto A. Weber. 2010. Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior. *Games and Economic Behavior* 68:614–25.
- Health Resources and Services Administration. 2019. Organ Donation Statistics. https://www.organdonor.gov/statistics-stories/statistics.html.
- Hein, Grit, Giorgia Silani, Kerstin Preuschoff, C. Daniel Batson, and Tania Singer. 2010. Neural Responses to Ingroup and Outgroup Members' Suffering Predict Individual Differences in Costly Helping. *Neuron* 68:149–60.
- Hoffman, Elizabeth, David L. Schwartz, Matthew L. Spitzer, and Eric L. Talley. 2017. Patently Risky: Framing, Innovation, and Entrepreneurial Preferences. Working paper. Iowa State University, Department of Economics, Ames.
- Imas, Alex. 2014. Working for "Warm Glow": On the Benefits and Limits of Prosocial Incentives. *Journal of Public Economics* 114:14–18.

- Jolls, Christine, and Cass R. Sunstein. 2006. The Law of Implicit Bias. *California Law Review* 94:969–96.
- Kamenica, Emir. 2012. Behavioral Economics and Psychology of Incentives. Annual Review of Economics 4:427–52.
- Krupnikov, Yanna, and Adam Seth Levine. 2014. Cross-Sample Comparisons and External Validity. *Journal of Experimental Political Science* 1:59–80.
- Kulkarni, Sanjay, Keren Ladin, Danielle Haakinson, Erich Greene, Luhang Li, and Yanhong Deng. 2019. Association of Racial Disparities with Access to Kidney Transplant after the Implementation of the New Kidney Allocation System. JAMA Surgery 154:618–25.
- Lemm, Kristi M., Kristin A. Lane, David N. Sattler, Saera R. Khan, and Brian A. Nosek. 2008. Assessing Implicit Cognitions with a Paper-Format Implicit Association Test. Pp. 123–46 in *The Psychology of Modern Prejudice*, edited by Melanie A. Morrison and Todd G. Morrison. Hauppauge, NY: Nova Science.
- List, John A., and Craig A. Gallet. 2001. What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values? *Environmental* and Resource Economics 20:241–54.
- McConahay, John B. 1986. Modern Racism, Ambivalence, and the Modern Racism Scale. Pp. 91–125 in *Prejudice*, *Discrimination*, and *Racism*, edited by John F. Dovidio and Samuel L. Gaertner. San Diego, CA: Academic Press.
- Meier, Stephan. 2006. The Economics of Non-selfish Behaviour: Decisions to Contribute Money to Public Goods. Northampton, MA: Edward Elgar.
- Morgan, Susan, and Jenny Miller. 2002. Communicating about Gifts of Life: The Effect of Knowledge, Attitudes, and Altruism on Behavior and Behavioral Intentions regarding Organ Donation. *Journal of Applied Communication Research* 30:163–78.
- Neidich, E. M., A. B. Neidich, J. T. Cooper, and K. A. Bramstedt. 2012. The Ethical Complexities of Online Organ Solicitation via Donor-Patient Websites: Avoiding the "Beauty Contest." *American Journal of Transplantation* 12:43–47.
- Paolacci, Gabriele, and Jesse Chandler. 2014. Inside the Turk: Understanding Mechanical Turk as a Participant Pool. Current Directions in Psychological Science 23:184–88.
- Pearson, Adam R., John F. Dovidio, and Samuel L. Gaertner. 2009. The Nature of Contemporary Prejudice: Insights from Aversive Racism. Social and Personality Psychology Compass 3:314–38.
- Peters, Thomas G., Jonathan S. Fisher, Robert G. Gish, and Richard J. Howard. 2016. Views of US Voters on Compensating Living Kidney Donors. JAMA Surgery 151:710–16.
- Reese, Peter P., Neil Boudville, and Amit X. Garg. 2015. Living Kidney Donation: Outcomes, Ethics, and Uncertainty. *Lancet* 385:2003–13.
- Richins, Marsha L. 2004. The Material Values Scale: Measurement Properties and Development of a Short Form. *Journal of Consumer Research* 31:209–19.

- Rodrigue, James R., Abby Swanson Kazley, Didier A. Mandelbrot, Rebecca Hays, Dianne LaPointe Rudow, and Prabhakar Baliga. 2015. Living Donor Kidney Transplantation: Overcoming Disparities in Life Kidney Donation in the US-Recommendations from a Consensus Conference. *Clinical Journal of the American Society of Nephrology* 10:1687–95.
- Rudman, Laurie A., and Richard D. Ashmore. 2007. Discrimination and the Implicit Association Test. Group Processes and Intergroup Relations 10:359–72.
- Ryan, Richard M., and Edward L. Deci. 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary Educational Psychology 25:54–67.
- Sandel, Michael J. 2012. What Money Can't Buy: The Moral Limits of Markets. New York: Farrar, Straus, & Giroux.
- Satz, Debra. 2010. Why Some Things Should Not Be for Sale: The Moral Limits of Markets. New York: Oxford University Press.
- Smith, Robert J., Justin D. Levinson, and Zoë Robinson. 2015. Implicit White Favoritism in the Criminal Justice System. Alabama Law Review 66:871–923.
- Stepanikova, Irena, Jennifer Triplett, and Brent Simpson. 2011. Implicit Racial Bias and Prosocial Behavior. Social Science Research 40:1186–95.
- Sutton, Victoria. 2010. Is There a Doctor (and a Lawyer) in the House? Why Our Good Samaritans Laws Are Doing More Harm than Good for a National Public Health Security Strategy: A Fifty-State Survey. *Journal of Health and Biomedical Law* 6:261–300.
- Titmuss, Richard M. 1972. *The Gift Relationship: From Human Blood to Social Policy*. New York: New Press.
- Underhill, Kristen. 2016. When Extrinsic Incentives Displace Intrinsic Motivation: Designing Legal Carrots and Sticks to Confront the Challenge of Motivational Crowding-Out. Yale Journal on Regulation 33:213–79.
 - 2018. Money That Costs Too Much: Regulating Financial Incentives. Indiana Law Journal 94:1109–62.
- US Census Bureau. 2008. Frequently Occurring Surnames from the Census 2000. File A: Top 1000 Names. https://www.census.gov/topics/population/genealogy /data/2000_surnames.html.
- Vermeule, Adrian. 2012. Precautionary Principles in Constitutional Law. Journal of Legal Analysis 4:181–222.