CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

Discussion Paper No. 2023-12

Luise Goerges, Tom Lane, Daniele Nosenzo and Silvia Sonderegger

November 2023

Equal before the (expressive power of) law?

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Samantha Stapleford-Allen
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 74 86214
Samantha.Stapleford-Allen@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

Equal before the (expressive power of) law?

Luise Goerges, Tom Lane, Daniele Nosenzo and Silvia Sonderegger[a]

November 28, 2023

Abstract

Building on findings showing that laws exert a causal effect on social norms, this paper investigates whether this "expressive power of law" differs by gender or race. We develop a model to show that such differences are theoretically plausible. We then use an incentivized vignette experiment to test whether these differences are empirically relevant. Results from an online sample of around 4000 subjects confirm that laws causally influence social norms. However, we find little evidence of a differential effect across gender or race, suggesting that gender and race biases in the legal system are driven by other mechanisms than differences in the expressive power of law.

JEL codes: C91, C92, D9, K1, K42

Keywords: Social Norms; Law; Expressive Function of Law; Gender Gap; Racial Bias

---

[a]Goerges: Leuphana University Lüneburg, Germany; luise.goerges@leuphana.de. Lane: Newcastle University Business School; Tom.Lane@newcastle.ac.uk. Nosenzo: Department of Economics and Business Economics, Aarhus University; daniele.nosenzo@econ.au.dk. Sonderegger: School of Economics and Centre for Decision Research and Experimental Economics (CeDEx), University of Nottingham; silvia.sonderegger@nottingham.ac.uk. *Acknowledgements:* We received helpful comments from Billur Aksoy, Aleksandr Alekseev, Francesco Fallucchi, Pavitra Govindan, Hanna Hoover, Wieland Müller, Noemi Peter, Eli Spiegelman, and participants at several seminars and workshops. This work was supported by the Luxembourg Institute of Socio-Economic Research (LISER) and the Aarhus University Research Foundation (AUFF Starting Grant 36835). The project underwent ethical review and was approved by the Nottingham School of Economics Research Ethics Committee. We have no conflicts of interest to declare.

1 Introduction

All are equal before the law and are entitled without any discrimination to equal protection of the law – so pronounces Article 7 of the Universal Declaration of Human Rights. The declaration establishes the right to fair treatment and protection by the law regardless of one's race, gender, religion, and a host of other characteristics. Enshrined in the constitutional guarantees of countries around the world are similar principles. Yet, tragically, the everyday reality observed in the very same countries is of large and persistent differences in outcomes between individuals belonging to different sub-groups of the population at all stages of the criminal justice system. Race and gender gaps are prevalent. In the United States, black people are more likely to be stopped and searched (Coviello and Persico, 2015; Pierson et al., 2020), criminally charged (Berdejó, 2018), convicted (Anwar et al., 2012) and given harsher sentences (Rehavi and Starr, 2014; Beckett and Evans, 2016; Alesina and La Ferrara, 2014) than whites. Relative to men, women are under-represented at every stage of the justice system and receive lighter sentences (Butcher et al., 2017).

Understanding the reasons for these gaps is of paramount importance, and several different explanations have been put forward, most prominently discrimination (e.g., Knowles et al., 2001; Alesina and La Ferrara, 2014) and the incentives for crime resulting from socioeconomic disadvantage (Becker, 1968). In this paper, we hone in on another, thus far unexplored, possible channel, which focuses instead on the "expressive power of the law" (Sunstein, 1996; McAdams, 2015). This draws upon the literature, in law and economics, which argues that the power of the law in discouraging criminal activity depends only in part on its deterrent effect of altering the material costs and benefits of such behavior but also greatly relies on its ability to induce stigma towards actions which it deems illegal. This "expressive" function of the law – to causally influence social norms - has been suggested by theorists for some time (e.g., Posner, 1998, 2000, 2002; Bénabou and Tirole, 2011; van der Weele, 2012), but only recently has its existence received strong empirical support (Tankard and Paluck, 2017; Aksoy et al., 2020; Casoria et al., 2020; Galbiati et al., 2020; Lane et al., 2023).

We examine the possibility that the expressive power of law may differ across different population sub-groups. That is, the law may succeed in stigmatizing the same behavior to different extents, depending on the identity of the perpetrator or victim of the crime. For instance, laws may succeed in creating strong social norms proscribing particular illegal actions for white but not for black people or for women but not for men. If so, this could have powerful and asymmetric effects on the cost-benefit calculations individuals from different identity groups make when deciding whether to engage in crime, so long as we accept that the stigma resulting from norm violation plays its part in such calculations. We provide the first paper to explore this channel.

In Section 2, we outline a theoretical model which illustrates compelling reasons for why the law might exert substantially different normative effects on behavior by or towards those belonging to different groups. We follow previous work by Bénabou and Tirole (2006, 2011) and Lane et al. (2023) to provide a simple theoretical framework where the law exerts expressive power. In our model, individuals care about material incentives as well as the externalities their behavior imposes on others (i.e., they are "prosocial"). Moreover, they also care about what their behavior signals (to others) about their prosociality, as they receive "stigma" and/or "esteem" for it. We can think of "social norms" as functions that describe the social incentives (stigma, esteem) associated with behavior that is observable by others. In Lane et al. (2023), we argue that because laws introduce sharp payoff discontinuities between legal and illegal behaviors (e.g., through the material penalties to lawbreakers), social norm functions generally exhibit a discontinuity between legal and illegal behaviors: the esteem associated with selecting an action changes discontinuously once the action becomes illegal, even if the negative externality associated with the action is kept the same. We interpret the ability of law to introduce sharp discontinuity in the norms of a society as a manifestation of its expressive power.

Crucially, the magnitude of these discontinuities may partly depend on the (observable) identity of the decision-maker or potential victim. Focusing specifically on the decision-maker's identity, this dependence arises because the inferences observers make about a person's prosociality may be conditioned not only on their behavior but also on any observable characteristics of the person, such as their gender and race. In Section 2, we detail several mechanisms whereby stigma and esteem may systematically differ between people belonging to different identity groups. For instance, if the probability of conviction for committing a crime differs exogenously between identity groups, then the stigma for engaging in criminal behavior will also differ (and be higher for the group that is more likely to be convicted). As another example, if one group faces a higher probability of miscarriage of justice, then inferences by observers about a person's prosociality upon evidence of a conviction will be milder (because the conviction's signal is jammed). Importantly, these differences in stigma and esteem translate into *different social incentives* to engage in criminal activity, which in turn may help to explain why we observe the large and persistent differences in criminal justice outcomes between individuals belonging to different identity groups.

Note that some of these mechanisms are intimately related to the explanations the extant literature has provided to explain race and gender gaps in the criminal justice system. For instance, in the world of Gary Becker's economic model of crime (Becker, 1968), decisions to breach the law are taken by weighing up the material costs and benefits of doing so, and these calculations may differ across those of differing economic circumstances. Another first-order explanation is that law enforcement systematically discriminates against certain groups (Coviello and Persico, 2015; Luh, 2022). This may translate into more aggressive law enforcement against individuals belonging to those groups and, hence, a higher probability of (true and false) convictions. Our model clarifies that, in addition to their direct effects, socio-economic disadvantage and discrimination also have the capacity to distort the expressive power of the law. For example, we show that for groups that face a higher probability of false convictions (which can be due to discrimination), the expressive power of the law is weaker compared to groups that do not face discrimination.

Importantly, our model also clarifies that intra-group differences in social incentives (and hence in criminal justice outcomes) can arise even in the absence of discrimination and socio-economic advantages. For instance, if groups differ in the underlying distribution of prosociality, as the literature argues in the case of gender (e.g., Croson and Gneezy, 2009; Engel, 2011; Falk et al., 2018; Bilén et al., 2021; Exley et al., 2022), then our model predicts that the same instance of criminal behavior will lead to different inferences about the prosociality of the individual engaging in it, depending on whether they are a man or a woman.

In summary, our model reveals that several plausible theoretical mechanisms could give rise to systematic differences in the expressive power of law across different population sub-groups. In the remainder of the paper, we then undertake an empirical assessment of whether the expressive power of law does in fact differ across groups. This is done using an incentivized vignette experiment, which we describe in Section 3. The experiment follows the approach introduced in Lane et al. (2023). Subjects are presented with scenarios in which hypothetical individuals engage in particular behaviors and then report their beliefs about the social appropriateness of the described behavior in an incentivized norm-elicitation task of the type devised by Krupka and Weber (2013). Across different vignette versions, we vary whether the described behavior is legal or illegal.

Crucial to our identification strategy, in the vignettes, we describe behaviors regulated by *legal thresholds*, such as speed or drink-driving limits. Focusing on such behaviors allows us to cleanly identify the causal effect of laws on norms since we can rely on the (relatively mild) assumption that actions that are close to one another but fall on different sides of the legal threshold (e.g., driving at a 71mph or 69mph speed on a 70mph road) are almost identical in all respects except for their legal status. Thus, akin to the local randomization assumption in a regression discontinuity design, we can assume that sharp differences in the normative evaluation of behaviors that fall in close proximity but on different sides of a legal threshold are caused by the differences in legality and are thus a manifestation of the expressive power of law.

This paper's innovation is based upon furthermore varying the identity (race or gender) of the person in the vignette whose behavior is being evaluated – or, in some instances, the identity of another person in the vignette who is affected by the behavior. In this way, we separately measure the expressive power of laws to stigmatize illegal actions depending on whether the perpetrators (or, in some cases, their victims) are black or white or male or female.

We employ several different vignettes to study the expressive effects of various laws. We selected areas of behavior where statistics suggest racial differences in justice system outcomes or underlying economic conditions exist in the United States – for instance, drug and gun ownership or interactions involving minimum wage legislation – or where prior research indicates the likelihood of gender differences in choices, such as risky decision-making, prosociality, and parental leave. We ran online experiments with more than 4,000 subjects to test for differential effects involving race and gender, using separate samples drawn from participants who, at the time of the experiment, were residents in Florida (for race) and Texas (for gender).

Our results, reported in Section 4, suggest that, by and large, there are no differences in the expressive power of law across gender or race groups. While we replicate the findings in Lane et al. (2023) of generally strong effects of laws on norms, we find that in most cases, these effects are very similar regardless of race or gender. Our analysis finds in only one of 12 vignettes a significant difference across groups in the expressive power of law: gun laws stigmatize illegal firearm ownership more strongly for black owners than whites. Thus, while theoretically a candidate explanation for unequal group-level justice system outcomes, our study seems to largely rule out differential expressive power of the law in practice, implying that other channels fully, or at least overwhelmingly, account for the phenomena.

In Section 5, we further explore the reasons behind our null effects. Monte Carlo simulations show we are well-powered to detect meaningful effects if they exist. A subsequent auxiliary experiment assuages concerns that our vignettes' method of revealing race and gender is too subtle to make this salient to subjects. Two more compelling possibilities relate to our theoretical analysis. The first is that the expressive power of law is equally strong across sub-populations because conflicting mechanisms offset one another. In our model, for instance, discrimination against one group by law enforcement, which results in false convictions, blurs the signal conveyed by being convicted. It, therefore, weakens the expressive power of law for members of this group. However, this could be offset if the same group also receives discrimination, resulting in an increased likelihood of rightful convictions, since this makes illegal behavior more costly for members of this group and, in consequence, enhances for them the negative signal of a criminal record.

The second possibility is that contrary to our model's assumption, individuals *do not* condition their inferences about prosociality on the decision-maker's observable characteristics, but only on their behavior. While this may be surprising from a theoretical point of view (since observers would disregard potentially useful information), it aligns with the principle of equality before the law established by the United States' and many other countries' constitutions and that individuals may feel compelled to follow in their normative judgments. That is, although different sub-populations may not experience equal treatment before the law, they may be truly equal before its expressive power.

2 Theoretical framework

2.1 Model

Our model follows the literature on social image concerns (e.g., Bernheim, 1994; Bénabou and Tirole, 2006, 2011; Ellingsen and Johannesson, 2008; Andreoni and Bernheim, 2009) and in particular the framework developed by Lane et al. (2023). Although the theoretical mechanisms described here apply generally, the framework is specifically geared towards decision settings regulated by *legal thresholds* that establish the cut-off value above (or below) which a behavior becomes illegal. As explained earlier and further discussed in Lane et al. (2023), the focus on legal thresholds is crucial for identifying empirically the causal influence that laws exert on social norms, eschewing the reverse causality and spurious correlation concerns that are otherwise pervasive in the empirical literature on the topic.

In our model, individuals must decide whether or not to take a randomly drawn opportunity for material gain that they are faced with. Taking the opportunity imposes on others a negative externality, the severity of which varies across different opportunities. Opportunity o creates a negative externality of size $o \in [o^{\min}, o^{\max}]$ where $o^{\min} > 0$. To model legal thresholds, we assume that there is a threshold $\overline{o}$ above which taking an opportunity is illegal and that this is common knowledge among all agents.

Utility depends on material payoff, a psychological cost for imposing negative externalities on others, and on the social esteem that accrues to an individual when he/she takes or leaves opportunity o. We let individuals belong to different observable groups (representing gender and race). The utility of an individual belonging to group g who chooses action $a \in \{0, 1\}$ is

$$u_a(o;\theta,g) = \tau_g^1 a - \tau_g^0(1-a) - [p_g a + \pi_g\,(1-a)]K_g I_{o>\overline{o}} - o\theta a + S_g(o,a). \quad (1)$$

where $a = 1$ if the opportunity is taken and $a = 0$ if the opportunity is rejected.[1] The indicator function $I_{o>\overline{o}}$ takes a value of one if the opportunity

[1]In some of the applications of our empirical analysis, a continuous-action model may be more natural. Lane et al. (2023) characterize the conditions under which our key theory

o exceeds the legal threshold ($o > \bar{o}$) and takes a value of zero otherwise.

Expression (1) captures the idea that taking an externality-generating opportunity o allows the agent to earn a material return of $\tau_g^1 > 0$ instead of $\tau_g^0 < \tau_g^1$. However, if the opportunity exceeds the legal threshold, seizing it also leads to a material penalty $K_g > 0$ with probability $p_g \in (0, 1]$, the probability of being convicted for breaking the law. We also allow for the possibility that, when confronted with an illegal opportunity, the individual may be unjustly convicted of having seized the opportunity even when he/she actually rejected it.[2] This happens with probability $\pi_g \in [0, 1]$. Intuitively, we can think of π_g as capturing the amount of discrimination against group g. We assume that, when an individual seizes an illegal opportunity, the probability of being convicted is higher than when the individual leaves it: $\pi_g < p_g$. This ensures that the material return from seizing the externality-generating opportunity when there is a law forbidding it is lower than when seizing the opportunity is legal.

In addition to utility from material payoff, the individual also suffers a psychological cost of size $o\theta$ when his/her actions cause a negative externality o that is detrimental to other individuals. We denote as θ the extent to which the individual cares about imposing negative externalities on others. We refer to this as the individual's "type" and assume that it is privately known to the individual alone. We assume that types are drawn from a distribution with continuous differentiable density $f_g(.)$ with mean μ_g and full support $\left[\theta_g^{\min}, \theta_g^{\max}\right]$, where $\theta_g^{\min} \geq 0$, and that the psychological cost of causing an

results extend to the continuous-action case.

[2] We assume that individuals cannot be unjustly convicted unless they face an illegal opportunity. This is a simplifying assumption that is, however, immaterial. Intuitively, for an individual who faces a probability π_g of being unjustly convicted when presented with a legal opportunity, the *difference* in the utility from seizing and from rejecting the opportunity would be the same as in the main text, since the individual would be facing the same risk of conviction π_g both when seizing and when rejecting the opportunity.

externality is larger for higher types.

Finally, the last term in (1) captures the social incentives ("stigma" and "esteem") that accrue to the individual for taking or leaving an opportunity. These depend on the inferences that other individuals ("observers") make about an individual's type θ when they observe action a, the opportunity o the individual is presented with and whether or not is above $\bar{o}$, and the individual's group g.

We expect that the esteem/stigma conferred to an individual who selects an action a will typically depend on the individual's observable characteristics, such as race or gender. Intuitively, these characteristics affect the expected material payoff from different actions, as well as the distribution from which an individual's type is drawn. Consequently, when observers update their beliefs about the individual's type, they should take their gender and race into account. Formally, the esteem conferred to an individual of group g who seizes opportunity o is $S_g(o, 1) \equiv E(\theta \mid o, g, a = 1)$, while the esteem conferred to an individual who leaves opportunity o is $S_g(o, 0) \equiv E(\theta \mid o, g, a = 0)$.

2.2 Analysis

We analyze the model using Perfect Bayesian Equilibrium as our equilibrium concept and restricting attention to interior solutions. Moreover, we assume monotonicity: we focus on equilibria where opportunities that generate stronger negative externalities are seized by a smaller share of individuals who are, on average, less prosocial (lower θ) than the individuals who seize opportunities with weaker negative externalities. All proofs are relegated to Appendix A.

Consider an individual of type θ and group g, and let $t_g \equiv \tau_g^1 - \tau_g^0 > 0$

denote the net material earning from seizing the opportunity. Taking social esteem as given, the *net utility* from seizing the opportunity, rather than rejecting it, is given by

$$u_1(o;\theta,g) - u_0(o;\theta,g) = t_g - \theta o - (p_g - \pi_g)\,K_g I_{o>\bar{o}} + S_g(o,1) - S_g(o,0),$$

decreasing in θ. For each opportunity o and group g, we can therefore identify the highest θ who takes o. We denote this as $\widehat{\theta}_g^o$ for legal opportunities and $\widetilde{\theta}_g^o$ for illegal opportunities. In equilibrium,

$$S_g(o,1) = \begin{cases} \mathcal{M}_g^-(\widehat{\theta}_g^o) \text{ if } o \leq \bar{o} \\ \mathcal{M}_g^-(\widetilde{\theta}_g^o) \text{ if } o > \bar{o} \end{cases} \text{ and } \; S_g(o,0) = \begin{cases} \mathcal{M}_g^+(\widehat{\theta}_g^o) \text{ if } o \leq \bar{o} \\ \mathcal{M}_g^+(\widetilde{\theta}_g^o)\text{if } o > \bar{o} \end{cases} \tag{2}$$

where $\mathcal{M}_g^-(\theta^o) \equiv E\left(\theta \mid \theta < \theta^o, g\right)$ and $\mathcal{M}_g^+(\theta^o) \equiv E\left(\theta \mid \theta > \theta^o, g\right)$. In what follows, we will primarily focus on $S_g(o,1)$, namely the esteem afforded to an individual who seizes an opportunity o since this is the focus of our empirical investigation. However, it should be clear that our results are based on the relationship between $\widehat{\theta}_g^o$ and $\widetilde{\theta}_g^o$, and, hence, they also apply straightforwardly to $S_g(o,0)$.

Using the notation $\Delta_g\left(\theta^o\right) \equiv \mathcal{M}_g^+(\theta^o) - \mathcal{M}_g^-(\theta^o) > 0$, the threshold type seizing a legal opportunity $o \leq \bar{o}$ satisfies

$$t_g - \widehat{\theta}_g^o o - \Delta_g(\widehat{\theta}_g^o) = 0. \tag{3}$$

while the threshold type seizing an illegal opportunity $o > \bar{o}$ satisfies

$$t_g - \widetilde{\theta}_g^o o - (p_g - \pi_g)\,K_g - \Delta_g(\widetilde{\theta}_g^o) = 0. \tag{4}$$

Under our assumptions, $\widetilde{\theta}_g^o < \widehat{\theta}_g^o$ for all o and g.[3] For a given opportunity o, the threshold type seizing o is higher if o is legal than if it is illegal. We now compare behavior when an individual is confronted with an opportunity $\overline{o}-\varepsilon$, which is marginally legal, or with an opportunity $\overline{o}+\varepsilon$, which is marginally illegal, for a vanishingly small ε.

Proposition 1 (Lane et al., 2023) *The esteem function $S(o,1)$ exhibits a downward discontinuity at $\overline{o}$:*

$$D_g \equiv \lim_{\varepsilon \to 0} \left[S_g(\overline{o}-\varepsilon,1) - S_g(\overline{o}+\varepsilon,1)\right] = \mathcal{M}_g^-(\widehat{\theta}_g^{\overline{o}}) - \mathcal{M}_g^-(\widetilde{\theta}_g^{\overline{o}}) > 0.$$

Intuitively, the expected net material payoff of seizing an illegal opportunity is discontinuously smaller than that of a legal opportunity for $K_g > 0$ and $p_g \in (0,1]$: $(t_g - (p_g - \pi_g)\,K_g < t_g)$. This implies that the pool of types willing to take a marginally illegal opportunity is discontinuously worse than the pool of types who take a marginally legal opportunity. Observers recognize and take this into account when forming beliefs about an individual's type. Therefore, seizing $\overline{o}+\varepsilon$ carries discontinuously less esteem than seizing $\overline{o}-\varepsilon$, despite these two opportunities generating very similar externalities. This discontinuity in the social incentives faced by individuals confronted with (marginally) legal and illegal opportunities is a manifestation of the expressive power of law. Laws exert expressive power on society by introducing sharp discontinuities in the stigma and esteem that individuals obtain for engaging in legal or illegal behaviors, which is the key result reported in Lane et al. (2023).

[3]Specifically, this follows from monotonicity and our focus on interior solutions. A sufficient condition for monotonicity is that, for all θ_o, $o_{\min} > -\Delta_g'(\theta_o)$. A sufficient condition for interior $\widehat{\theta}_g^o$ and $\widetilde{\theta}_g^o$ is that $t_g - o^{\max}\theta_g^{\min} - (p_g - \pi_g)K_g - \mu_g + \theta_g^{\min} > 0 > t_g - o^{\min}\theta_g^{\max} + \mu_g - \theta_g^{\max}$.

2.3 The Expressive Power of Law Across Gender and Race Groups

The innovation of this paper compared to Lane et al. (2023) is that our model allows observers to condition their inferences about an individual's type on his/her observable characteristics, such as gender and race. In turn, this allows for social incentives to differ between individuals belonging to different groups g. In this subsection, we show a number of plausible mechanisms that can give rise to such differences.

We start by performing a series of simple comparative statics exercises to illustrate how the size of the discontinuity of Proposition 1 varies with different parameters of our model for a given group g.[4] We next argue that these parameters may indeed vary for individuals belonging to different groups g. Hence, the expressive power of law may not be equal across groups.

2.3.1 Comparative Statics

Our comparative statics analysis focuses on three factors: (1) the probability of conviction p_g and severity of sanctions K_g; (2) the probability of false conviction π_g; and (3) the net material gain from taking an opportunity t_g.

Probability of conviction and severity of sanctions. The expected material payoff of taking an illegal opportunity decreases in the size of the penalty K_g and in the likelihood of being convicted p_g. This, in turn, worsens the pool of types willing to engage in illegal behavior. The expected material payoff of taking a legal opportunity is instead unchanged, and so is the pool of types willing to take it. Observers take this into account when drawing inferences about the individual's type, which produces larger discontinuities

[4]Some of these comparative statics are already introduced in Lane et al. 2023.

for crimes that are associated with a larger K_g and/or p_g.

Lemma 1: *Ceteris paribus, D_g is increasing in K_g and in p_g.*

Probability of false conviction. A higher probability of being convicted even when an individual does not seize an illegal opportunity lowers the expected return of rejecting the opportunity and, therefore, in relative terms, increases the appeal of seizing it. This improves the pool of types willing to engage in illegal behavior and reduces the size of the downward discontinuity in social esteem at the legal limit.

Lemma 2 *Ceteris paribus, D_g is decreasing in π_g.*

Material return from the opportunity. A higher net material return from seizing the opportunity makes any opportunity more attractive. Therefore, *both* the pool of types willing to seize a legal opportunity and the pool of types willing to seize an illegal opportunity improve. This exerts countervailing forces on the size of the discontinuity at the legal limit. The end result is ambiguous.

Lemma 3 *Ceteris paribus, D_g may be increasing, decreasing or invariant in t_g.*

2.3.2 Comparing actors from different groups

We now look at the implications of our analysis for the esteem obtained by individuals from different groups. In general, it is not implausible that groups may systematically differ in the three factors discussed above. In fact, the existing literature has advanced explanations for the observed gender and race gaps in criminal justice outcomes that are closely related to these factors. Differences in socio-economic disadvantage generate different material incentives to engage in criminal activity in the standard economics approach to crime (Becker, 1968). This corresponds to differences in the net material returns t_g for taking an externality-generating opportunity. Discrimination can be modeled as stricter law enforcement against the discriminated group, which in turn may translate into more severe sanctions against that group (K_g), as well as into a higher likelihood of arrest and conviction both in case the discriminated individual has committed the crime (p_g) and when he/she is innocent (π_g).

Our model clarifies that these phenomena also have implications regarding the social incentives individuals face to commit crimes. Consider two groups g and g' that differ in the severity of the penalty incurred when convicted and in the probability of being convicted when breaking the law (and this is known to observers). A direct implication of Lemma 1 is that the discontinuity in social stigma between legal and illegal actions is more pronounced for the group that faces stronger penalties and higher conviction probability.

Corollary 1 *Suppose that $K_g \geq K_{g'}$ and $p_g \geq p_{g'}$ with at least one strict inequality, while in all other respects group g and g' are identical. Then, D_g $> D_{g'}$.*

In contrast, suppose that discrimination takes the form of a higher likelihood of a *false* conviction. The two groups g and g' now differ (only) in π_g. It follows from Lemma 2 that the group facing a higher π_g also faces a smaller discontinuity in stigma at the legal limit.

Corollary 2 *Suppose that $\pi_g > \pi_{g'}$, while in all other respects group g and g' are identical. Then, $D_g < D_{g'}$.*

Interestingly, taken together, Corollary 1 and 2 also imply that if discrimination takes the form of *both* higher penalties and (rightful) conviction probability *and* higher false conviction probability, the relationship between D_g and $D_{g'}$ is ambiguous, as the two effects offset one another.

Moreover, Lemma 3 shows that in our framework, socio-economic disadvantage may also not have a clear-cut effect on social incentives. Individuals from a disadvantaged group will typically face a higher net material gain from seizing externality-generating opportunities because the payoff they enjoy if they do not seize the opportunity is especially low. Lemma 3 shows that this may result in stronger or weaker discontinuities in social stigma at the legal limit, with ambiguous effects on behavior.

Corollary 3 *Suppose that $t_g > t_{g'}$, while in all other respects group g and g' are identical. The relationship between D_g and $D_{g'}$ is ambiguous. We may have $D_g > D_{g'}$, $D_g < D_{g'}$ or $D_g = D_{g'}$.*

Our analysis so far illustrates how previous explanations for gender and race gaps in criminal justice outcomes (discrimination, socio-economic disadvantage) may produce differences in social incentives between gender and racial groups. However, our model can generate differences in social incentives between groups even in the absence of discrimination and socio-economic disadvantage. In particular, consider the distribution of types in a group, $f_g(\theta)$. In our previous analysis, we have kept this constant across groups. However, in practice, this may differ between gender and/or racial groups. For instance, there is some evidence that women are (and are believed to be) on average more prosocial than men (e.g., Croson and Gneezy, 2009; Engel, 2011; Falk et al., 2018; Bilén et al., 2021; Exley et al., 2022).

Differences in the type distribution may impact the size of the discontinuity at the legal threshold through two channels. First, the type distribution determines the highest types seizing a marginally legal and a marginally illegal opportunity. Second, the type distribution determines the expected type conditional on being lower than a given threshold and, thus, the esteem afforded to an individual who seizes a marginally legal or illegal opportunity. Generally, we therefore expect that when the distributions of types for two groups g and g' differ, the size of the downward discontinuity at the legal threshold will also differ. This also generally applies to distributions that belong to the same family but differ in their mean and/or variance.[5] The following lemma focuses on a natural statistic for comparing different groups, the mean of the type distribution. As the lemma shows, looking at the relationship between the means of two distributions g and g' provides little guidance on the relationship between D_g and $D_{g'}$. This is true even if the two groups are identical in all other respects except for their type distributions.

Lemma 4 *Consider two groups g and g' that are identical in all respects except for their type distributions f_g and $f_{g'} \neq f_g$. Denote as μ_g and $\mu_{g'}$ the respective means. (i) Even if $\mu_g = \mu_{g'}$, we may have $D_g \neq D_{g'}$; (ii) Suppose that $\mu_g > \mu_{g'}$. Both $D_g > D_{g'}$ and $D_g < D_{g'}$ are possible, depending on the nature of the type distributions.*

Finally, note that while our discussion has focused on the identity of the perpetrators, the individual(s) who suffer the negative externality when the opportunity is seized may also belong to different groups. This may also trigger differences in social incentives. For instance, a crime where the victim belongs to a disadvantaged group may be punished less harshly (and may attract fewer resources devoted to identifying the perpetrator) than if the victim belongs to a privileged, dominant group. In this case, the perpetrator may face a lower probability of being convicted when breaking the law and/or

[5] An exception is the case of uniform distributions, where $\Delta_g(\theta)$ is a constant and $\mathcal{M}_g^-(\theta)$ is linear, so that $D_g = D_{g'} = (p-\pi)K/\bar{o}$ for all g and g' with uniformly distributed types.

a lower penalty when convicted, with effects that are analogous to those discussed in Corollary 1. As another example, the return for the perpetrator from seizing the opportunity may differ depending on the group of the victim, with effects analogous to those discussed in Corollary 3. Thus, in our model, differences in social incentives may arise not only across different groups of perpetrators but also across perpetrators who commit crimes against different groups of victims.

3 Experimental design

In order to measure whether the law exerts a uniform expressive power across gender and race groups, we designed an experiment following the approach introduced by Lane et al. (2023). This involves using vignettes - hypothetical scenarios in which a person behaves in a particular way - to evaluate the social appropriateness of this behavior. Between subjects, we vary some feature of the behavior described, allowing the estimation of a norm function which maps how changes in social appropriateness result from changes in this feature of behavior. These norm functions embody the stigma and esteem that accrue to a person for taking actions that are deemed socially appropriate or inappropriate.

The key to the approach is that we focus on types of behavior that are regulated by legal thresholds, where the feature of behavior we vary is the one that determines on which side of the law the behavior falls. We study behaviors very close to the threshold but on either side of it, and are therefore able to identify causal effects of laws on norms by testing for differences in appropriateness between behaviors which are in all respects almost identical except for their legal status. This allows us to overcome the empirical difficulty in unraveling cause and effect that has typically been an obstacle to scholars investigating relationships between laws and norms.

To explain in more detail, we designed different versions of our vignettes, which placed the hypothetical person's behavior at one of four possible points on either side of the legal threshold. For instance, in our Drink-Driving Vignette, we described a person who had been drinking alcohol at home and then drove to a bar and varied whether the person had a blood-alcohol content (BAC) of 0.076%, 0.077%, 0.078%, 0.079%, 0.081%, 0.082%, 0.083% or 0.084% at the time of driving. Given that the legal limit for driving is 0.08%, this creates variation in whether or not the described behavior is legal and how far from the threshold it lies.

By eliciting the social appropriateness of this behavior at each BAC level, we can estimate appropriateness as a function of BAC and, importantly, observe whether this function exhibits a discontinuity as BAC crosses the legal limit. In line with the theoretical analysis of the previous section, we infer such a discontinuity to represent the causal effect of the law on the norm, i.e., the difference in appropriateness between two otherwise identical actions lying on either side of the threshold at arbitrarily low distances from it. This identification strategy rests on a very mild assumption – akin to the local randomization assumption employed in regression discontinuity designs with naturally occurring data – that, apart from its legality, there is no other variable associated with the behavior (e.g., the potential harm to bystanders) which exhibits a discontinuity when behavior crosses the legal threshold (see Lane et al., 2023, for further discussion on this point).

The novelty of the current study is that in the vignettes, we also varied the identity of the fictional person whose behavior was to be evaluated – or, in some cases, the identity of another fictional person who was affected by this behavior. In the Gender Experiment, we varied whether this person was described as male or female. In the Race Experiment, we varied whether they were described as African American or White American. In both cases, identity was varied between-subject, so as to minimize experimenter demand effects.

For each vignette, we can, therefore, plot two different norm functions, depending on the identity of the person described, and separately estimate the discontinuity at the legal threshold for each. Our primary interest is in whether these discontinuities differ in strength depending on the gender or race that is manipulated; if they do, we can conclude that laws exert different causal effects on norms regulating identical sets of behavior but involving different identity groups of people. These differences in causal effects could stem from any of the mechanisms described in the theoretical analysis of Section 2.

We incentivized norm-elicitations using the method of Krupka and Weber (2013). For each vignette, subjects were told to report how socially appropriate they considered the described behavior by selecting one option on an ordered scale, ranging from 'Very socially appropriate' to 'Very socially inappropriate.'[6] Note that since social appropriateness was explained to subjects to refer to behavior that "you think most people would agree is the right thing to do," the task asks subjects to report second-order beliefs (rather than personal beliefs) about appropriateness, which reflects how social norms have generally been conceptualized (e.g., Bicchieri, 2006; Krupka

[6]The scale contained either four items (half of the sessions in the Gender Experiment) or six items (Race Experiment and the other half of the sessions in the Gender Experiment). Both types of scales have been used in the previous literature. We moved from a four- to a six-point scale to address concerns that the four-point scale would not have enough granularity to detect small differences in responses. In fact, our data show no systematic differences in responses across versions of the scale (see Appendix F).

and Weber, 2013).[7]

For incentive compatibility, subjects were eligible to earn a bonus payment from their evaluation of behavior in a vignette only if their rating was the same as that selected by the most other subjects who saw the same (version of the) vignette. This transforms the task into a coordination game and provides material incentives for subjects to truthfully report beliefs about how behavior is regarded within society, assuming that such truthful revelation represents the salient coordination strategy (see Krupka and Weber, 2013).[8]

3.1 Vignettes

For each experiment, six vignettes were devised to study the effects of laws on norms covering a range of different types of behaviors. In each case, four of these vignettes manipulated the identity of the person whose behavior was to be evaluated, while the other two manipulated the identity of a second person affected by this behavior. Half of these vignettes were adapted from

[7]Our instructions provided subjects with a lengthy explanation of what we meant by socially appropriate behavior. Here, we emphasized that subjects should not necessarily understand it to mean the equivalent of appropriate behavior in the eyes of the law. Screenshots from the experiments showing the instructions subjects received are available in Appendix B.

[8]The Krupka-Weber technique has been the subject of methodological discussion (see Görges and Nosenzo, 2020; Fallucchi and Nosenzo, 2021). A debate has centered on whether the incentive to coordinate responses does indeed result in subjects revealing true beliefs about social appropriateness of behavior or whether it may instead result in them attempting alternative strategies to match one another's answers. Such concerns are particularly relevant to our current design, as subjects could use legality to coordinate their answers (i.e., report all legal behavior to be socially appropriate and all illegal behavior to be socially inappropriate, even if this does not reflect their actual beliefs). In Lane et al. (2023) we tested the effects of laws on norms using both the Krupka-Weber method and an alternative 'opinion-matching' method, which did not rely on the use of a coordination game. The two methods produced virtually identical results. This, along with other evidence on the robustness of the Krupka-Weber method (Fallucchi and Nosenzo, 2021), strongly suggests it is an appropriate method for measuring norms in the contexts our paper considers.

those in Lane et al. (2023), while the others were newly introduced for this study. We deliberately selected settings where prior literature has established group-level differences in behavior or outcomes.

For each experiment, there were a total of 16 different versions of each vignette, with 8 different conditions for the precise behavior (4 on each side of the legal threshold) crossed with 2 group identity conditions (male or female in the Gender Experiment; African American or White American in the Race Experiment). We employed a between-subject design, i.e., each subject was only exposed to one randomly selected version of each vignette. [9]

3.1.1 Gender Experiment

In the Gender Experiment, we study three sets of behaviors representative of traits and preferences for which gender differences are common or commonly suspected. The first set relates to risk-taking and risky decision-making, for which gender differences have been established (e.g., Dohmen et al., 2011; Falk et al., 2018). We used the Drink-Driving Vignette as described above, varying whether the person driving to the bar was male or female. We also used a Speeding Vignette in which a person is driving on a highway; we varied the gender of the driver and whether his/her speed was above or below the legal limit.

Our second set of vignettes is focused on other-regarding behavior. Two vignettes devised settings where a person did business with a youth either just below or above the minimum legal age for engaging in the activity. In the Alcohol to Youth Vignette, a storekeeper sells alcohol to a young customer; we varied whether the youth is a few days younger or older than the minimum drinking age and whether the storekeeper is male or female.

[9] See Appendix C for the full wording of all vignettes.

The Casino Vignette is similar but instead focuses on a casino employee admitting a young customer onto the premises, where we vary the gender of the employee and whether the customer is a few days above or below the legal age for gambling. In both cases, an important consequence of illegal behavior (doing business with a minor) is that it may have negative effects on the minor and potentially others in society, so the decision to engage in such behavior may be related to a person's other-regarding preferences. As mentioned earlier, previous literature has found females to be more other-regarding than males (e.g., Croson and Gneezy, 2009; Engel, 2011; Falk et al., 2018).

Finally, we designed two vignettes focused on labor market interactions, motivated by findings of gender differences in labor market outcomes (e.g., Blau and Kahn, 2017). In these vignettes, we varied the gender of someone affected by the action of the person whose behavior subjects evaluated. In the Minimum Wage Vignette, a manager hires a custodial worker; we varied the gender of the worker and the hourly wage paid by the manager so that it stood at either a few cents above or below the legal minimum. In the Parental Leave Vignette, a manager receives a request for a period of unpaid leave from an employee expecting his/her first child and responds by offering the employee a certain number of weeks of unpaid leave, after which the employee must return to work or be fired; we varied whether the employee was male or female and the number of weeks the manager offered such that it was either just above or below the legal minimum number of weeks of unpaid, job-protected leave.

3.1.2 Race Experiment

We reused three vignettes from the Gender Experiment: the Drink-Driving, Speeding, and Minimum Wage Vignettes. In this experiment, we varied whether the drivers in the Drink-Driving and Speeding Vignettes, and the custodial worker in the Minimum Wage Vignette, were described as African American or White American (their gender was fixed as female in all cases).

We then added three new vignettes. In the Gun Possession Vignette, a father buys a gun and gives it to his son; we varied the son's age, such that he was either a few days below or above the legal minimum age for gun possession, and also whether the father was African American or White American. In the Marijuana Vignette, a male medical marijuana user buys some marijuana from a licensed outlet; we varied whether the user was African American or White American and also varied the user's existing stock of marijuana such that his new purchase would take his total possession to a level either just below or above the legal limit. In the Age of Consent Vignette, a male college student has sex with a female high school student, whose age we vary such that she is either a few days above or below the legal age of consent. In this vignette, we ask subjects to evaluate the behavior of the college student, whose race is fixed as African American, while we vary whether the high school student is presented as African American or White American.

These vignettes are motivated not by expected racial differences in traits and preferences but by economic or crime statistics showing racial differences in outcomes. African Americans earn lower wages than White Americans on average (Gould, 2021), are more likely to be apprehended while driving (Stanford Open Policing Project, 2021), more likely to be arrested for drug offenses (Mitchell and Caudy, 2015), and more likely to be arrested, charged and convicted of rape (FBI, 2021; Shaw and Lee, 2019), while guns bought by African Americans are more likely to later be identified as involved in crimes than those bought by White Americans (Koper, 2014). In the case of the Age of Consent Vignette, we are also motivated by the appalling history of vigilante justice enacted against African American men who have engaged in sexual activity with specifically white women, indicating that the race of the woman has been treated as a relevant criterion concerning the appropriateness of the man's behavior in such cases (see also Alesina and La Ferrara, 2014, who study capital punishment appeals in the US and find that courts give out more severe sentences to minority defendants who killed white victims).

3.1.3 Presentation of race/gender to participants

In the Race Experiment, the race of the hypothetical person whose identity we manipulated was conveyed to subjects so that this information would be salient yet presented naturally. We explicitly told subjects the race of the person but provided this information along with other details that we held constant across versions of the vignette (for instance, in the Minimum Wage Vignette, the custodial worker is described as a 40-year-old, White/African American woman, living in a small town, who has been unemployed for six months). We also gave the person a name likely associated with a particular race. We selected 10 strongly American American-associated names and ten strongly White American-associated names for each gender for use in our experiment following a pilot study.[10]

Noting that names might be signifiers not only of race but also of socioeconomic status (Bertrand and Mullainathan, 2004), we ran a second pilot to measure perceptions of socioeconomic status associated with each of the

[10]See Appendix E for details and Appendix B.5 for screenshots.

names chosen (screenshots provided in Appendix B.6. This found there was indeed variation across the chosen names in perceived status (see Appendix E). We, therefore, designed the experiment so that we could account for such differences by randomizing which of the ten selected names was presented to a subject and controlling for the socioeconomic status associated with that name in our analyses.

In the Gender Experiment, we did not use names but simply described the person in the vignette as ‘a man’ or ‘a woman’ alongside some other personal details (age, occupation), as we believed subjects would consider such descriptions perfectly natural.

3.1.4 Additional vignettes and further design details

In addition to the six vignettes designed to measure the effects of laws on norms, subjects in each experiment were presented with nine filler vignettes, whose purpose was to obfuscate our research objective by reducing the salience of legal thresholds and the two specific racial groups studied in our Race Experiment. The fillers described actions that were unregulated by laws (such as choosing whether to give money to charity) or were regulated but not by means of a threshold (such as driving without wearing a seatbelt). In the Race Experiment, we presented the names and race or ethnicity of the characters in the fillers but included Asian Americans and Hispanic Americans in addition to African Americans and White Americans. We did not manipulate the fillers, so all subjects in a given experiment saw the same versions.

Therefore, each subject in our experiments evaluated a total of 15 vignettes. These appeared in random order, except for three fillers, which were always the first three subjects encountered. We placed these at the beginning for training purposes, to teach subjects to differentiate social appropriateness from legality. The vignettes described behavior that was unlikely to be considered very inappropriate and included cases where the behavior was regulated but legal, regulated and illegal, or not regulated by law.

After subjects completed the last vignette, they filled out a questionnaire that collected their demographic information. At this point, we also added questions based on the US General Social Survey (GSS, Smith et al., 2019) to measure subjects' levels of sexism in the Gender Experiment (eight questions, as used in Charles et al., 2022), and four GSS questions to measure racism in the Race Experiment.

Each of our vignettes was carefully worded so as to subtly remind subjects of the relevant laws and, therefore, inform them whether the behavior to be evaluated was legal or not. The vignettes also made it clear that the fictional character under evaluation knew the law (for instance, in the Parental Leave Vignette, the manager consults the company's legal department before acting). This was done to remove any possible influence that (perceived) ignorance of the law might exert on our results. The full wording of our vignettes can be found in Appendix B.

3.2 Implementation

Both experiments were run online. For the Gender Experiment, subjects from Texas were recruited; the Race Experiment enlisted subjects from Florida. We restricted each experiment to one state because many laws, including some of those in our vignettes, differ across state lines in the US. Our vignettes were specifically tailored to the laws and legal thresholds of the relevant state (for example, our Marijuana Vignette was based on the Floridian law that allows marijuana for medical usage and sets a possession limit of 4 ounces for users), and we made it clear to subjects through the use of place names that the scenarios were assumed to be taking place there. Restricting participation to those from the given states ensures that subjects evaluate behavior regulated by real laws that also apply to them and that they are likely to be familiar with state-specific norms. It was made clear to subjects that the other participants, with whom they needed to coordinate, had been recruited in the same way and, therefore, were also from the state.

The states were chosen on the basis of two criteria. First, Texas and Florida are both populous, facilitating the online recruitment of large sample sizes. Second, previous research has identified Texas as a relatively sexist state (Charles et al., 2022) and Florida as a relatively racist one (Stephens-Davidowitz, 2014).

The Gender Experiment was run between December 2019 and August 2020, while the Race Experiment was conducted in June and July 2020. Each experiment was programmed in Qualtrics and distributed in separate waves using the recruitment platforms Prolific and CloudResearch. Subjects recruited on CloudResearch received a participation fee of \$1 and were eligible to earn a bonus of \$4 on the basis of their performance in the norm-elicitation task. It was explained to subjects that after the experiment, we would randomly pick one of the 15 vignettes and would pay them the bonus if they had selected the most common response for that vignette. As Prolific requires the use of British currency, for subjects recruited via this platform we adjusted the participation fee to £1 and the bonus to £4.[11]

In total, we received 2,516 completions for the Gender Experiment and 2,447 for the Race Experiment. As we could identify respondents only by

[11] At the time of our experiment, the exchange rate was roughly \$1 = £0.80. We considered it more natural to hold constant the numerical payoffs as integers rather than convert them into values introducing decimal points.

their CloudResearch or Prolific ID, we could not, by design, exclude respondents who had already completed the experiment on the other platform. However, we included a question at the end of the experiment asking subjects if they were registered on the other platform and, if they answered yes, whether they had already participated in the study there (to encourage truthful responses, we assured subjects this would not affect their payment). Roughly 18 percent of subjects reported being registered on both platforms, and 4-5 percent reported having participated in the study previously on the other platform. To minimize the chances of repeat observations, we drop both types of participants from our analysis (463 subjects in the Gender Experiment and 435 in the Race Experiment). We discuss the small effects this has on our results in Appendix F. Summary statistics about respondent characteristics are presented in Appendix D.

4 Results

4.1 Gender

We begin by examining the raw means of the appropriateness ratings.[12] We plot the norm functions for the six legal threshold situations in the Gender Experiment in Figure 1. These functions represent the average social appropriateness ratings given by participants to the behaviors evaluated in our experiment. In accordance with the social norms literature, we assign equally-spaced values of +1, +0.6, +0.2, -0.2, -.6, and -1 to the ordered ratings 'Very socially appropriate,' 'Socially appropriate,' 'Somewhat socially appropriate,' 'Somewhat socially inappropriate,' 'Socially inappropriate,' and

[12] See Appendix D for the full distributions of appropriateness ratings.

'Very socially inappropriate,' respectively.[13] As a result, the norm functions have positive values for actions evaluated as appropriate on average and negative values for inappropriate actions. The black norm functions show appropriateness ratings for the vignettes where we presented male protagonists, and the red functions show those for female protagonists. The panels in the first row of the Figure show the norm functions for risky behavior, the second for prosocial behavior, and the third for labor market behavior. The figure includes a dashed black line in each panel indicating the legal threshold position at which an action crosses from legal (left) to illegal (right).

Overall, we note that the norm functions generally follow an intuitive pattern, with behaviors rated as more appropriate when they are legal as opposed to illegal. Indeed, for behaviors that were studied in Lane et al. (2023) (drink-driving, speeding, alcohol-to-youth), the functions look very similar. In particular, as in their paper, we also find that the law has strong expressive power in the case of the alcohol-to-youth vignette but much weaker power in the drink-driving and speeding vignettes.[14] In our three new vignettes (casino, minimum wage, parental leave), we observe strong downward discontinuities at the legal thresholds in all cases. Most importantly for our research question, we note that in all six vignettes, the norm functions for men and women look very similar and, in particular, do not provide visual evidence for a difference in the discontinuities at the threshold.

[13] We collected a portion of the data (approximately 41 percent, see Appendix D - Table 1) using a four-point scale, with the categories 'Very socially appropriate,' 'Somewhat socially appropriate,' 'Somewhat socially inappropriate' and 'Very socially inappropriate.' Our analysis does not reveal systematic differences in responses between subjects using the four-point scale compared to the six-point scale. We pool the data by assigning the same values to the four-point scale as we do for the six-point scale, i.e., $+1, +.6, -.6, -1$, and verify in robustness checks that assigning $(-)0.33$ to 'Somewhat socially (in)appropriate' instead does not alter our results.

[14] Lane et al. (2023) show that these differences can be related to different perceptions of the intentionality and measurability of behavior across these vignettes (speeding and drink-driving are perceived as less intentional and more difficult to measure accurately).

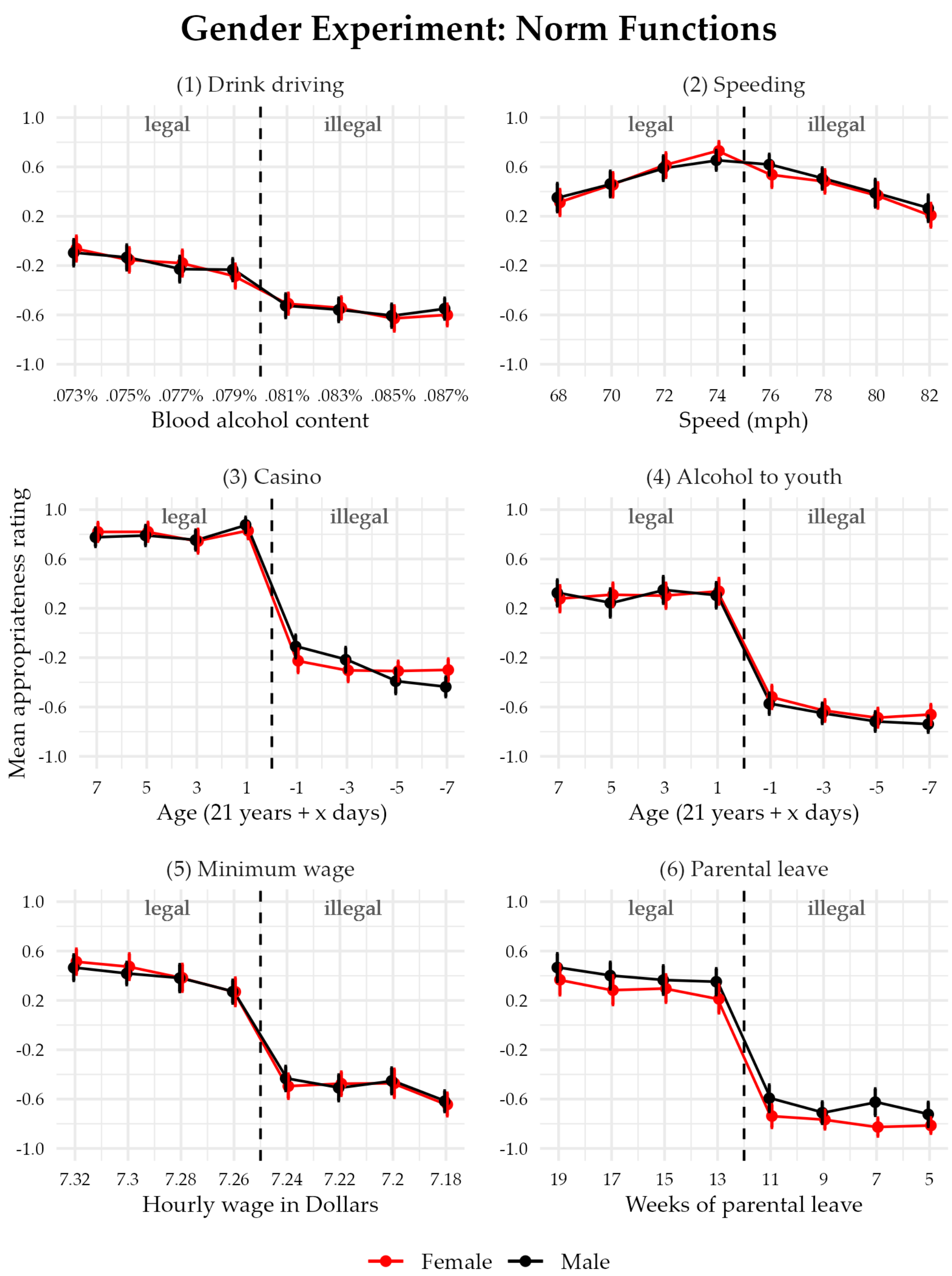


Figure 1: Raw means and 95%-confidence intervals for appropriateness ratings (Gender Experiment)

We formally test for differences in these discontinuities using regression analyses. Our main specification is the following OLS model that assumes linear trends in the distance to the threshold.[15]

$rating_{i,v} = \beta_0 + \beta_1 male_v + \beta_2 illegal_v + \beta_3 male_v \times illegal_v + \beta_4 absdist_v + \beta_5 male_v \times absdist_v + \beta_6 illegal_v \times absdist_v + \beta_7 male_v \times illegal_v \times absdist_v + X_i + \epsilon.$

The dependent variable, $rating_{i,v}$, is the social appropriateness rating subject i assigns to a behavior in vignette v. Variables indexed v are ones that we vary experimentally across vignettes: $male_v$ is an indicator equal to one if the person we describe in vignette v is male and zero otherwise; $illegal_v$ is one if the behavior is illegal and zero otherwise; $absdist_v$ is the absolute distance from the legal threshold (an integer $\in \{1, 3, 5, 7\}$). These three variables and their interactions allow us to capture the full variation across vignettes in a regression discontinuity style, where β_2 measures the discontinuity at the legal threshold for the reference group (women) and β_3, our coefficient of interest, measures the difference in the discontinuity at the threshold for men relative to women.

The vector X_i includes the following subject-level controls: the order in which the subject encountered a vignette during the experiment, the subject's age, and indicators for whether the subject was recruited on CloudResearch (as opposed to Prolific), was presented with a 4-point appropriateness scale (as opposed to the 6-point scale), was female, a US-citizen, born in Texas, had a middle or a high income (as opposed to low), held sexist views. Appendix F shows all regressions without these controls.

The regressions confirm that the law exerts systematic expressive power

[15] In robustness analyses, reported in Appendix F, we have estimated Ordered Logit models and have allowed for flexible trends in the distance to the threshold; the results are very similar.

Table 1: Social appropriateness in the GENDER experiment

	Drink driving	Speeding	Casino	Alcohol to youth	Minimum wage	Parental leave
male==1 (M), β_1	.05	$-.09$	.05	$-.04$	.02	$.13^{*}$
	(.07)	(.07)	(.06)	(.07)	(.07)	(.07)
illegal==1 (I), β_2	$-.16^{**}$	$-.18^{**}$	-1.03^{***}	$-.87^{***}$	$-.67^{***}$	$-.93^{***}$
	(.07)	(.07)	(.06)	(.07)	(.07)	(.07)
M x I, β_3	$-.11$	.15	.13	.00	$-.01$	.00
	(.10)	(.10)	(.09)	(.10)	(.10)	(.11)
absolute distance (AD), β_4	$.03^{***}$	$-.07^{***}$	.00	$-.01$	$.04^{***}$	$.02^{*}$
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
M x AD, β_5	$-.01$	.02	$-.01$	.01	$-.01$	$-.01$
	(.02)	(.02)	(.01)	(.01)	(.02)	(.02)
I x AD, β_6	$-.06^{***}$	.01	$-.01$	$-.02$	$-.07^{***}$	$-.03^{**}$
	(.02)	(.02)	(.01)	(.01)	(.02)	(.02)
M x I x AD, β_7	.03	$-.02$	$-.03^{*}$	$-.01$	.02	.00
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Intercept	$-.27^{***}$	$.63^{***}$	$.90^{***}$	$.31^{***}$	$-.04$	.14
	(.08)	(.09)	(.08)	(.08)	(.09)	(.09)
Adj. R^2	.13	.07	.57	.44	.41	.47
Num. obs.	2053	2053	2053	2053	2053	2053

Note: Table shows OLS regression coefficients, with standard errors reported in parentheses. Stars indicate significance levels at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All regressions include the full set of controls: the order in which the vignette was evaluated by the subject, the subject's age and indicators for whether the subject had been recruited via CloudResearch (as opposed to Prolific), had assessed social appropriateness on a 4-point scale (as opposed to 6-point), was female, a US-citizen, born in Texas, reported a middle or a high income (as opposed to low), and above-median sexist views. These controls are omitted from the table for ease of presentation. Appendix F shows the full regression output and all regressions without additional controls.

on norms in all vignettes and that the manipulation of the gender of the person in the vignette had no impact on it. In all six regressions, our estimates of β_3 are very small and insignificantly different from zero, regardless of whether or not we perform a Benjamini-Hochberg correction to the six p-values. Our estimates of β_2 are always negative and statistically significant, indicating systematic downward discontinuities in the norm function in all vignettes. However, as in Lane et al. (2023), the magnitude of these discontinuities is much weaker in the vignettes related to driving behaviors (and in the case of drink-driving and speeding, the effect is only significant at the 5 percent level). Overall, the results of the Gender Experiment show no evidence that observers condition their judgment of appropriateness on the gender of the person being evaluated (or being affected by the evaluated person's behavior).

4.2 Race

Figure 2 plots the raw means from the Race Experiment.[16] The figure has the same structure as Figure 1, except that black functions now measure the average appropriateness in the vignettes where we used African American protagonists, while the red functions correspond to vignettes with White American protagonists. Three vignettes are also different, as described earlier (we included medical marijuana, age of consent, and gun possession vignettes in place of parental leave, casino, and alcohol-to-youth vignettes).

[16] We report the full distributions of appropriateness ratings in Appendix D.

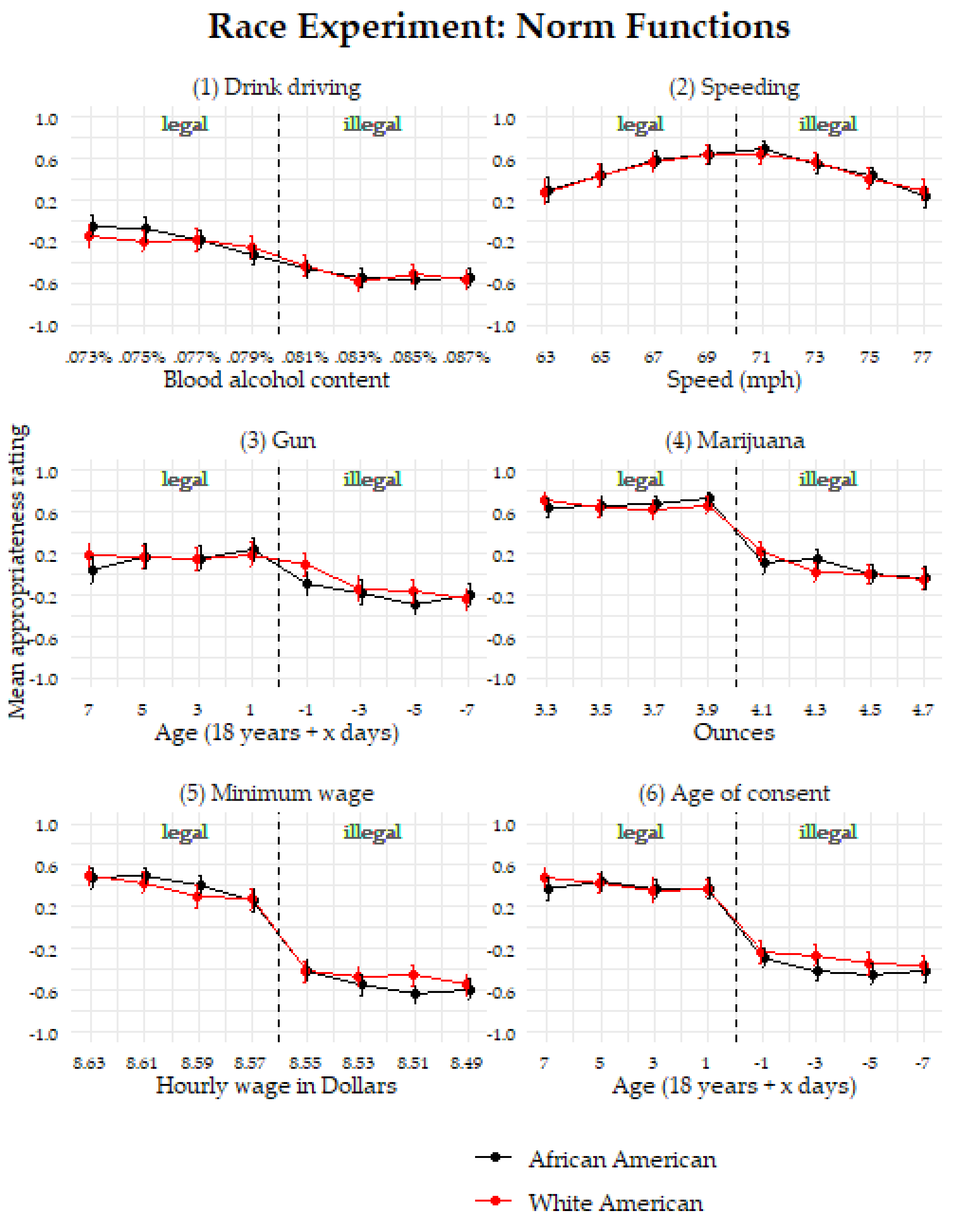


Figure 2: Raw means and 95%-confidence intervals for appropriateness ratings (Race Experiment)

The figure paints a similar picture as Figure 1. We observe strong discontinuities at the legal threshold for the Minimum Wage, Age of Consent, and Marijuana Vignettes. As before, the discontinuities are weaker for the vignettes related to driving behavior. The discontinuity is also somewhat smaller in the Gun Possession Vignette. In this vignette, there also seems to be a difference in the discontinuity of functions relating to African and White Americans. The other five vignettes display virtually no difference between the norm functions of the two racial groups.

To formally estimate the difference in the discontinuity at the threshold, we run OLS models as we did for our Gender Experiment. The general regression model specification is similar, except with dummies for race rather than gender. However, there are some notable differences in the set of control variables. As in the Gender Experiment, we control for the order in which the subject evaluated the vignette, the subject's age, and indicators for whether the subject has been recruited on CloudResearch (as opposed to Prolific), is female, a US citizen, and has a middle or a high income (as opposed to low). Because we recruited subjects residing in Florida for the Race Experiment, we control for whether subjects self-report having been born in Florida and residing there currently. Because our focus is on race and not gender, we do not control for sexist attitudes (nor did we measure them). Instead, we include indicators for whether subjects are themselves non-white and whether they report racist views. We also control for the perceived "Whiteness" of the name of the individual described in the vignette and the perceived socioeconomic status of that name. These variables were obtained in a pilot study, as described in Section E.[17]

In line with the visual evidence observed in Figure 2, the regressions show no difference in the evaluation of appropriateness between African and White

[17] Appendix F presents all regressions without controls included, as well as additional robustness analyses. Our interpretation of results does not change with these additional analyses.

Table 2: Social appropriateness in the RACE experiment

	Drink driving	Speeding	Gun	Marijuana	Minimum wage	Age of consent
white==1 (W), β_1	.29	$-.15$	$-.04$	.01	$-.32$	.06
	(.25)	(.25)	(.20)	(.16)	(.24)	(.25)
illegal==1 (I), β_2	$-.10$	.05	$-.37^{***}$	$-.55^{***}$	$-.69^{***}$	$-.68^{***}$
	(.07)	(.07)	(.08)	(.07)	(.07)	(.07)
W x I, β_3	$-.09$	$-.07$	$.29^{***}$	.14	.11	.14
	(.10)	(.10)	(.11)	(.09)	(.10)	(.10)
absolute distance (AD), β_4	$.05^{***}$	$-.06^{***}$	$-.03^{**}$	$-.01$	$.03^{***}$	.00
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
W x AD, β_5	$-.03^{*}$	$-.00$	.03	.02	.01	.02
	(.02)	(.02)	(.02)	(.01)	(.02)	(.02)
I x AD, β_6	$-.06^{***}$	$-.02$	.01	$-.02$	$-.06^{***}$	$-.03$
	(.02)	(.02)	(.02)	(.01)	(.02)	(.02)
W x I x AD, β_7	.03	.02	$-.06^{**}$	$-.03$	.00	$-.02$
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Intercept	$-.34^{**}$	$.76^{***}$	$.49^{***}$	$.89^{***}$	$-.15$	$.69^{***}$
	(.17)	(.17)	(.18)	(.15)	(.17)	(.18)
Adj. R^2	.12	.06	.08	.28	.42	.32
Num. obs.	2014	2014	2014	2014	2014	2014

Note: Table shows OLS regression coefficients, with standard errors reported in parentheses. Stars indicate significance levels at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All regressions include the full set of controls: the order in which the vignette was evaluated by the subject, the relative 'whiteness' of the name used in the vignette as well as the SES associated with the name (both measured in pilot studies), the subject's age and indicators for whether the subject had been recruited via CloudResearch (as opposed to Prolific), was female, non-white, a US-citizen, born in Florida, reported a middle or a high income (as opposed to low), and above-median racist views. These controls are omitted from the table for ease of presentation. Appendix F shows the full regression output as well as all regressions without controls included.

Americans in all vignettes, except the Gun Possession Vignette. There, our estimate of β_3 is positive and significant (magnitude: .295, $p = 0.010$). The corresponding estimate of β_2 is negative and also significantly different from zero (magnitude: $-.374$, $p < 0.001$). Thus, while the law does seem to have expressive power in the case of African Americans, the discontinuity in the norm function for White Americans does not differ from zero statistically (magnitude: $-.079$, $p = .314$ based on the linear restriction test that $\beta_2 + \beta_3 = 0$). This result should, however, be taken with caution, as this is the only significant effect out of the 12 tests we conducted across our two experiments. Performing a Benjamini-Hochberg correction to the six p-values obtained from the Race Experiment yields $p = .057$ for the Gun Possession Vignette.

5 Discussion and Conclusions

Despite the compelling theoretical arguments for the existence of differences in the expressive power of law across gender and race, our experiments reveal negligible differences between men and women, or African and White Americans. In this concluding section, we discuss several possible explanations for this null result.

The first possibility is that our statistical analysis does not have sufficient power to detect meaningful differences between the groups. We ran simulations to gauge how much this may be a concern (see Appendix G for details). With our sample sizes of approximately 2,000 observations per experiment, we have roughly 80% power to detect differences in appropriateness ratings between groups of about -0.47 in terms of Hedges' g, a popular measure of effect size. For comparison, the magnitude of the discontinuities at the thresholds observed in our Gender and Race experiments are considerably larger than this (averaging -1.12 for Gender and -0.66 for Race). As another benchmark, the original Krupka and Weber (2013) study found that varying the frame (give vs. take) with which actions in a dictator game are described to subjects generated differences in appropriateness ratings of about -0.40 in terms of Hedges' g, which is comparable to the minimum detectable effect size in our study. Thus, although the size of the effects we are powered to detect is not trivial, it does not strike us as prohibitively large to the point of rendering our null results meaningless. Note also that the minimum detectable effect size we calculated is based on assuming the need to adjust p-values for multiple hypothesis testing; in terms of raw p-values, we would be powered to detect effects that are smaller still.

Another possibility is that our treatment manipulations – subtly varying the gender or race of the people described in the vignettes – may not have been sufficiently salient to the subjects. For instance, if subjects did not pay sufficient attention to the names of the vignettes' protagonists in the Race Experiment, then our treatment manipulation would not have had bite, which could explain the null results. To test whether subjects paid attention to the manipulations, we ran follow-up studies in which the experiment proceeded exactly as in our main study, except that immediately after supplying their response to one of the six vignettes, subjects were given an unexpected memory test about the vignette, for which they would receive a bonus payment if they provided the correct answer.[18] The memory test asked subjects to recall the gender (male/female) or race/ethnicity (African/Asian/Hispanic/White American) of the person in the vignette. Overall, participants had a high degree of recall: the percentage of correct answers was 74.4 percent in the Race Experiment and 96 percent in the Gender Experiment.[19]

There are also several possible theoretical explanations for the null results. First, as discussed earlier, our model allows for multiple channels through which the law may exert different expressive power across gender and race. Some of these channels have effects that go in opposite directions and that thus may potentially cancel each other out. For instance, consider the possibility that law enforcement may discriminate against one group.

[18]The follow-up studies were conducted from September to October 2021 in Prolific with Texas-registered subjects for the Gender Experiment and Florida-registered subjects for the Race Experiment. Only subjects who had not already participated in our main experiment were allowed to participate. We recruited 101 subjects for the Gender Experiment and 203 for the Race Experiment. The bonus payment for correctly answering the memory test was £1. Screenshots are provided in Appendix subsections B.3 and B.4.

[19]Note that choosing at random across the different multiple-choice options would result in a correct recall of 25 percent in the Race and 50 percent for the Gender Experiment.

Discrimination may take the form of a higher rate of wrongful convictions and a higher rate of correct convictions. As shown in our theoretical analysis, the former reduces the social incentives against criminal behavior while the latter sharpens them.

Moreover, our theoretical analysis is based on the assumption that observers are aware of the existence of differences between groups. However, empirically, it is not obvious that this is the case. For instance, although academic research suggests that women may, on average, be more prosocial than men, not everyone in the lay population may believe these differences exist. As another example, although discrimination has been shown to be a key factor in driving differences in criminal justice outcomes between African and White Americans, not everyone we sampled in our experiments may believe that discrimination even exists.

Finally, another possible, intriguing explanation for our findings is that people may *deliberately* refrain from using gender and race as the basis of inference when they are asked to cast normative judgments about a person's behavior. Normative and moral judgments may automatically trigger a duty of impartiality, even when this comes at a cost regarding informational efficiency. In other words, while in our theory we assumed that observers would condition their normative inferences on all information available to them – including any group-level characteristics of the person they are evaluating – the fact that these judgments are *normative* may per se induce observers to deliberately disregard such characteristics and cast inferences only on the basis of expected costs, benefits and types as estimated at the aggregate population level (i.e. subscript g would drop out of the model), which may be less efficient but more impartial.

Although merely speculation, we note that the concept of impartial normative judgments aligns with the principle of equality before the law enshrined in the constitutional guarantees of the US and many other countries worldwide. While it is debatable if these countries consistently abide by this principle in their law enforcement (and evidence of discrimination against certain groups suggests otherwise, as noted earlier), our findings may offer a silver lining: even if the law and its enforcers may not always be blind to gender and race differences, its expressive power truly is, as neither race nor gender factors in people's perception of the (in)appropriateness of behaviors regulated by legal rules.

References

Aksoy, C. G., Carpenter, C. S., De Haas, R., and Tran, K. D. (2020). Do laws shape attitudes? evidence from same-sex relationship recognition policies in europe. *European Economic Review*, 124:103399.

Alesina, A. and La Ferrara, E. (2014). A test of racial bias in capital sentencing. *American Economic Review*, 104(11):3397–3433.

Andreoni, J. and Bernheim, D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Anwar, S., Bayer, P., and Hjalmarsson, R. (2012). The impact of jury race in criminal trials. *The Quarterly Journal of Economics*, 127(2):1017–1055.

Becker, G. (1968). Crime and punishment - an economic approach. *Journal of Political Economy*, 76(2):169–217.

Beckett, K. and Evans, H. (2016). Race, death, and justice: Capital sentencing in washington state, 1981-2014. *Colum. J. Race & L.*, 6:77.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–78.

Bénabou, R. and Tirole, J. (2011). Laws and norms. NBER Working Paper N. 17579.

Berdejó, C. (2018). Criminalizing race: Racial disparities in plea-bargaining. *BCL Rev.*, 59:1187.

Bernheim, D. (1994). A theory of conformity. *Journal of Political Economy*, 102(5):841–877.

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review*, 94(4):991–1013.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms.* Cambridge University Press.

Bilén, D., Dreber, A., and Johannesson, M. (2021). Are women more generous than men? a meta-analysis. *Journal of the Economic Science Association*, 7:1–18.

Blau, F. and Kahn, L. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Butcher, K. F., Park, K. H., and Piehl, A. M. (2017). Comparing apples to oranges: Differences in women's and men's incarceration and sentencing outcomes. *Journal of Labor Economics*, 35(S1):S201–S234.

Casoria, F., Galeotti, F., and Villeval, M. C. (2020). Perceived social norm and behavior quickly adjusted to legal changes during the covid-19 pandemic in france. SSRN Working Paper N. 3670895.

Charles, K. K., Guryan, J., and Pan, J. (2022). The effects of sexism on american women: The role of norms vs. discrimination. *Journal of Human Resources*, pages 0920–11209R3.

Coviello, D. and Persico, N. (2015). An economic analysis of black-white disparities in the new york police department's stop-and-frisk program. *The Journal of Legal Studies*, 44(2):315–360.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–474.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Ellingsen, T. and Johannesson, M. (2008). Pride and prejudice: The human side of incentive theory. *American Economic Review*, 98(3):990–1008.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*, 14(4):583–610.

Exley, C., Hauser, O., Moore, M., and Pezzuto, J.-H. (2022). Beliefs about gender differences in social preferences. Discussion papers 2204, university of exeter, department of economics.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, u. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics*, 133(4):1645–1692.

Fallucchi, F. and Nosenzo, D. (2021). The coordinating power of social norms. *Experimental Economics*, forthcoming.

FBI (2021). Arrests by race and ethnicity, 2016. https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/topic-pages/tables/table-21.

Galbiati, R., Henry, E., Jacquemet, N., and Lobeck, M. (2020). How laws affect the perception of norms: Empirical evidence from the lockdown. SSRN Working Paper N. 3684710.

Gould, E. (2021). State of working america wages 2018. Economic policy institute report Paper available at https://www.epi.org/publication/state-of-american-wages-2018/:.

Görges, L. and Nosenzo, D. (2020). Measuring social norms in economics: Why it is important and how it is done. *Analyse & Kritik*, 42(2):285–311.

Knowles, J., Persico, N., and Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of political economy*, 109(1):203–229.

Koper, C. (2014). Crime gun risk factors: buyer, seller, firearm, and transaction characteristics associated with gun trafficking and criminal gun use. *Journal of quantitative criminology*, 30(2):285–315.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Lane, T., Nosenzo, D., and Sonderegger, S. (2023). Law and norms: Empirical evidence. *American Economic Review*, 113(5):1255–93.

Luh, E. (2022). Not so black and white: Uncovering racial bias from systematically misreported trooper reports. *Available at SSRN 3357063*.

McAdams, R. (2015). *The Expressive Powers of Law*. Harvard University Press.

Mitchell, O. and Caudy, M. (2015). Examining racial disparities in drug arrests. *Justice Quarterly*, 32(2):288–313.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., et al. (2020). A large-scale analysis of racial disparities in police stops across the united states. *Nature human behaviour*, 4(7):736–745.

Posner, E. A. (1998). Symbols, signals, and social norms in politics and the law. *The Journal of Legal Studies*, 27(S2):765–797.

Posner, E. A. (2000). Law and social norms: The case of tax compliance. *Virginia Law Review*, 86(8):1781–1819.

Posner, E. A. (2002). *Law and Social Norms.* Harvard University Press.

Rehavi, M. M. and Starr, S. B. (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy*, 122(6):1320–1354.

Shaw, J. and Lee, H. (2019). Race and the criminal justice system response to sexual assault: A systematic review. *American journal of community psychology*, 64(1-2):256–278.

Smith, T. W., Davern, M., Freese, J., and Morgan, S. L. (2019). *General Social Surveys, 1972-2018: Cumulative Codebook.* Chicago: NORC.

Stanford Open Policing Project (2021). Findings – the results of our nationwide analysis of traffic stops and searches. Paper available at:.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Public Economics*, 118:26–40.

Sunstein, C. R. (1996). On the expressive function of law. *University of Pennsylvania Law Review*, 144(5):2021–53.

Tankard, M. and Paluck, E. (2017). The effect of a supreme court decision regarding gay marriage on social norms and personal attitudes. *Psychological Science*, 28(9):1334–1344.

van der Weele, J. (2012). The signaling power of sanctions in social dilemmas. *The Journal of Law, Economics, and Organization*, 28(1):103–26.

Online Appendices

A Theory Appendix

Proof of Proposition 1 This follows directly from Proposition 2.2 in Lane et al. (2023) where the parameter $p > 0$ has been replaced by $p_g - \pi_g > 0$.

Proof of Lemma 1 This follows from Proposition A2 in Lane et al. (2023).

Proof of Lemma 2 This follows from Lemma 1 since

$$\frac{dD_g}{d\pi_g} = -\frac{dD_g}{dp_g}. \tag{5}$$

QED

Proof of Lemma 3 First, note that

$$\frac{dD_g}{dt_g} = \mathcal{M}^{-\prime}(\widehat{\theta}_g^{\overline{o}})\frac{d\widehat{\theta}_g^{\overline{o}}}{dt_g} - \mathcal{M}^{-\prime}(\widetilde{\theta}_g^{\overline{o}})\frac{d\widetilde{\theta}_g^{\overline{o}}}{dt_g}. \tag{6}$$

Second, from the definitions of $\widehat{\theta}_g^o$ and $\widetilde{\theta}_g^o$, we have

$$\frac{d\widehat{\theta}_g^{\overline{o}}}{dt_g} = \frac{1}{\overline{o} + \Delta'(\widehat{\theta}_g^{\overline{o}})} \text{ and } \frac{d\widetilde{\theta}_g^{\overline{o}}}{dt_g} = \frac{1}{\overline{o} + \Delta'(\widetilde{\theta}_g^{\overline{o}})}. \tag{7}$$

where $\overline{o} + \Delta'(.) > 0$ by monotonicity (see Lane et al. (2023), proof of Proposition 2.1). We can now proceed to prove the lemma by means of three examples. For simplicity, we normalize $\overline{o} = 1$. (1) When θ is uniformly distributed, $\mathcal{M}^{-\prime}(.)$ is a constant and $\Delta'(.)$=0. Hence, $\frac{dD_g}{dt_g} = 0$. (2) Suppose now that θ is distributed according to a triangular distribution on $[0,1]$ with mode equal to 1. This implies that $f(\theta) = 2\theta$ and $F(\theta) = \theta^2$, delivering $\mathcal{M}^{+}(\theta) = \frac{2}{3}\frac{1-\theta^3}{1-\theta^2}$ and $\mathcal{M}^{-}(\theta) = \frac{2}{3}\theta$. As a result, $\mathcal{M}^{-\prime}(\theta) = \frac{2}{3}$ and

$\Delta'(\theta) = -\frac{2}{3}\frac{1}{(1+\theta)^2}$ and hence, once we account for $\mathcal{M}^{-\prime}(.)$ and $\Delta'(.)$, we obtain $\frac{dD_g}{dt_g} < 0$. (3) Finally, suppose that θ is distributed on $[0,1]$ according to the following distribution:

$$f(\theta) = \begin{cases} \frac{3}{4} \text{ for } \theta \leq 1/2 \\ \frac{3}{4} - 3(\theta - \frac{1}{2})^2 \ \text{ for } \theta > 1/2 \end{cases}$$

Suppose further that $0 < \widetilde{\theta} < \widehat{\theta} < \frac{1}{2}$.[1] This generates $\mathcal{M}^{+}(\theta) = \frac{1}{16}\frac{24\theta^2-17}{3\theta^2-4}$ and $\mathcal{M}^{-}(\theta) = \frac{\theta}{2}$ so that $\mathcal{M}^{-\prime}(\theta) = \frac{1}{2}$ and $\Delta'(\theta) = -\frac{77}{16(3\theta-4)^2}$. In that case, after substituting for $\mathcal{M}^{-\prime}(.)$ and $\Delta'(.)$, we find that $\frac{dD_g}{dt_g} > 0$. QED

Proof of Lemma 4 We prove the lemma by means of four numerical examples. In all these examples, we set $\overline{o} = t = 1$ and $(p-\pi)K = 0.25$. (i) Consider first a triangular distribution on $[0,1]$ with mode at 1. We have $f(\theta) = 2\theta$ and $F(\theta) = \theta^2$, so that $E(\theta) = \frac{2}{3}$ and $Var(\theta) = \frac{1}{18}$. Furthermore, $\mathcal{M}^{+}(\theta) = \frac{2}{3}\frac{1-\theta^3}{1-\theta^2}$ and $\mathcal{M}^{-}(\theta) = \frac{2}{3}\theta$, implying that $\Delta(\theta) = \frac{2}{3(\theta+1)}$. It is straightforward to compute $\widehat{\theta} = 0.58$ and $\widetilde{\theta} = 0.19$. Substituting for these, we obtain $D = 0.26$, henceforth denoted as D_1. (ii) Consider now a triangular distribution on $[0,1]$ with mode at 0. We have $f(\theta) = 2(1-\theta)$ and $F(\theta) = (1-\theta)^2$, so that $E(\theta) = \frac{1}{3}$ and $Var(\theta) = \frac{1}{18}$. Furthermore, $\mathcal{M}^{+}(\theta) = \frac{2\theta+1}{3}$ and $\mathcal{M}^{-}(\theta) = \frac{1}{3}\theta\frac{2\theta-3}{\theta-2}$, implying that $\Delta(\theta) = \frac{2}{3(2-\theta)}$. It is straightforward to compute $\widehat{\theta} = 0.54$ and $\widetilde{\theta} = 0.35$. Substituting for these, we obtain $D = 7.6 \times 10^{-2}$, henceforth denoted as D_2. (iii) Consider now a triangular distribution on $[0,1]$ with mode at 0.5. We have $f(\theta) = 4\theta$ for $\theta \leq 0.5$ and $f(\theta) = 4(1-\theta)$ for $\theta > 0.5$. This implies that $F(\theta) = 2\theta^2$ for $\theta \leq 0.5$ and $F(\theta) = 1 - 2(1-\theta)^2$ for $\theta > 0.5$. We have $E(\theta) = \frac{1}{2}$ and $Var(\theta) = \frac{1}{24}$. For $\theta \leq 0.5$, $\mathcal{M}^{+}(\theta) = \frac{4}{3}\frac{\theta^3-1}{2\theta^2-1}$ and $\mathcal{M}^{-}(\theta) = \frac{2}{3}\theta$, so that $\Delta(\theta) = \frac{2}{3}\frac{\theta-2}{2\theta^2-1}$. For

[1] This happens for appropriate parameter values. An example is $t = 0.45$ and $(p-\pi)K = 0.1$.

$\theta > 0.5$, $\mathcal{M}^{+}(\theta) = \frac{1}{3}(2\theta + 1)$ and $\mathcal{M}^{-}(\theta) = \frac{2}{3}\theta^2 \frac{2\theta-3}{1-2(1-x)^2}$, so that $\Delta(\theta) = \frac{1}{3}\frac{1-2\theta}{1-2(1-x)^2}$. It is straightforward to compute $\widehat{\theta} = 0.79$ and $\widetilde{\theta} = 0.63$. Substituting for these, we obtain $D = 1.4 \times 10^{-2}$, henceforth denoted as D_3.
(iv) Finally, consider a uniform distribution on $[0, 1]$ so that $E(\theta) = \frac{1}{2}$ and $Var(\theta) = \frac{1}{12}$. In this case $\mathcal{M}^{+}(\theta) = \frac{1+\theta}{2}$ and $\mathcal{M}^{-}(\theta) = \frac{\theta}{2}$, so that $\Delta(\theta) = \frac{1}{2}$. It is straightforward to compute $\widehat{\theta} = 0.5$ and $\widetilde{\theta} = 0.25$. Substituting for these, we obtain $D = 0.125$, henceforth denoted as D_4.

Part (i) of the lemma is proved by comparing D_3 and D_4. Consider now part (ii). Comparing D_1 and D_2 proves that we may have $D_g > D_{g'}$. Comparing D_2 and D_3 proves that we may have $D_g < D_{g'}$. QED

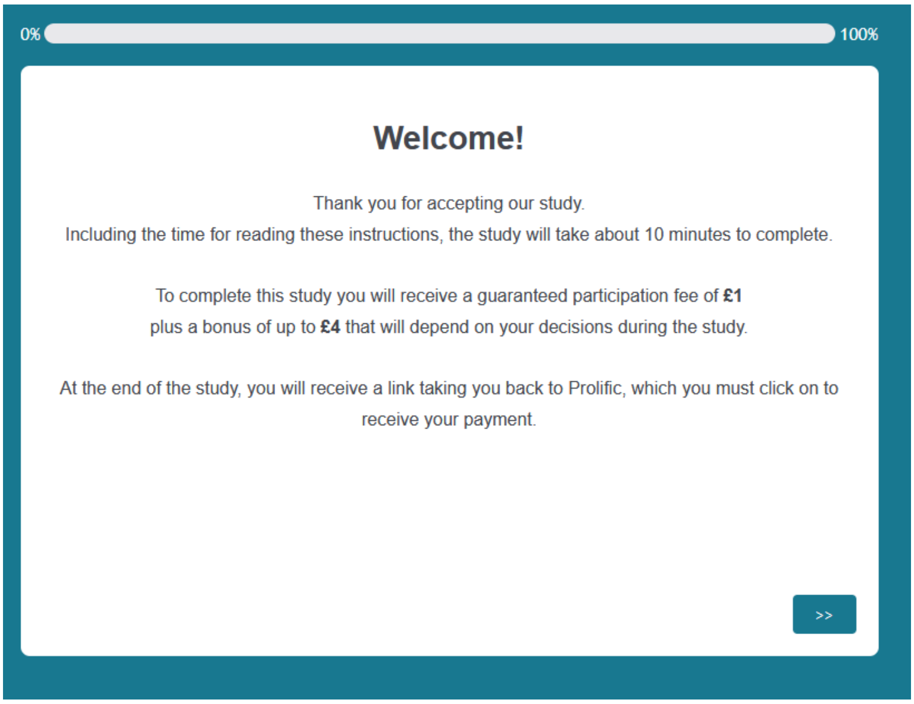
B Experiment Instructions

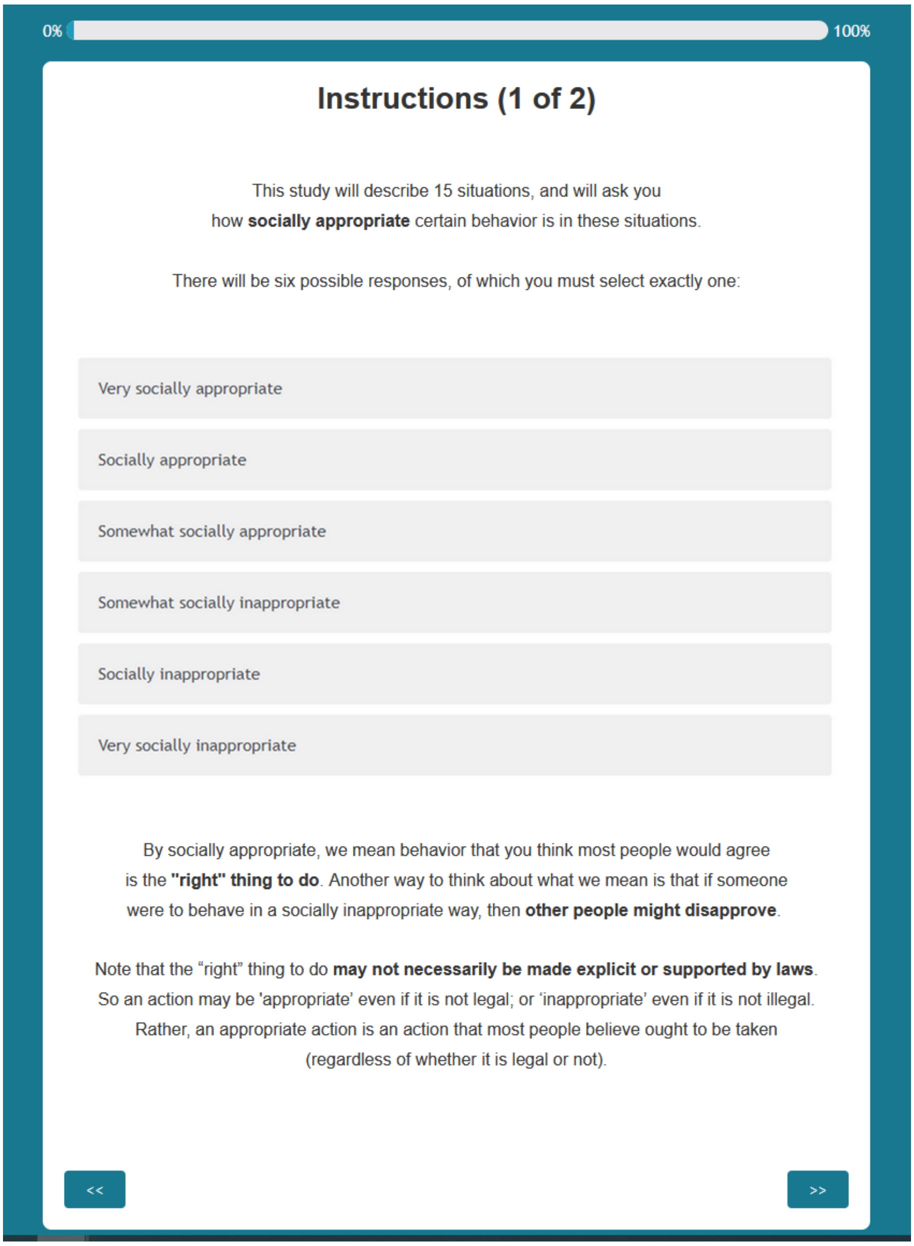
In this section, we present screenshots from the Gender Experiment (B.1) and the Race Experiment (B.2). The vignettes are presented in random order, just as they were to subjects in the experiment. The displayed versions of the manipulated vignettes are randomly selected – see Appendix C for full details of how these vignettes differed across conditions. For the main experiment, the screens shown are those displayed to subjects who accessed the study via Prolific.[2]

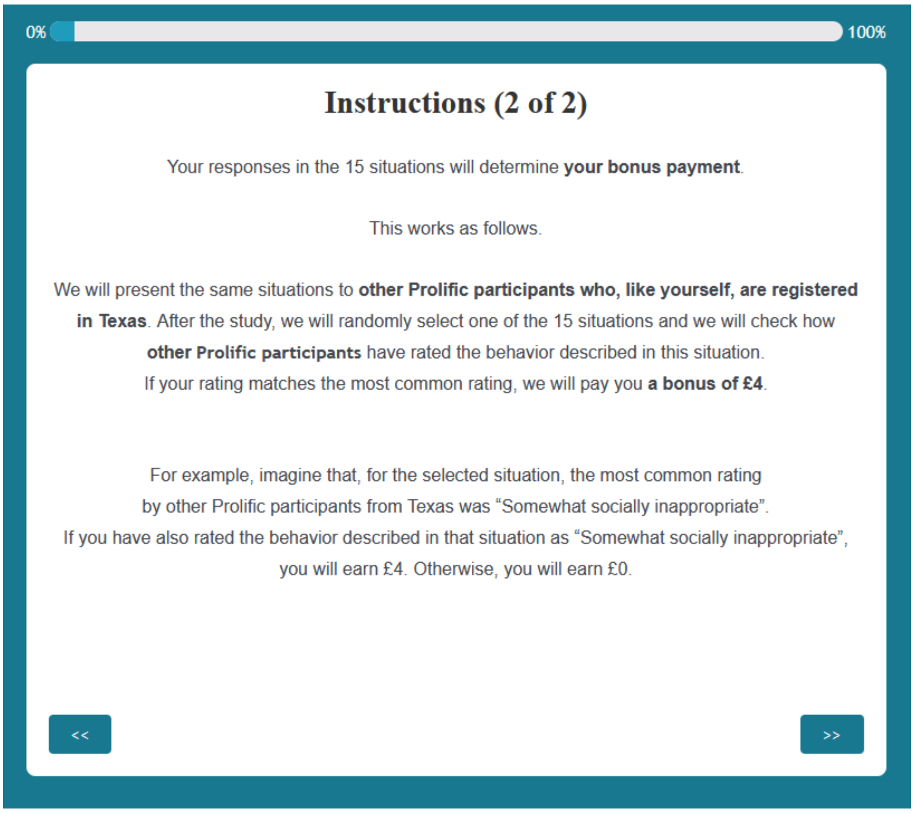
Subsections B.3 and B.4 present the additional screens we included in the follow-up experiments designed to check subjects' attention to the gender/race of the person described in a vignette. Those were inserted at a random position during the main experiment. Finally, subsections B.5 and B.6 show, respectively, the screenshots from the two pilot studies we ran to determine the relative whiteness of different names and the names' associated socioeconomic status. In the first pilot, subjects were randomly assigned to evaluate the likelihood that names belonged to an African American or a White American; shown below is the African American version. In the second pilot, subjects were randomly assigned to evaluate the socioeconomic status of the names of either White Americans or African Americans; shown below is the White name version.

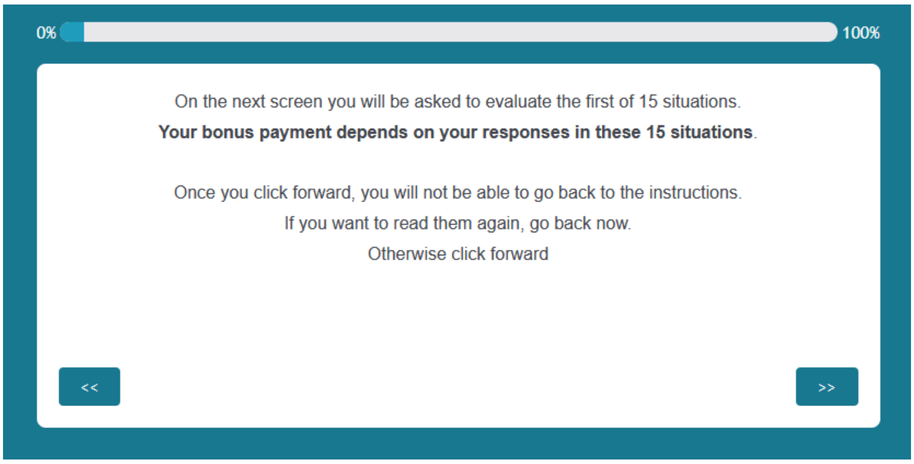
[2]The CloudResearch versions were different in that they referred to other subjects also being MTurkers, the payments were specified in US dollars rather than British Pounds, and there were different explanations regarding the logistics of subjects receiving their earnings. In the CloudResearch versions, we asked at the end of the study whether subjects had Prolific accounts rather than vice versa (we did not include this question for the first wave of CloudResearch data collection in the Gender Experiment). The first CloudResearch wave of data collection in the Gender Experiment also differed in that the social appropriateness scale contained four rather than six points, with the options "Socially appropriate" and "Socially inappropriate" excluded.

B.1 Gender Experiment




0%
100%
Instructions (1 of 2)
This study will describe 15 situations, and will ask you how **socially appropriate** certain behavior is in these situations.
There will be six possible responses, of which you must select exactly one:
Very socially appropriate
Socially appropriate
Somewhat socially appropriate
Somewhat socially inappropriate
Socially inappropriate
Very socially inappropriate
By socially appropriate, we mean behavior that you think most people would agree is the **"right" thing to do**. Another way to think about what we mean is that if someone were to behave in a socially inappropriate way, then **other people might disapprove**.
Note that the "right" thing to do **may not necessarily be made explicit or supported by laws**. So an action may be 'appropriate' even if it is not legal; or 'inappropriate' even if it is not illegal. Rather, an appropriate action is an action that most people believe ought to be taken (regardless of whether it is legal or not).
<<
>>


0%
100%
Instructions (2 of 2)
Your responses in the 15 situations will determine **your bonus payment**.
This works as follows.
We will present the same situations to **other Prolific participants who, like yourself, are registered in Texas**. After the study, we will randomly select one of the 15 situations and we will check how **other Prolific participants** have rated the behavior described in this situation.
If your rating matches the most common rating, we will pay you **a bonus of £4**.
For example, imagine that, for the selected situation, the most common rating by other Prolific participants from Texas was “Somewhat socially inappropriate”.
If you have also rated the behavior described in that situation as “Somewhat socially inappropriate”, you will earn £4. Otherwise, you will earn £0.
<<
>>


0%
100%
On the next screen you will be asked to evaluate the first of 15 situations.
Your bonus payment depends on your responses in these 15 situations.
Once you click forward, you will not be able to go back to the instructions.
If you want to read them again, go back now.
Otherwise click forward
<<
>>

0% 100%

A woman has saved up $2,000 which she intends to spend on a luxury beach vacation. Just before she books the vacation, she reads a news report about a charity providing aid for hungry people in an impoverished African country. The woman decides she should donate the $2,000 to the charity instead of going on vacation. However, she then changes her mind and books the beach vacation, and does not donate any money to charity.

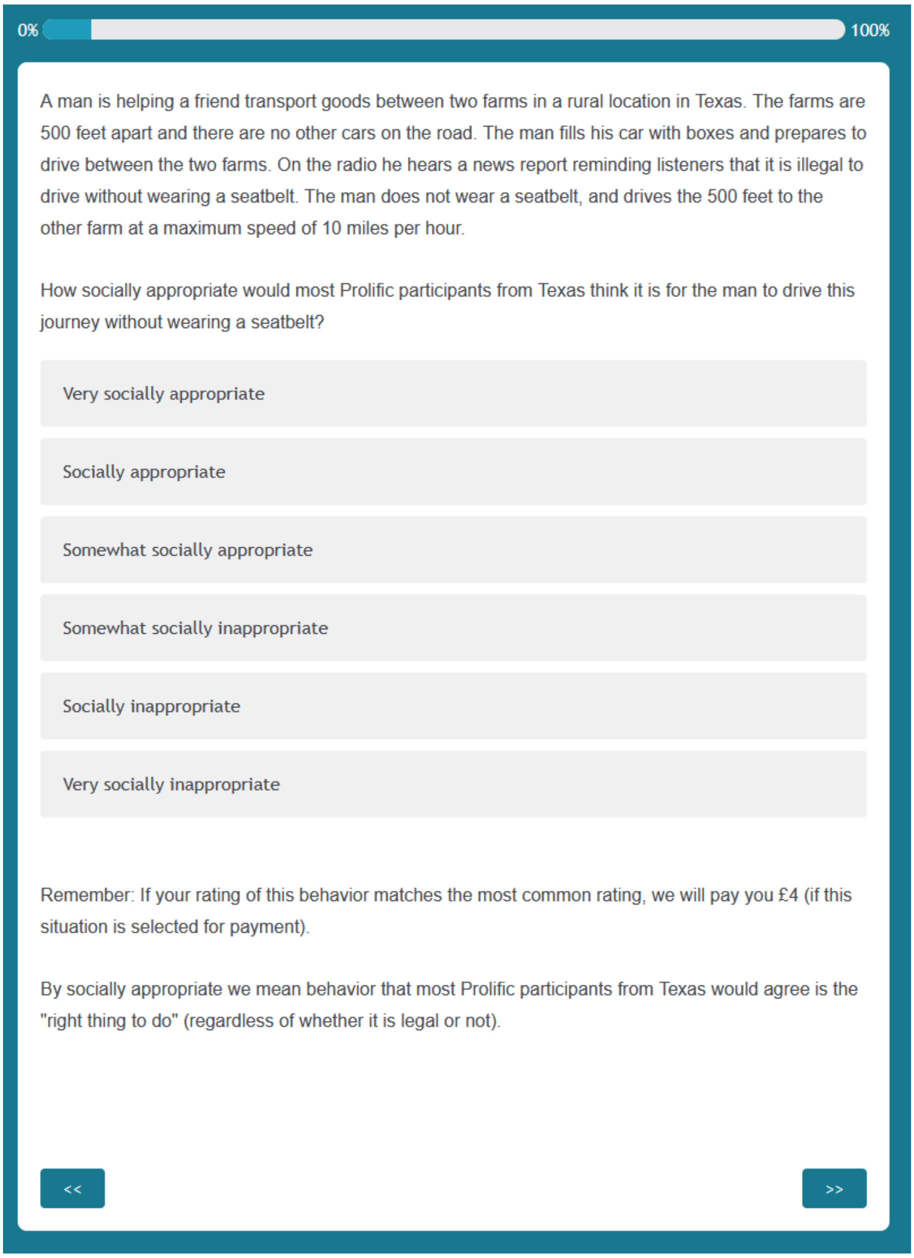
How socially appropriate would most Prolific participants from Texas think it is for the woman to book the beach vacation and not donate any money to charity?

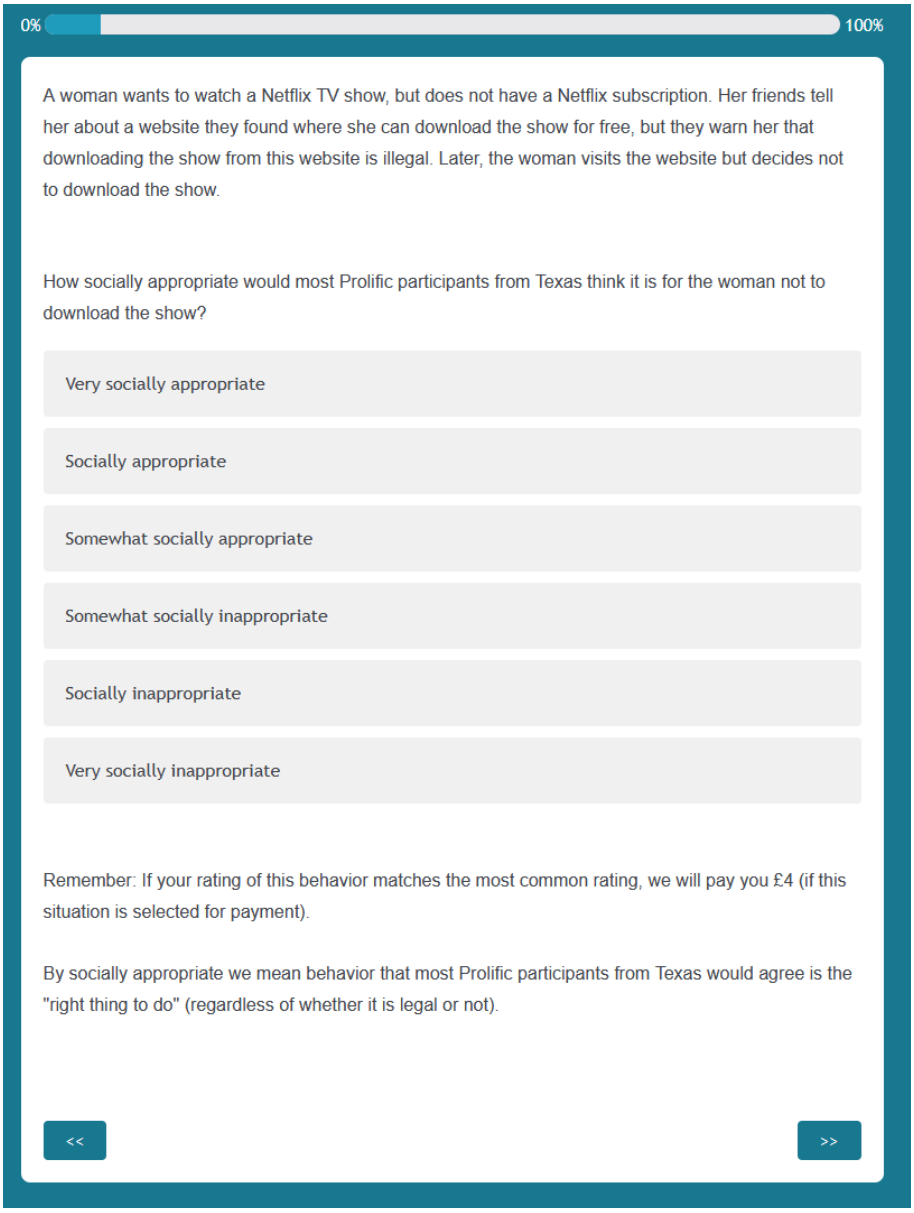
- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

>>


0%
100%
A man is helping a friend transport goods between two farms in a rural location in Texas. The farms are 500 feet apart and there are no other cars on the road. The man fills his car with boxes and prepares to drive between the two farms. On the radio he hears a news report reminding listeners that it is illegal to drive without wearing a seatbelt. The man does not wear a seatbelt, and drives the 500 feet to the other farm at a maximum speed of 10 miles per hour.
How socially appropriate would most Prolific participants from Texas think it is for the man to drive this journey without wearing a seatbelt?
Very socially appropriate
Socially appropriate
Somewhat socially appropriate
Somewhat socially inappropriate
Socially inappropriate
Very socially inappropriate
Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).
By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).
<<
>>


0%

100%

A woman wants to watch a Netflix TV show, but does not have a Netflix subscription. Her friends tell her about a website they found where she can download the show for free, but they warn her that downloading the show from this website is illegal. Later, the woman visits the website but decides not to download the show.

How socially appropriate would most Prolific participants from Texas think it is for the woman not to download the show?

Very socially appropriate

Socially appropriate

Somewhat socially appropriate

Somewhat socially inappropriate

Socially inappropriate

Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<<

>>

0% 100%

A man owns a store in a small town in Texas. One day, a young customer enters the store with the intention of buying some beer. The customer sees a sign in the store reminding customers that in the United States it is illegal for store owners to sell alcohol to people under the age of 21. The store owner is the father of a classmate of the customer and knows that the customer turned 21 7 days ago. He also knows that the customer often gets drunk and vandalizes property in the neighborhood. The customer brings a 24-pack of alcoholic beer up to the counter. The man looks at the customer who appears sober. He then sells the beer to the customer.

How socially appropriate would most Prolific participants from Texas think it is for the man to sell the beer to the customer?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

>>

0% 100%

A man works at a large warehouse in Texas. The man expects his first child and requests a period of unpaid leave from his manager. The manager consults with the company's legal department about the man's legal rights. The legal department tells the manager that the federal Family and Medical Leave Act requires the company to offer the man at least 12 weeks of unpaid, job-protected leave. The manager then phones the man, and offers him 15 weeks of unpaid leave, after which he must immediately return to work – if he refuses to do so, he will be dismissed.

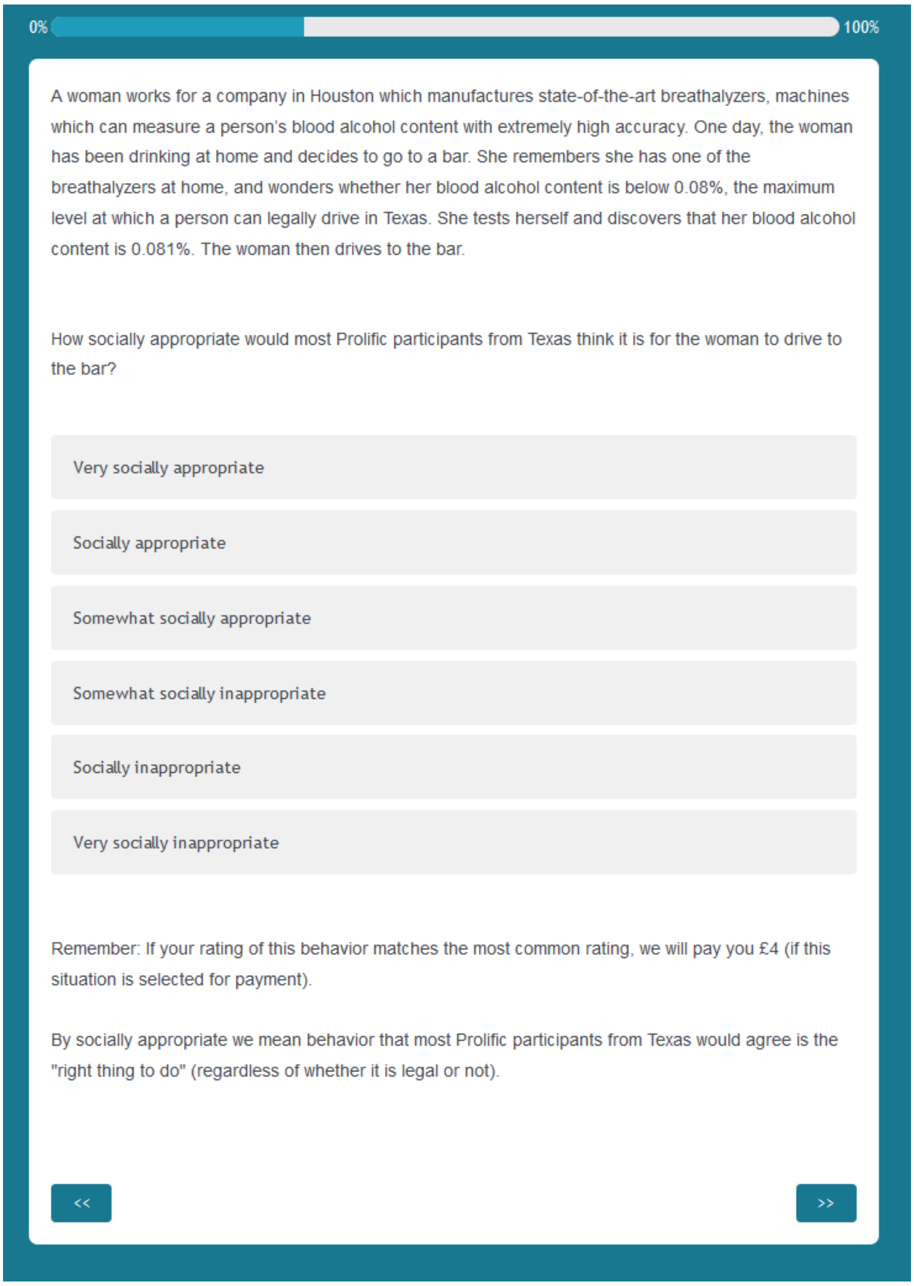
How socially appropriate would most Prolific participants from Texas think it is for the manager to make this demand to the man?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>


0% 100%

A woman works for a company in Houston which manufactures state-of-the-art breathalyzers, machines which can measure a person's blood alcohol content with extremely high accuracy. One day, the woman has been drinking at home and decides to go to a bar. She remembers she has one of the breathalyzers at home, and wonders whether her blood alcohol content is below 0.08%, the maximum level at which a person can legally drive in Texas. She tests herself and discovers that her blood alcohol content is 0.081%. The woman then drives to the bar.

How socially appropriate would most Prolific participants from Texas think it is for the woman to drive to the bar?

Very socially appropriate

Socially appropriate

Somewhat socially appropriate

Somewhat socially inappropriate

Socially inappropriate

Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

A woman is on her way to meet a friend when she sees an injured man lying on the sidewalk. The woman stops to ask the man if he is OK; the man tells her he has been attacked and had his wallet and cell phone stolen. The woman phones the police and waits with the man until they arrive.

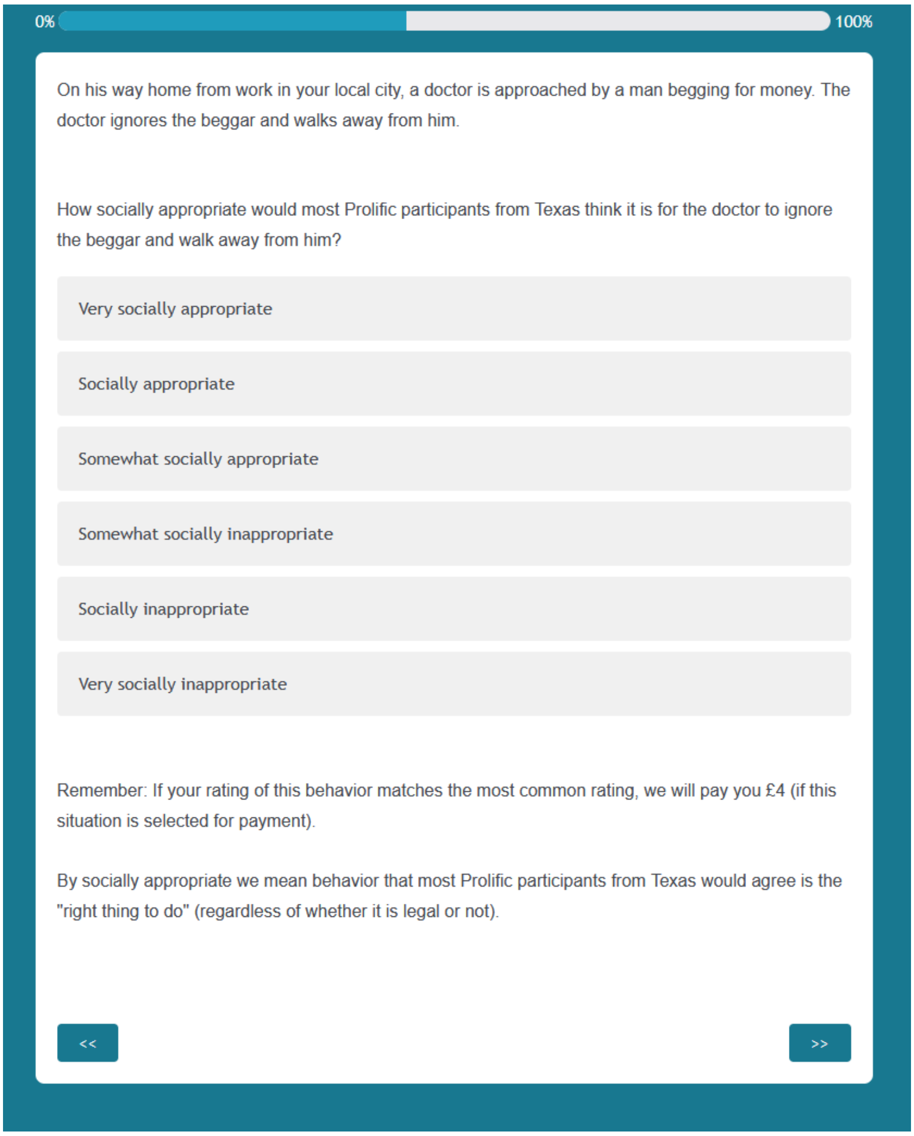
How socially appropriate would most Prolific participants from Texas think it is for the woman to help the man in this way?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>


0% 100%

On his way home from work in your local city, a doctor is approached by a man begging for money. The doctor ignores the beggar and walks away from him.

How socially appropriate would most Prolific participants from Texas think it is for the doctor to ignore the beggar and walk away from him?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

A homeowner wakes in the middle of the night and finds a thief attempting to steal his television. He catches the thief and beats him severely, breaking the thief's arm before throwing him out through the front door.

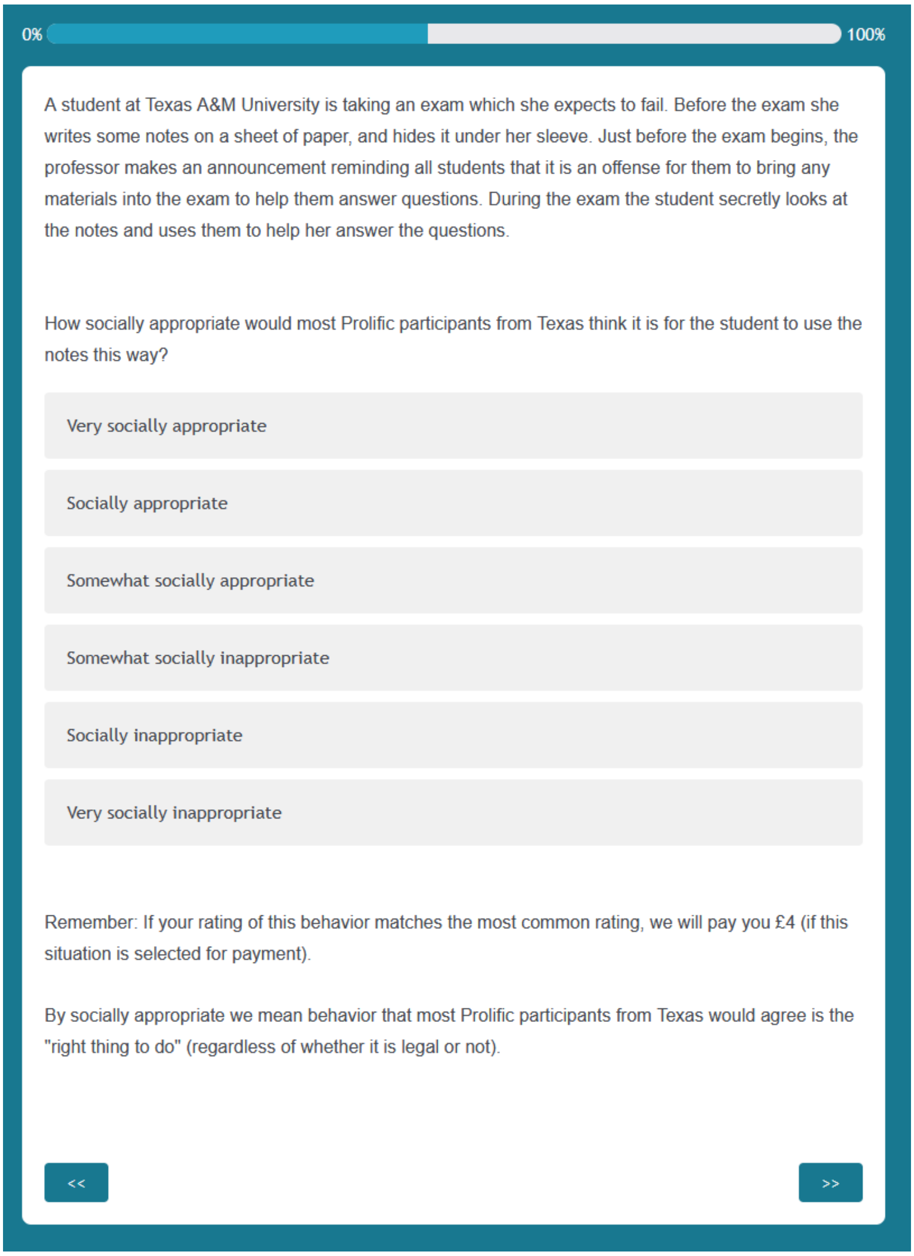
How socially appropriate would most Prolific participants from Texas think it is for the homeowner to treat the thief this way?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>


0% 100%

A student at Texas A&M University is taking an exam which she expects to fail. Before the exam she writes some notes on a sheet of paper, and hides it under her sleeve. Just before the exam begins, the professor makes an announcement reminding all students that it is an offense for them to bring any materials into the exam to help them answer questions. During the exam the student secretly looks at the notes and uses them to help her answer the questions.

How socially appropriate would most Prolific participants from Texas think it is for the student to use the notes this way?

Very socially appropriate

Socially appropriate

Somewhat socially appropriate

Somewhat socially inappropriate

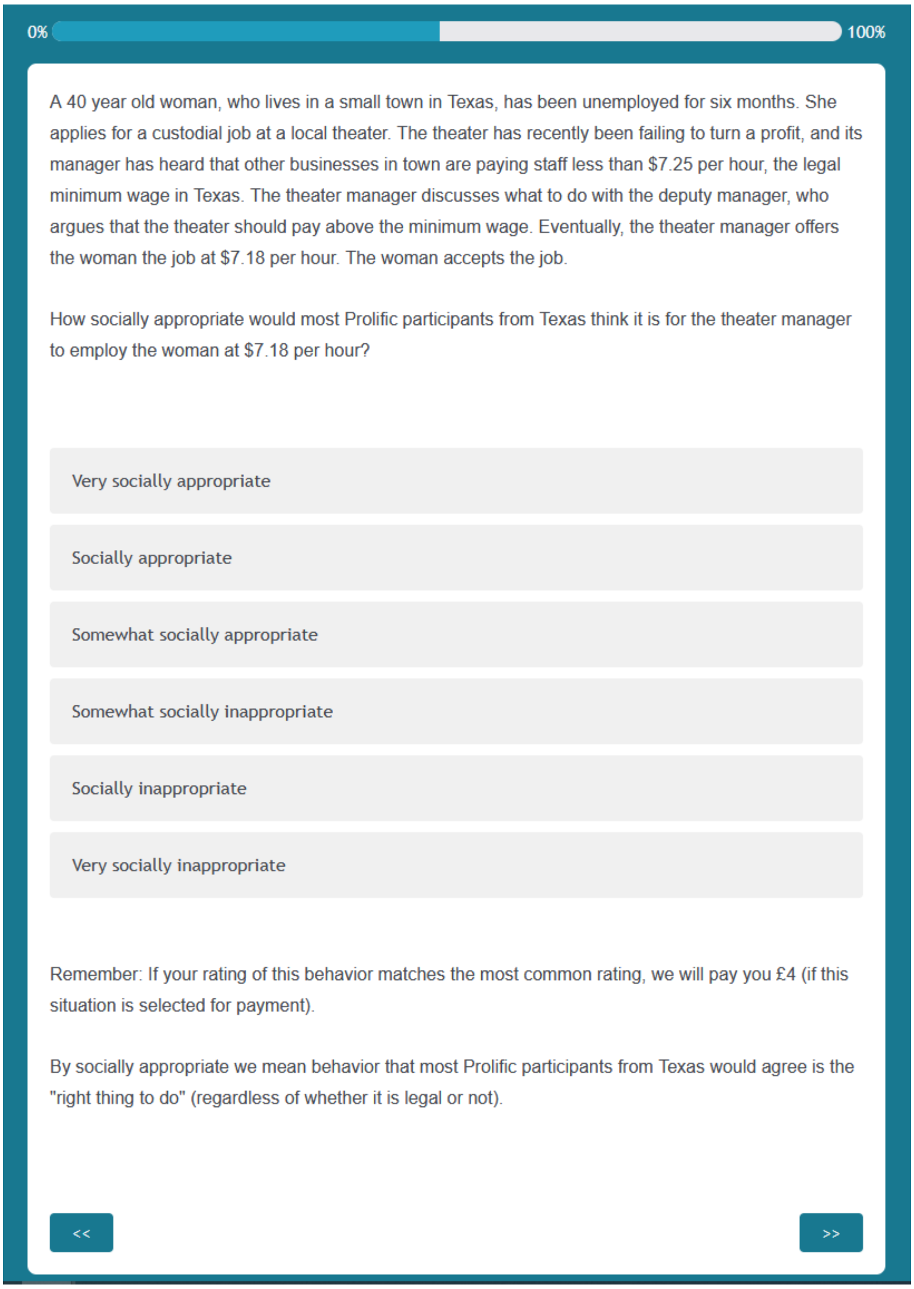
Socially inappropriate

Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>


0% 100%

A 40 year old woman, who lives in a small town in Texas, has been unemployed for six months. She applies for a custodial job at a local theater. The theater has recently been failing to turn a profit, and its manager has heard that other businesses in town are paying staff less than $7.25 per hour, the legal minimum wage in Texas. The theater manager discusses what to do with the deputy manager, who argues that the theater should pay above the minimum wage. Eventually, the theater manager offers the woman the job at $7.18 per hour. The woman accepts the job.

How socially appropriate would most Prolific participants from Texas think it is for the theater manager to employ the woman at $7.18 per hour?

Very socially appropriate

Socially appropriate

Somewhat socially appropriate

Somewhat socially inappropriate

Socially inappropriate

Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

A woman is having lunch at a coffee shop in a small town in Texas. When she gets up to leave, she notices a wallet unattended on the floor. The woman checks to see nobody is looking, and then picks the wallet up and walks out of the coffee shop with it.

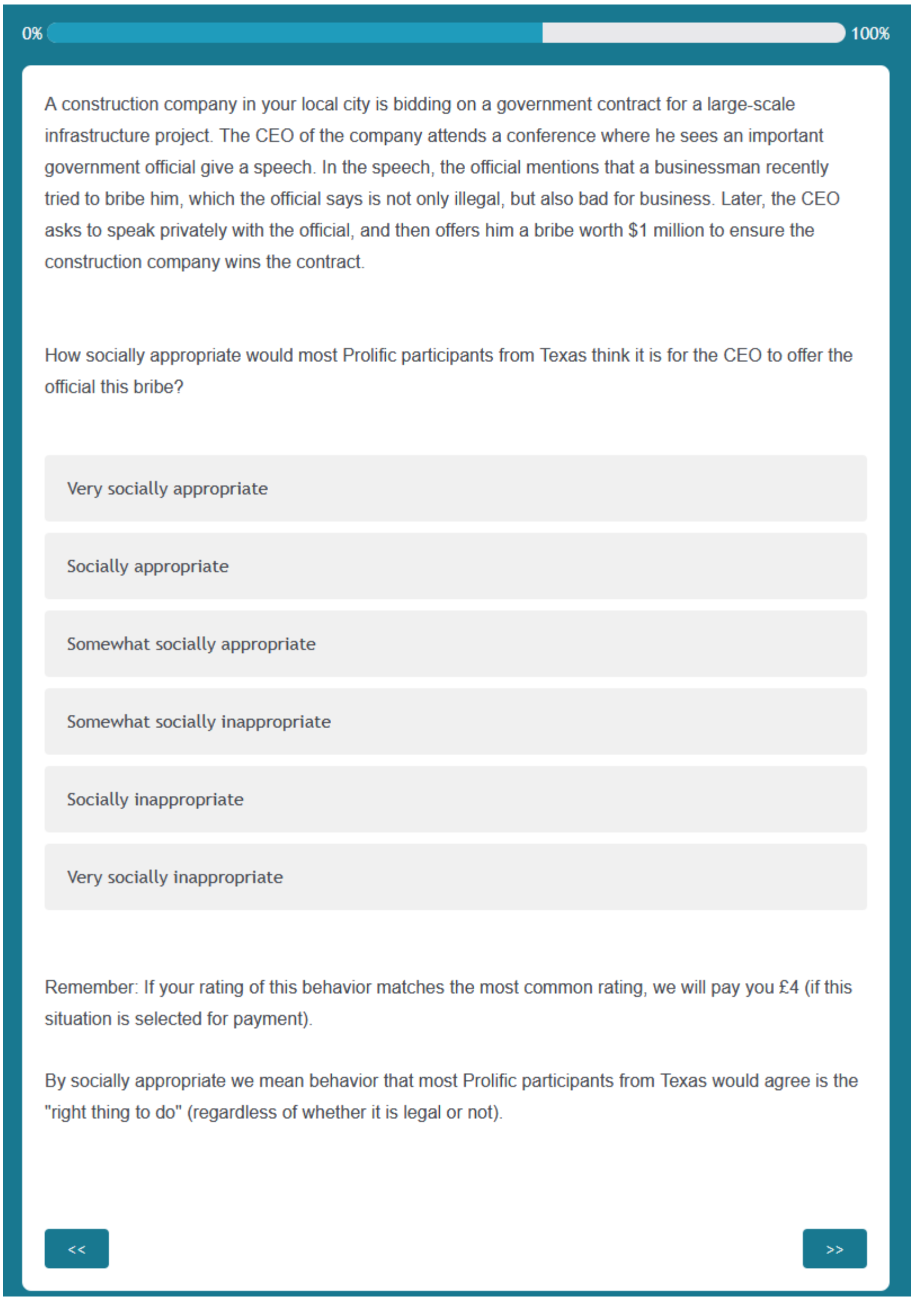
How socially appropriate would most Prolific participants from Texas think it is for the woman to pick the wallet up and walk out of the coffee shop with it?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>


0% 100%

A construction company in your local city is bidding on a government contract for a large-scale infrastructure project. The CEO of the company attends a conference where he sees an important government official give a speech. In the speech, the official mentions that a businessman recently tried to bribe him, which the official says is not only illegal, but also bad for business. Later, the CEO asks to speak privately with the official, and then offers him a bribe worth $1 million to ensure the construction company wins the contract.

How socially appropriate would most Prolific participants from Texas think it is for the CEO to offer the official this bribe?

Very socially appropriate

Socially appropriate

Somewhat socially appropriate

Somewhat socially inappropriate

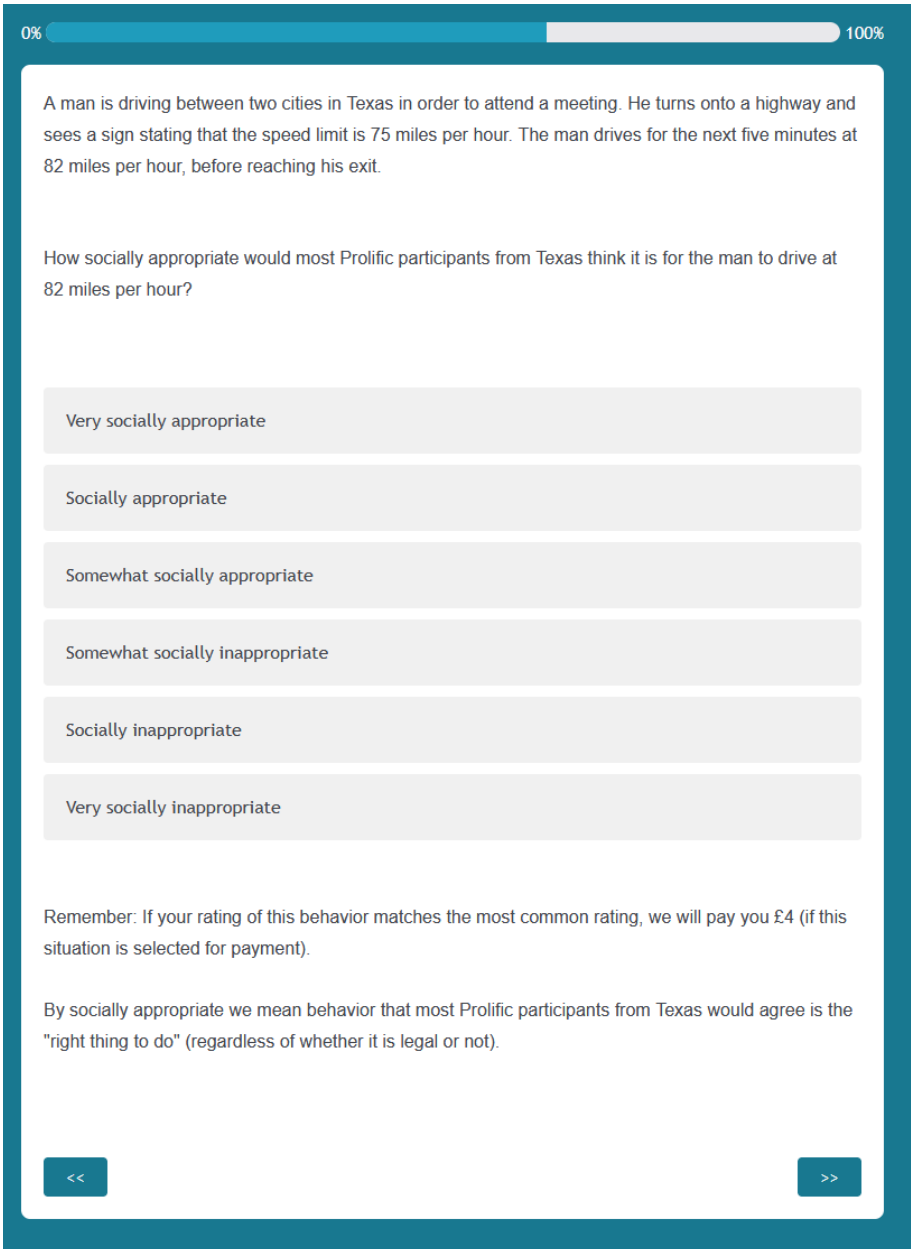
Socially inappropriate

Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>


0%
100%
A man is driving between two cities in Texas in order to attend a meeting. He turns onto a highway and sees a sign stating that the speed limit is 75 miles per hour. The man drives for the next five minutes at 82 miles per hour, before reaching his exit.
How socially appropriate would most Prolific participants from Texas think it is for the man to drive at 82 miles per hour?
Very socially appropriate
Socially appropriate
Somewhat socially appropriate
Somewhat socially inappropriate
Socially inappropriate
Very socially inappropriate
Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).
By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).
<<
>>

0% 100%

A man works for a casino in Texas. His job is to check customers' ID and prevent those below the legal gambling age of 21 from entering the casino. One evening, he sees a young customer in the line and asks for ID. The ID shows the customer will turn 21 in 5 days. The man lets the customer enter the casino.

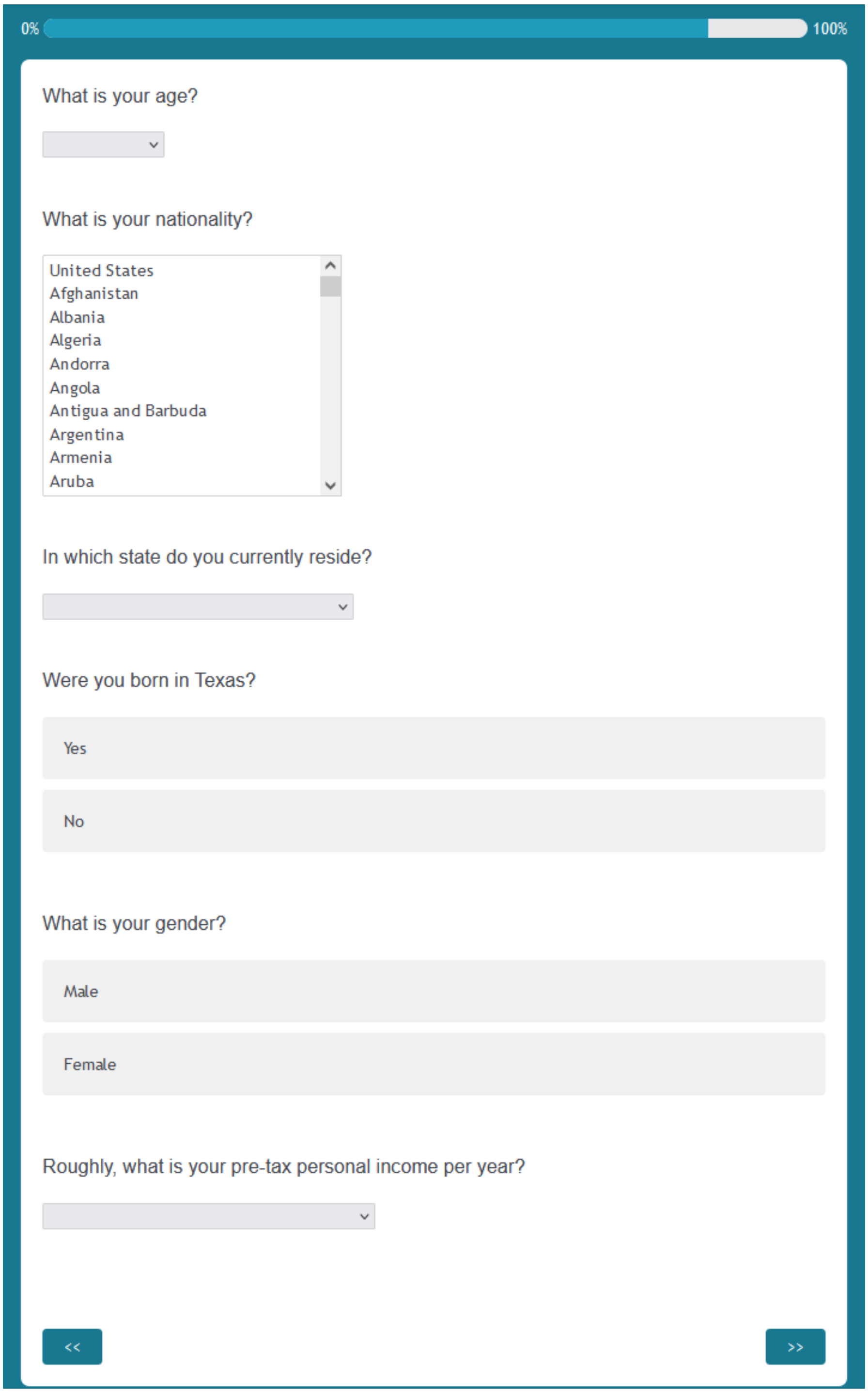
How socially appropriate would most Prolific participants from Texas think it is for the man to let this customer enter the casino?

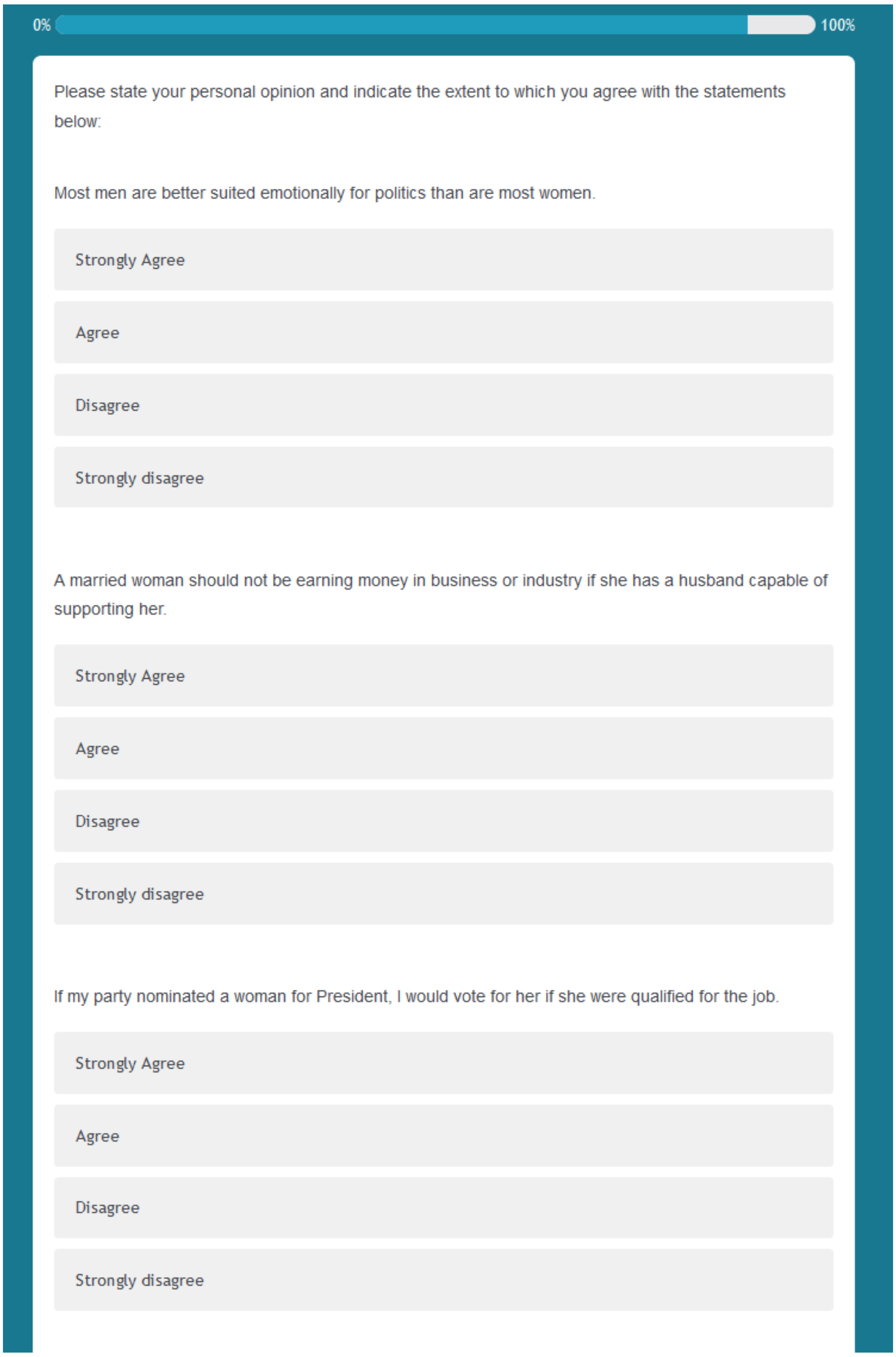
- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Texas would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>


0%
100%
What is your age?
What is your nationality?
United States
Afghanistan
Albania
Algeria
Andorra
Angola
Antigua and Barbuda
Argentina
Armenia
Aruba
In which state do you currently reside?
Were you born in Texas?
Yes
No
What is your gender?
Male
Female
Roughly, what is your pre-tax personal income per year?
<<
>>


0% 100%

Please state your personal opinion and indicate the extent to which you agree with the statements below:

Most men are better suited emotionally for politics than are most women.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

A married woman should not be earning money in business or industry if she has a husband capable of supporting her.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

If my party nominated a woman for President, I would vote for her if she were qualified for the job.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

Women should take care of running their homes and leave running the country up to men.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

It is more important for a wife to help her husband's career than to have one herself.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

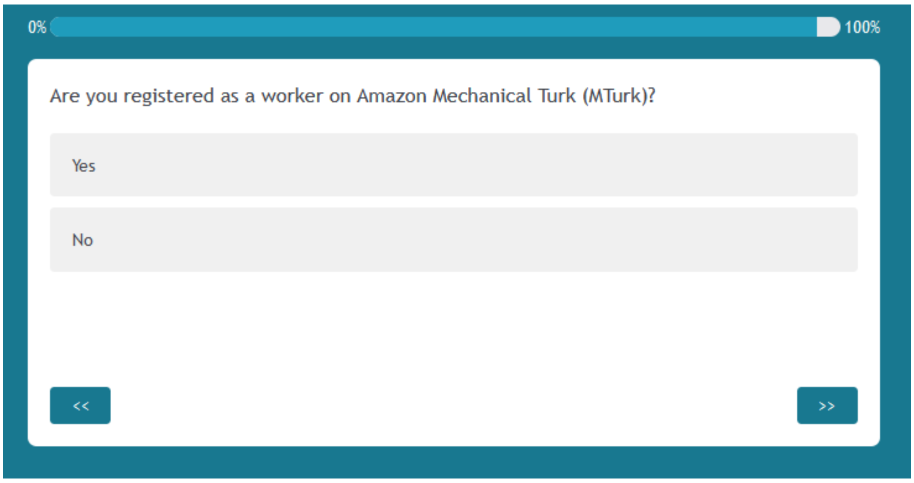
A preschool child is likely to suffer if his or her mother works.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

<< >>



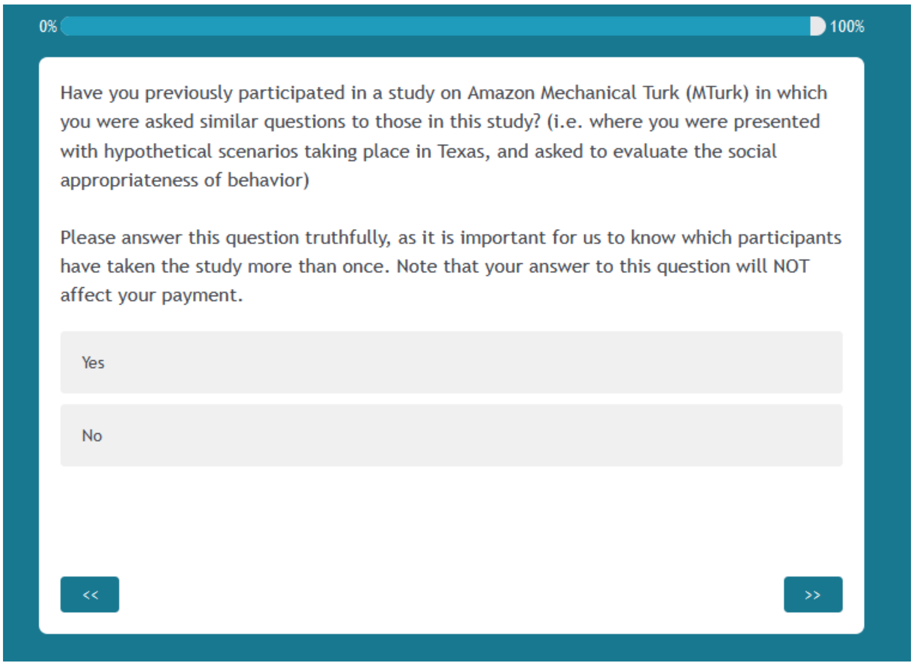
0% 100%

Are you registered as a worker on Amazon Mechanical Turk (MTurk)?

Yes

No

<< >>



0% 100%

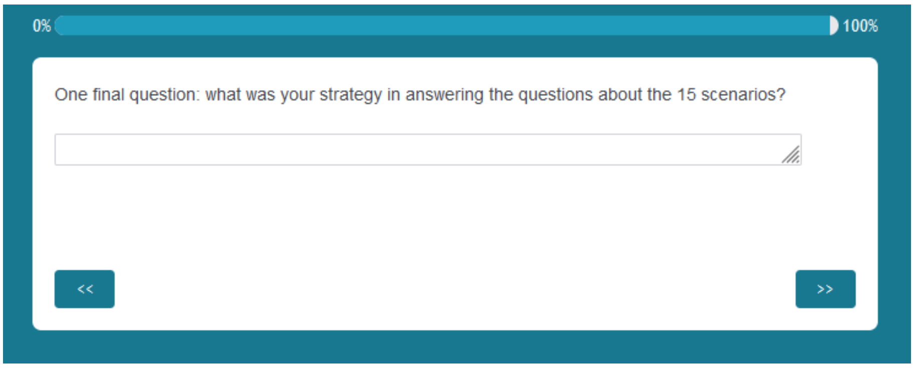
Have you previously participated in a study on Amazon Mechanical Turk (MTurk) in which you were asked similar questions to those in this study? (i.e. where you were presented with hypothetical scenarios taking place in Texas, and asked to evaluate the social appropriateness of behavior)

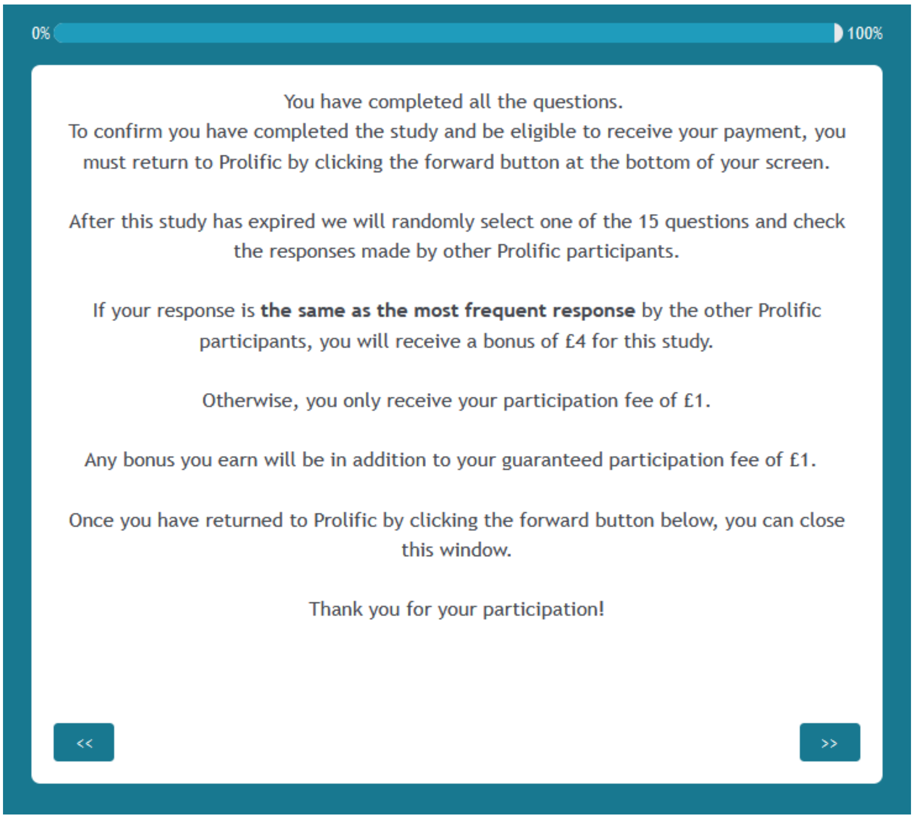
Please answer this question truthfully, as it is important for us to know which participants have taken the study more than once. Note that your answer to this question will NOT affect your payment.

Yes

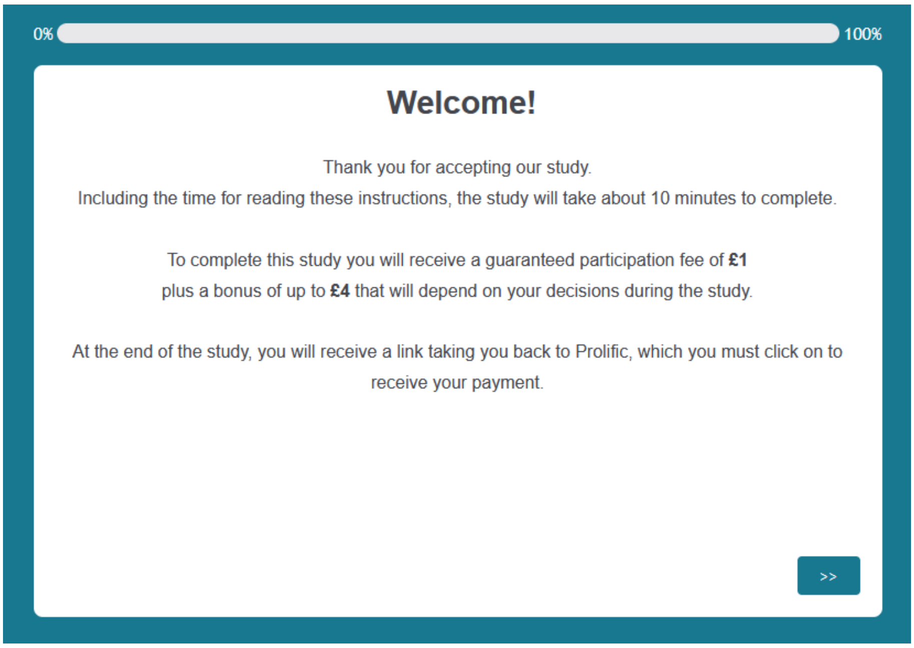
No

<< >>


0%
100%
One final question: what was your strategy in answering the questions about the 15 scenarios?
<<
>>


0%
100%
You have completed all the questions.
To confirm you have completed the study and be eligible to receive your payment, you must return to Prolific by clicking the forward button at the bottom of your screen.
After this study has expired we will randomly select one of the 15 questions and check the responses made by other Prolific participants.
If your response is the same as the most frequent response by the other Prolific participants, you will receive a bonus of £4 for this study.
Otherwise, you only receive your participation fee of £1.
Any bonus you earn will be in addition to your guaranteed participation fee of £1.
Once you have returned to Prolific by clicking the forward button below, you can close this window.
Thank you for your participation!
<<
>>

B.2 Race Experiment



0% 100%

Instructions (1 of 2)

This study will describe 15 situations, and will ask you how **socially appropriate** certain behavior is in these situations.

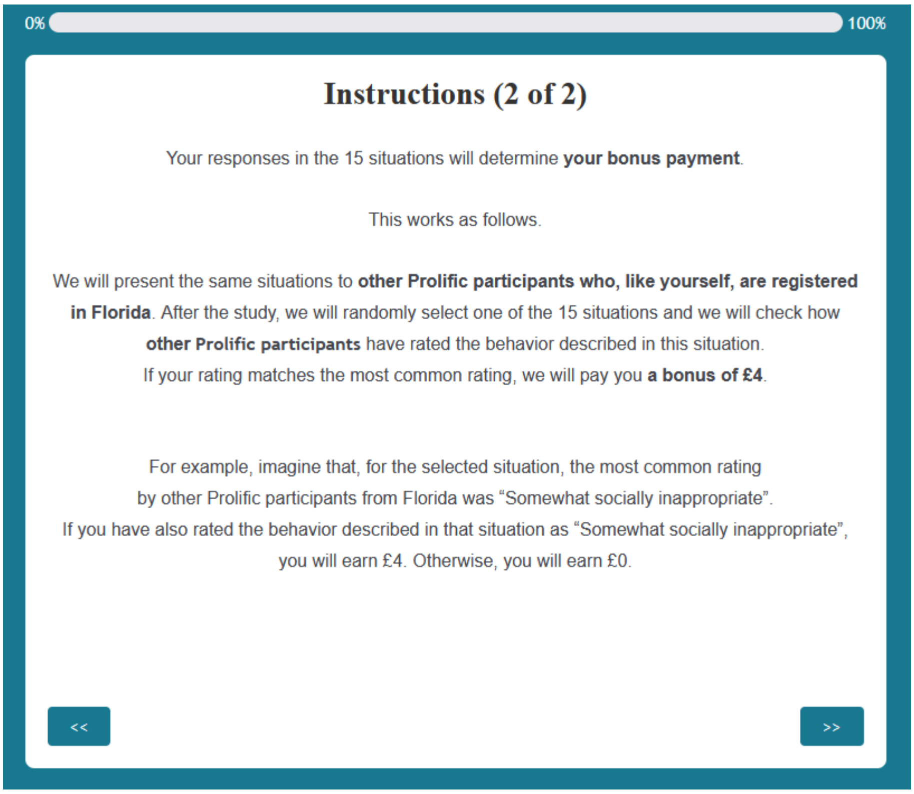
There will be six possible responses, of which you must select exactly one:

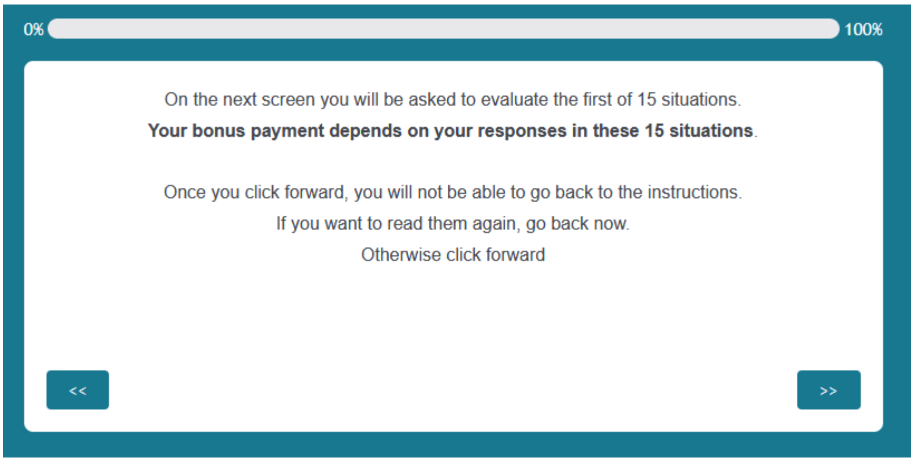
- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

By socially appropriate, we mean behavior that you think most people would agree is the **"right" thing to do**. Another way to think about what we mean is that if someone were to behave in a socially inappropriate way, then **other people might disapprove**.

Note that the “right” thing to do **may not necessarily be made explicit or supported by laws**. So an action may be 'appropriate' even if it is not legal; or ‘inappropriate’ even if it is not illegal. Rather, an appropriate action is an action that most people believe ought to be taken (regardless of whether it is legal or not).

<< >>


0%
100%
Instructions (2 of 2)
Your responses in the 15 situations will determine **your bonus payment**.
This works as follows.
We will present the same situations to **other Prolific participants who, like yourself, are registered in Florida**. After the study, we will randomly select one of the 15 situations and we will check how **other Prolific participants** have rated the behavior described in this situation.
If your rating matches the most common rating, we will pay you **a bonus of £4**.
For example, imagine that, for the selected situation, the most common rating by other Prolific participants from Florida was "Somewhat socially inappropriate". If you have also rated the behavior described in that situation as "Somewhat socially inappropriate", you will earn £4. Otherwise, you will earn £0.
<<
>>


0%
100%
On the next screen you will be asked to evaluate the first of 15 situations.
Your bonus payment depends on your responses in these 15 situations.
Once you click forward, you will not be able to go back to the instructions.
If you want to read them again, go back now.
Otherwise click forward
<<
>>

0% 100%

Cheng is a 50 year old Asian American farm worker, living in rural Florida. He is helping a friend transport goods between two local farms. The farms are 500 feet apart and there are no other cars on the road. Cheng fills his car with boxes and prepares to drive between the two farms. On the radio he hears a news report reminding listeners that it is illegal to drive without wearing a seatbelt. Cheng does not wear a seatbelt, and drives the 500 feet to the other farm at a maximum speed of 10 miles per hour.

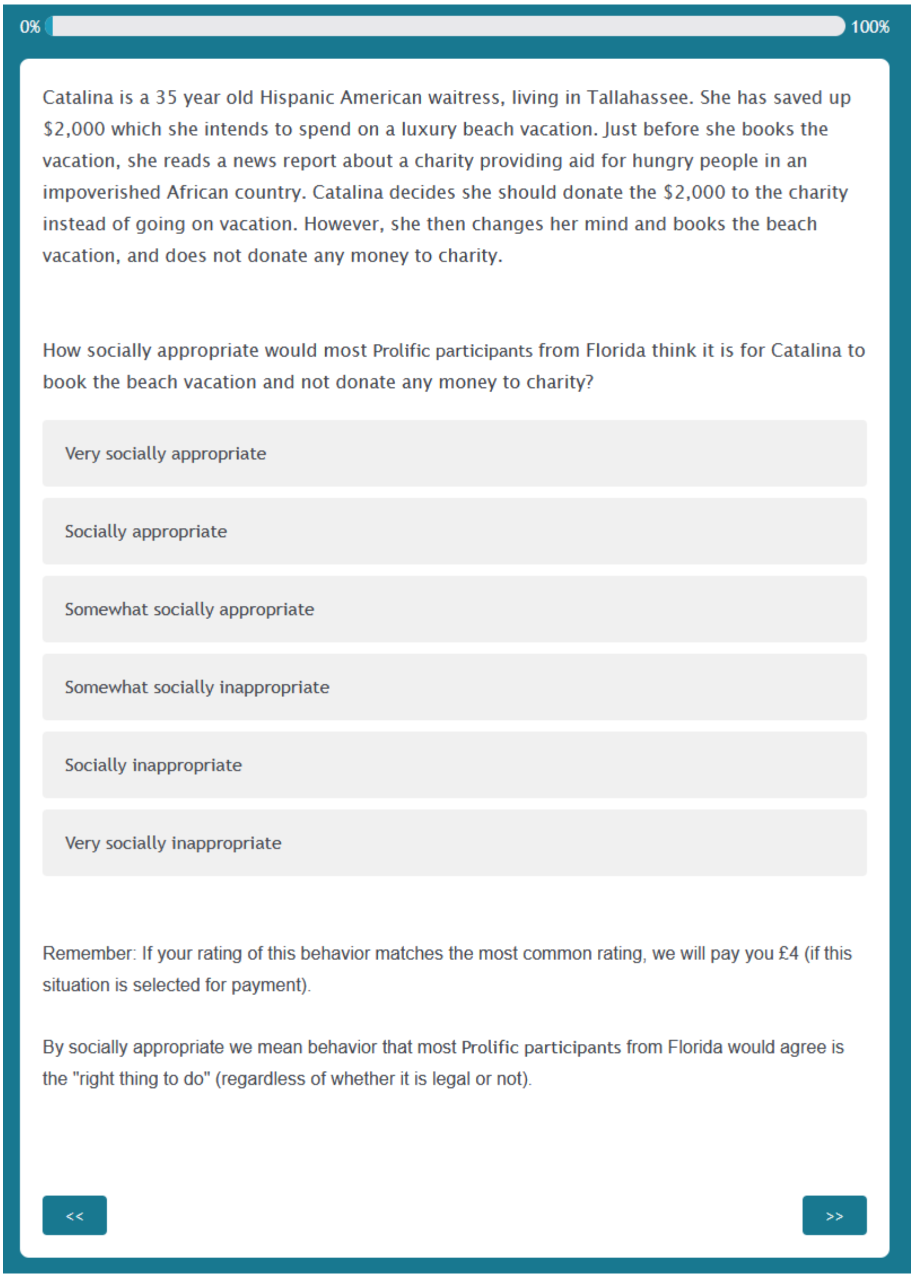
How socially appropriate would most Prolific participants from Florida think it is for Cheng to drive this journey without wearing a seatbelt?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

>>



0% 100%

Catalina is a 35 year old Hispanic American waitress, living in Tallahassee. She has saved up $2,000 which she intends to spend on a luxury beach vacation. Just before she books the vacation, she reads a news report about a charity providing aid for hungry people in an impoverished African country. Catalina decides she should donate the $2,000 to the charity instead of going on vacation. However, she then changes her mind and books the beach vacation, and does not donate any money to charity.

How socially appropriate would most Prolific participants from Florida think it is for Catalina to book the beach vacation and not donate any money to charity?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

Jessica is a 22 year old Asian American student living in Jacksonville. She wants to watch a Netflix TV show, but does not have a Netflix subscription. Her friends tell her about a website they found where she can download the show for free, but they warn her that downloading the show from this website is illegal. Later, Jessica visits the website but decides not to download the show.

How socially appropriate would most Prolific participants from Florida think it is for Jessica not to download the show?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

Ebony is a young, African American sales representative from Tampa. She is driving between two cities in Florida in order to attend a meeting. She turns onto a highway and sees a sign stating that the speed limit is 70 miles per hour. Ebony drives for the next five minutes at 67 miles per hour, before reaching her exit.

How socially appropriate would most Prolific participants from Florida think it is for Ebony to drive at 67 miles per hour?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

>>

0% 100%

Cadence is a middle-aged African American lawyer, living in Tampa Bay. On her way to meet a friend, she sees an injured man lying on the sidewalk. Cadence stops to ask the man if he is OK; the man tells her he has been attacked and had his wallet and cell phone stolen. Cadence phones the police and waits with the man until they arrive.

How socially appropriate would most Prolific participants from Florida think it is for Cadence to help the man in this way?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

Daniela is a 19 year old Hispanic American student at Florida State University, taking an exam which she expects to fail. Before the exam she writes some notes on a sheet of paper, and hides it under her sleeve. Just before the exam begins, the professor makes an announcement reminding all students that it is an offense for them to bring any materials into the exam to help them answer questions. During the exam Daniela secretly looks at the notes and uses them to help her answer the questions.

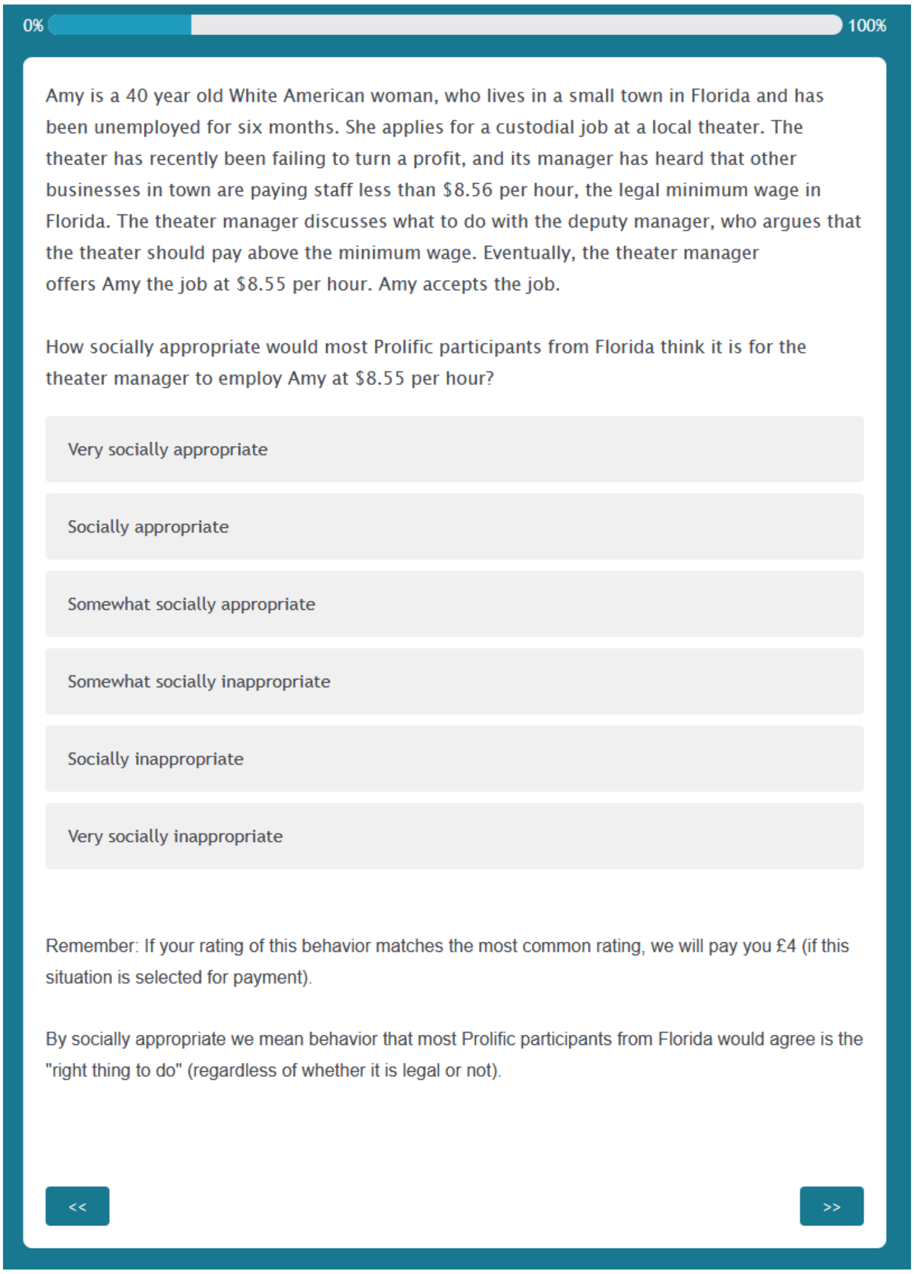
How socially appropriate would most Prolific participants from Florida think it is for Daniela to use the notes this way?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>



0% 100%

Amy is a 40 year old White American woman, who lives in a small town in Florida and has been unemployed for six months. She applies for a custodial job at a local theater. The theater has recently been failing to turn a profit, and its manager has heard that other businesses in town are paying staff less than $8.56 per hour, the legal minimum wage in Florida. The theater manager discusses what to do with the deputy manager, who argues that the theater should pay above the minimum wage. Eventually, the theater manager offers Amy the job at $8.55 per hour. Amy accepts the job.

How socially appropriate would most Prolific participants from Florida think it is for the theater manager to employ Amy at $8.55 per hour?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

Michael is a 53 year old Asian American doctor. On his way home from work in your local city, he is approached by a man begging for money. Michael ignores the beggar and walks away from him.

How socially appropriate would most Prolific participants from Florida think it is for Michael to ignore the beggar and walk away from him?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

Cody is a young White American cashier, living in Miami, who is a registered medical marijuana user. Today he has an appointment at his local Medical Marijuana Treatment Center. Before going, he weighs his remaining stock of marijuana at home, and finds that he has 2 ounces. At the treatment center, a staff member asks him how much marijuana he would like to buy. "As much as possible!" Cody replies. The staff member tells him: "The limit you can buy is 2.5 ounces. And the total limit you can legally possess is 4 ounces, so you should only buy 2.5 ounces if you don't already possess more than 1.5 ounces." Cody mentally computes that buying 2.5 ounces will leave him with 0.5 ounces more than the legal limit for possession of 4 ounces. Cody then buys 2.5 ounces.

How socially appropriate would most Prolific participants from Florida think it is for Cody to buy 2.5 ounces on this occasion?

Very socially appropriate

Socially appropriate

Somewhat socially appropriate

Somewhat socially inappropriate

Socially inappropriate

Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

Logan is a White American carpenter, living in a dangerous district of Miami with his wife and one son. His son wants a gun for protection and asks Logan to buy one for the family. Logan agrees. When he buys the gun, the store owner draws his attention to a notice at the counter: “IT IS UNLAWFUL TO STORE OR LEAVE A FIREARM IN ANY PLACE WITHIN THE REACH OR EASY ACCESS OF A MINOR UNDER 18 YEARS OF AGE.” Logan mentions that his son turned 18 7 days ago. He returns home and gives his son the gun.

How socially appropriate would most Prolific participants from Florida think it is for Logan to give his son the gun?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

Emma is a 35 year old White American woman working for a company in Orlando which manufactures state-of-the-art breathalyzers, machines which can measure a person's blood alcohol content with extremely high accuracy. One day, Emma has been drinking at home and decides to go to a bar. She remembers she has one of the breathalyzers at home, and wonders whether her blood alcohol content is below 0.08%, the maximum level at which a person can legally drive in Florida. Emma tests herself and discovers that her blood alcohol content is 0.083%. Emma then drives to the bar.

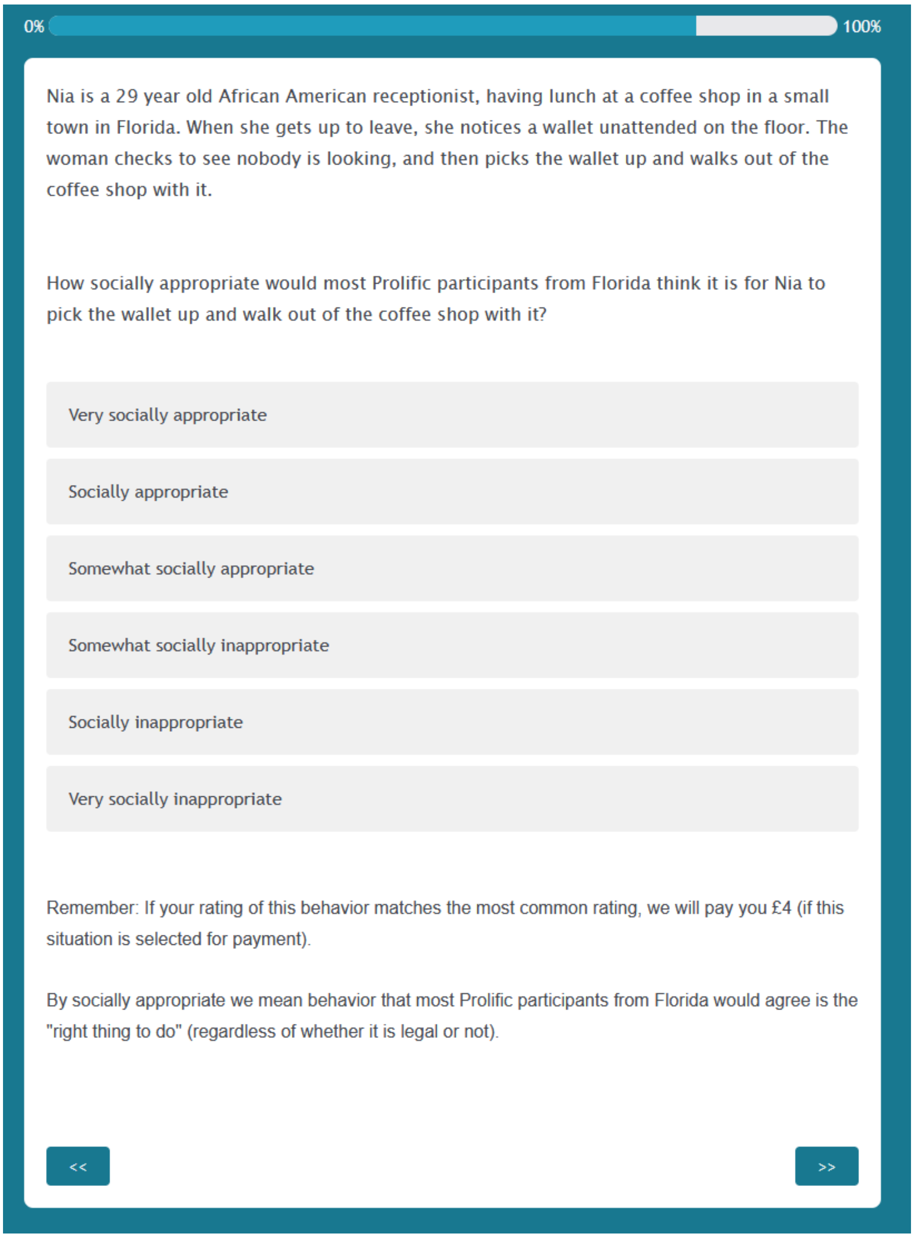
How socially appropriate would most Prolific participants from Florida think it is for Emma to drive to the bar?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>


0% 100%

Nia is a 29 year old African American receptionist, having lunch at a coffee shop in a small town in Florida. When she gets up to leave, she notices a wallet unattended on the floor. The woman checks to see nobody is looking, and then picks the wallet up and walks out of the coffee shop with it.

How socially appropriate would most Prolific participants from Florida think it is for Nia to pick the wallet up and walk out of the coffee shop with it?

Very socially appropriate

Socially appropriate

Somewhat socially appropriate

Somewhat socially inappropriate

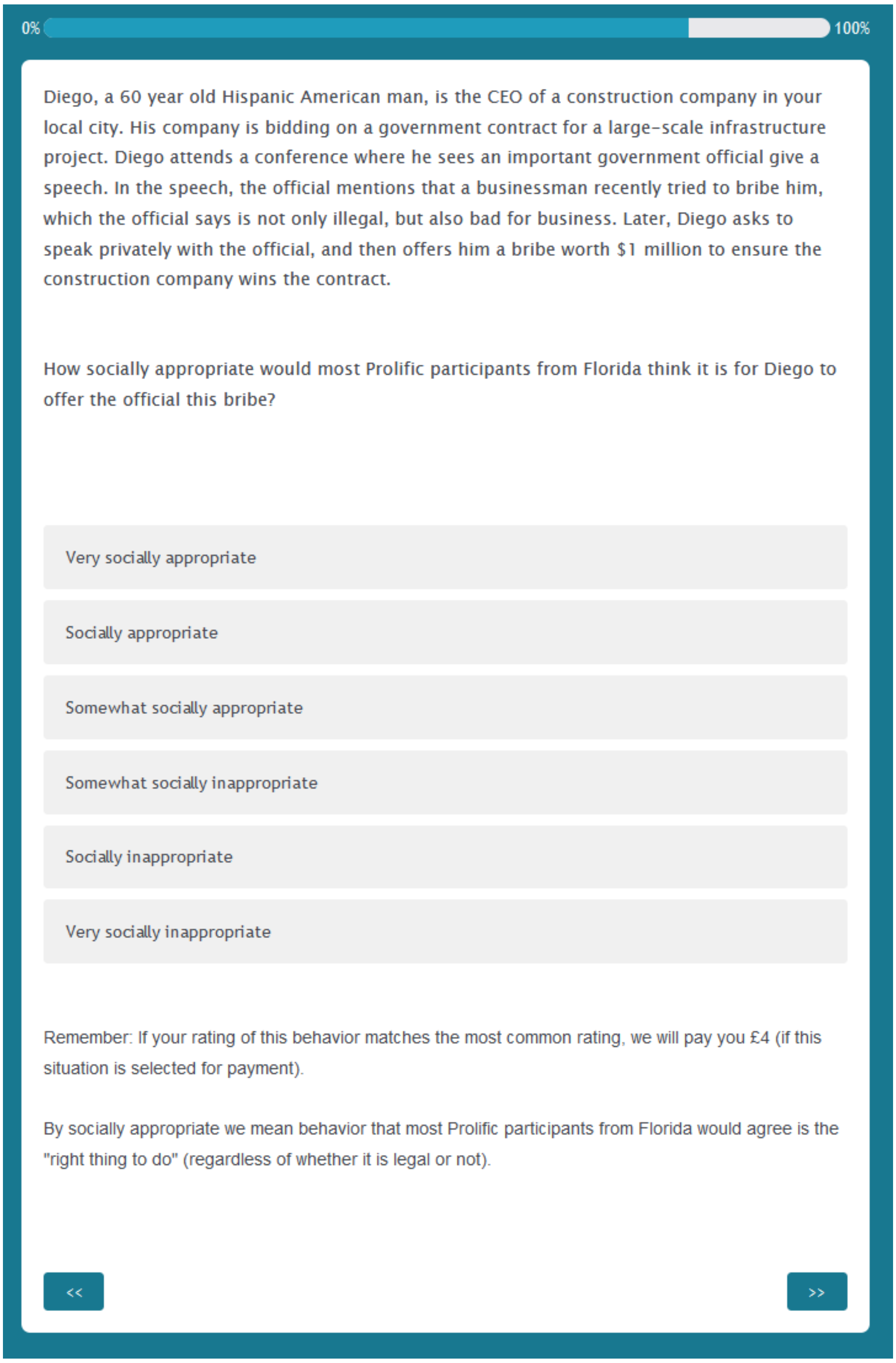
Socially inappropriate

Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>


0%
100%
Diego, a 60 year old Hispanic American man, is the CEO of a construction company in your local city. His company is bidding on a government contract for a large-scale infrastructure project. Diego attends a conference where he sees an important government official give a speech. In the speech, the official mentions that a businessman recently tried to bribe him, which the official says is not only illegal, but also bad for business. Later, Diego asks to speak privately with the official, and then offers him a bribe worth $1 million to ensure the construction company wins the contract.
How socially appropriate would most Prolific participants from Florida think it is for Diego to offer the official this bribe?
Very socially appropriate
Socially appropriate
Somewhat socially appropriate
Somewhat socially inappropriate
Socially inappropriate
Very socially inappropriate
Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).
By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).
<<
>>

0% 100%

Jermaine is a 24 year old African American student at Florida State University. One night, he meets Tanisha, an African American high school student at a party. Jermaine invites Tanisha to come to his home, and she agrees. At his home, Jermaine tells Tanisha he wants to have sex with her, but that she looks young, and asks whether she is above the legal age of consent of 18 years. Tanisha tells Jermaine that she has turned 18 7 days ago, and shows him an ID card which confirms this. She tells him that she wants to have sex with him. Jermaine then has sex with Tanisha.

How socially appropriate would most Prolific participants from Florida think it is for Jermaine to have sex with Tanisha?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>

0% 100%

Gary is a middle-aged White American management consultant in Tampa Bay. He wakes in the middle of the night and finds a thief attempting to steal his television. Gary catches the thief and beats him severely, breaking the thief's arm before throwing him out through the front door.

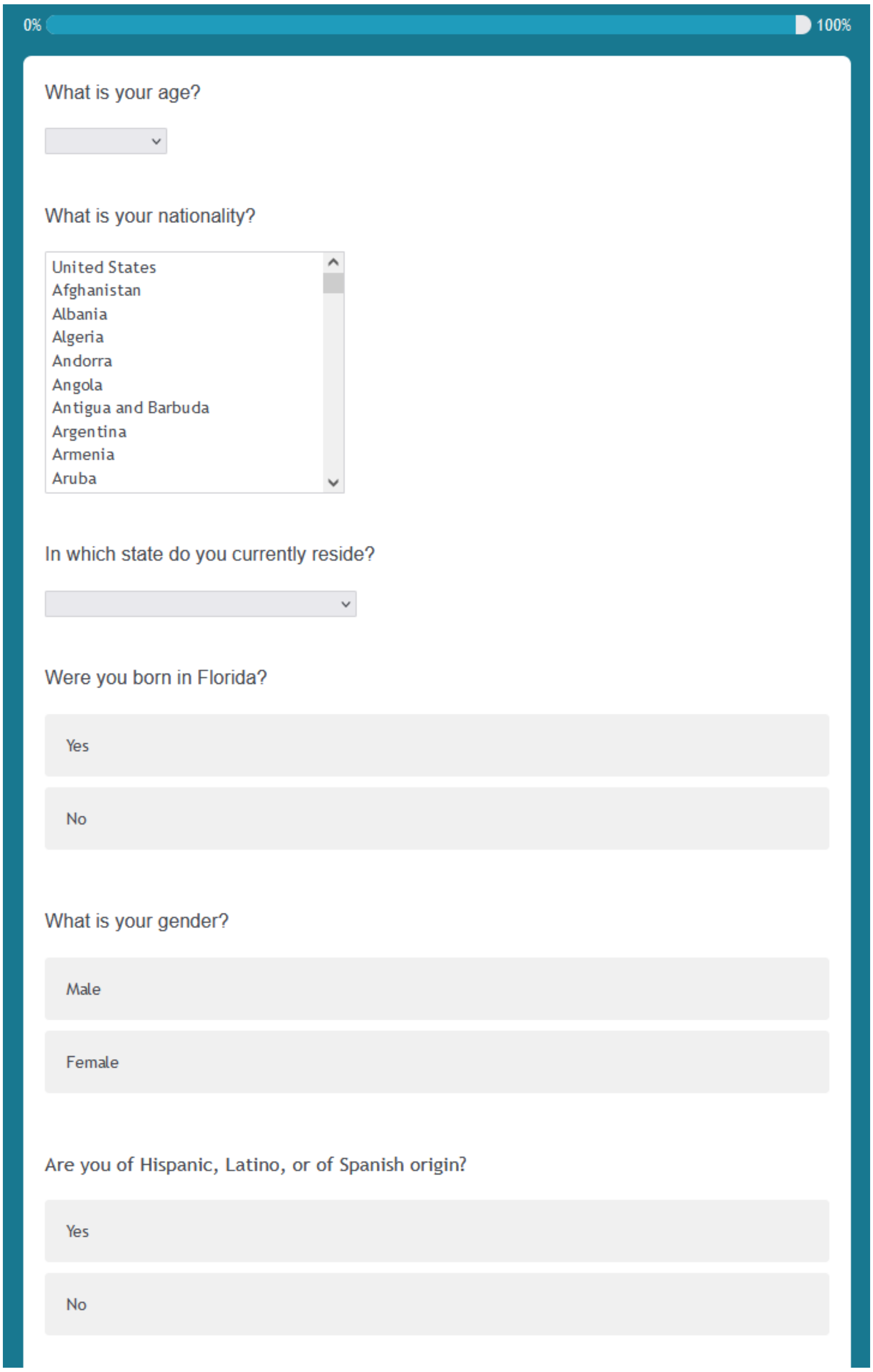
How socially appropriate would most Prolific participants from Florida think it is for Gary to treat the thief this way?

- Very socially appropriate
- Socially appropriate
- Somewhat socially appropriate
- Somewhat socially inappropriate
- Socially inappropriate
- Very socially inappropriate

Remember: If your rating of this behavior matches the most common rating, we will pay you £4 (if this situation is selected for payment).

By socially appropriate we mean behavior that most Prolific participants from Florida would agree is the "right thing to do" (regardless of whether it is legal or not).

<< >>



0% 100%

What is your age?

What is your nationality?

United States
Afghanistan
Albania
Algeria
Andorra
Angola
Antigua and Barbuda
Argentina
Armenia
Aruba

In which state do you currently reside?

Were you born in Florida?

Yes

No

What is your gender?

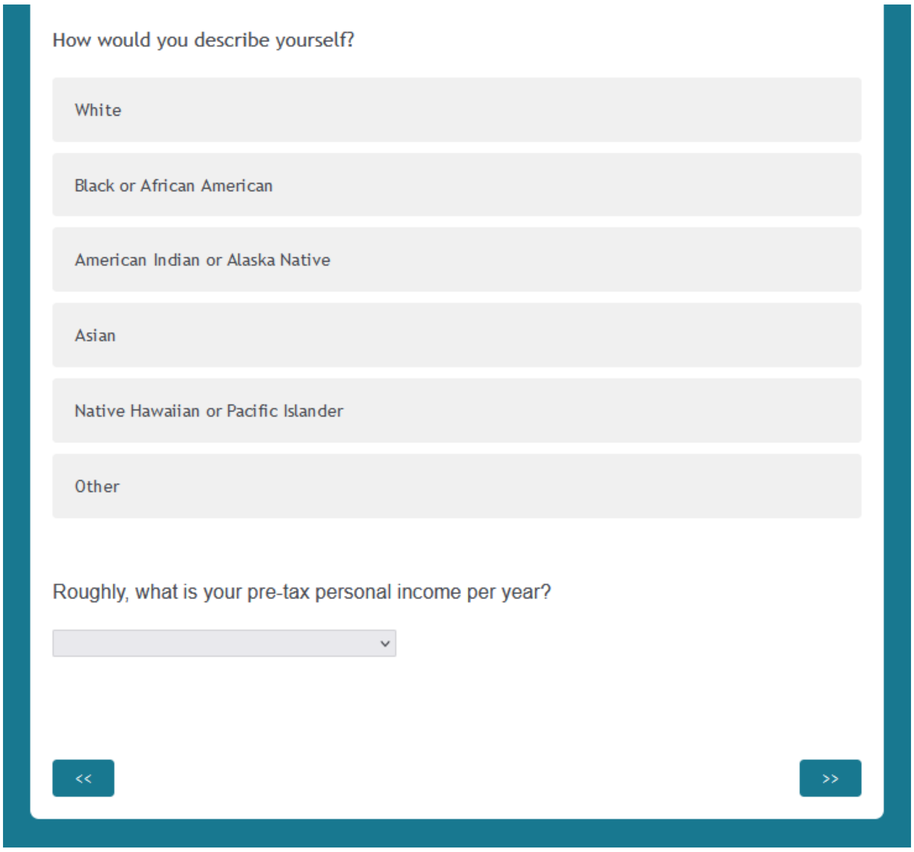
Male

Female

Are you of Hispanic, Latino, or of Spanish origin?

Yes

No



How would you describe yourself?

- White
- Black or African American
- American Indian or Alaska Native
- Asian
- Native Hawaiian or Pacific Islander
- Other

Roughly, what is your pre-tax personal income per year?

<< >>

0% 100%

We now read you a series of statements and we ask you to indicate whether or not you agree with them.

On average African Americans have worse jobs, income, and housing than White Americans.

This is because...

...Of discrimination against African Americans.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

...Most African Americans don't have the chance for education that it takes to rise out of poverty.

- Strongly Agree
- Agree
- Disagree
- Strongly disagree

...Most African Americans just don't have the motivation or will power to pull themselves up out of poverty.

Strongly Agree

Agree

Disagree

Strongly disagree

...Most African Americans have less in-built ability to learn.

Strongly Agree

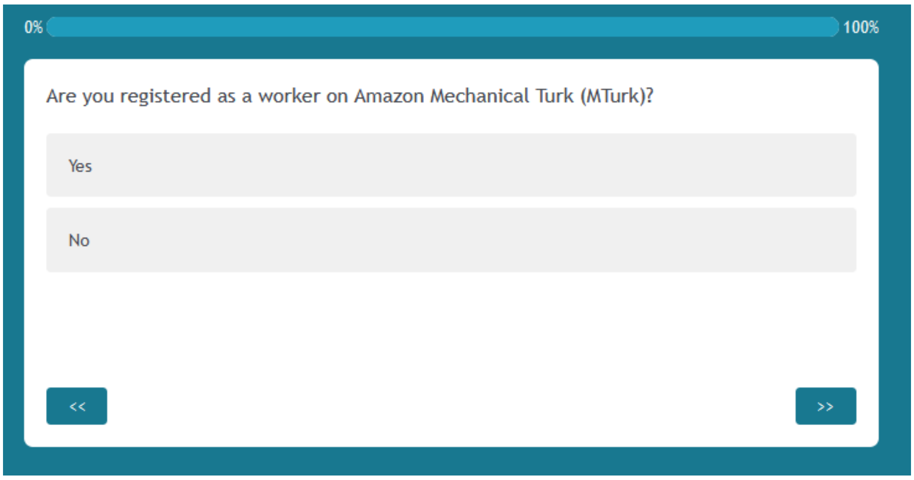
Agree

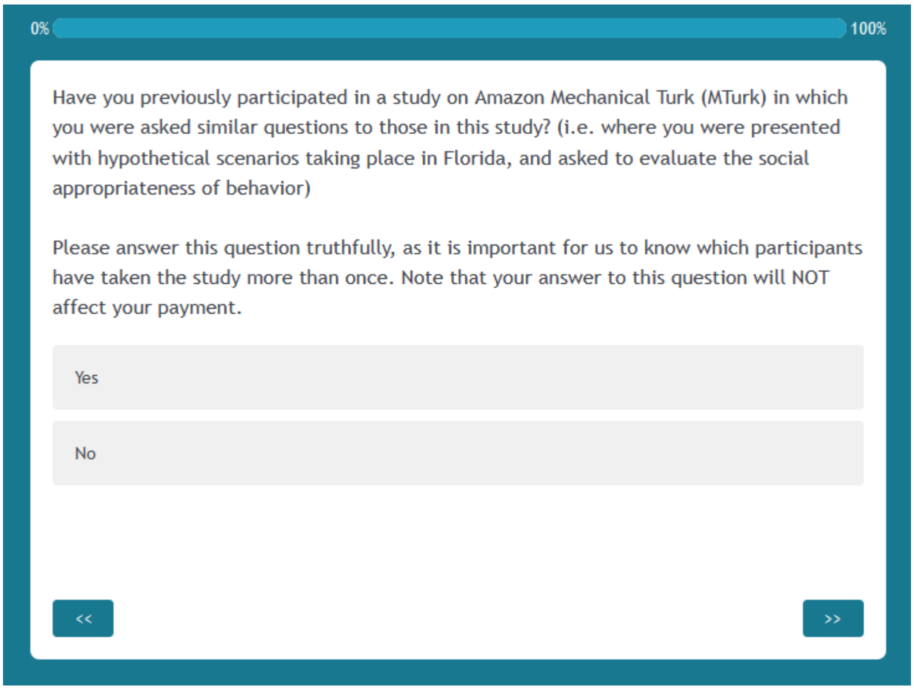
Disagree

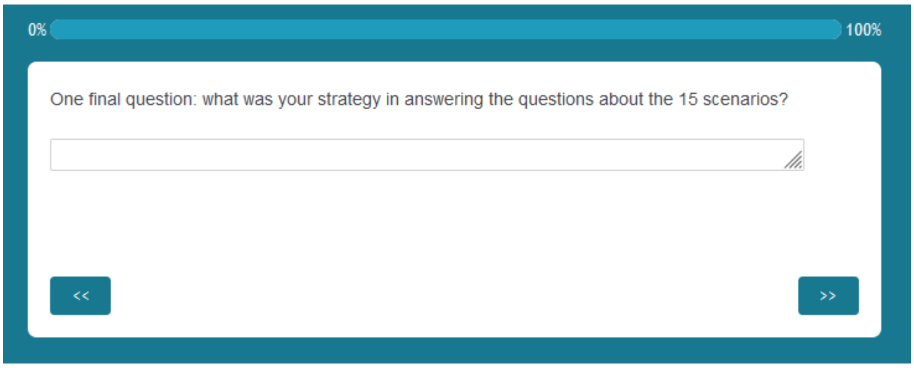
Strongly disagree

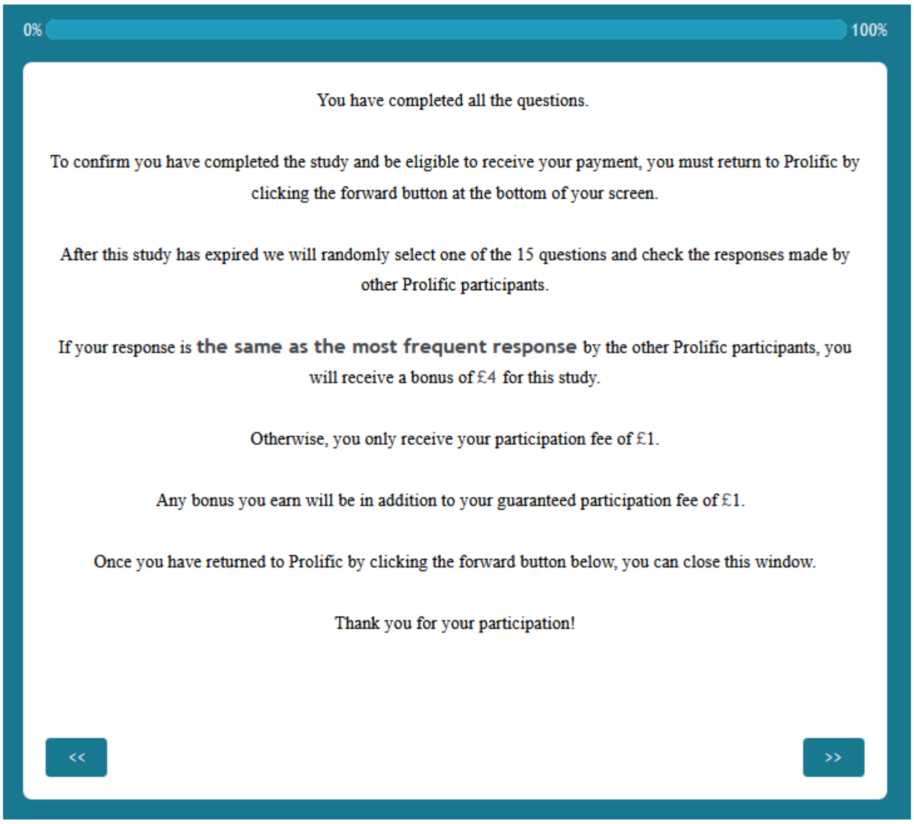
<<

>>

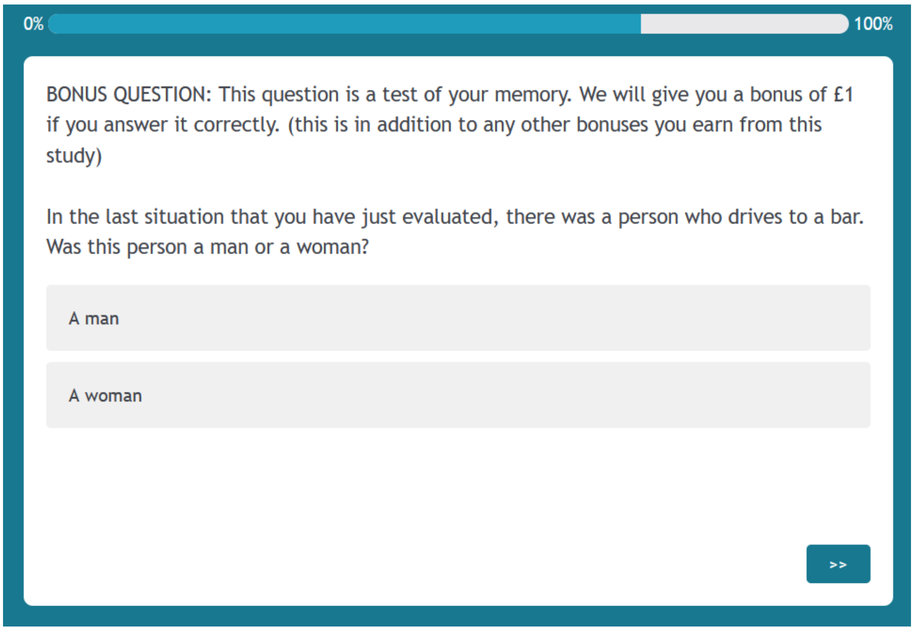

0%
100%
Are you registered as a worker on Amazon Mechanical Turk (MTurk)?
Yes
No
<<
>>


0%
100%
Have you previously participated in a study on Amazon Mechanical Turk (MTurk) in which you were asked similar questions to those in this study? (i.e. where you were presented with hypothetical scenarios taking place in Florida, and asked to evaluate the social appropriateness of behavior)
Please answer this question truthfully, as it is important for us to know which participants have taken the study more than once. Note that your answer to this question will NOT affect your payment.
Yes
No
<<
>>

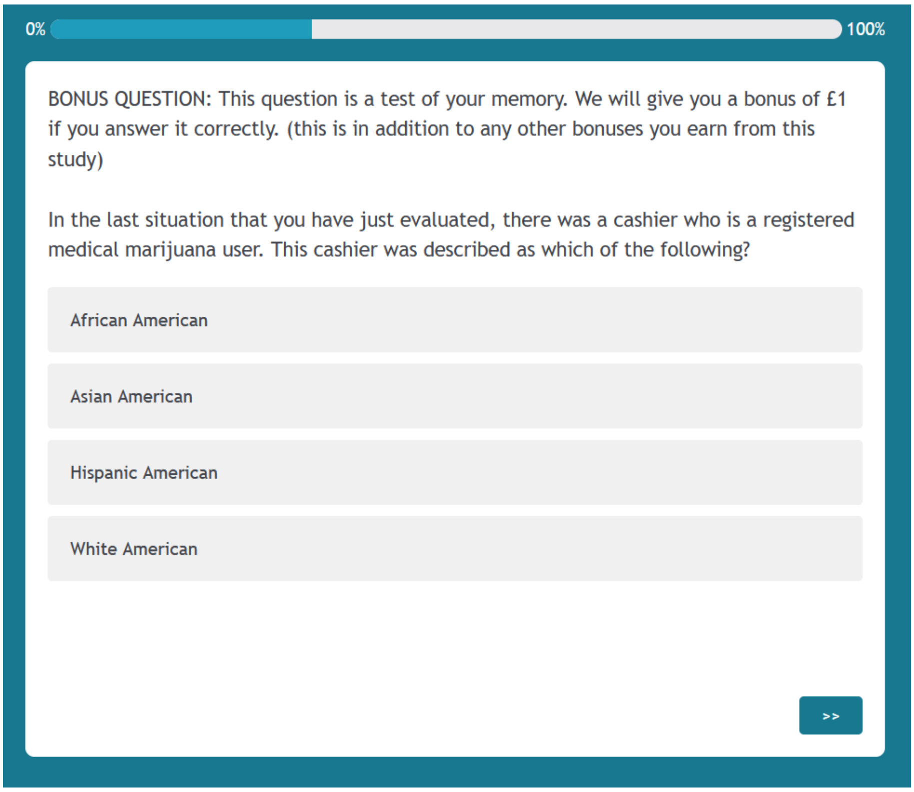

0%
100%
One final question: what was your strategy in answering the questions about the 15 scenarios?
<<
>>


0%
100%
You have completed all the questions.
To confirm you have completed the study and be eligible to receive your payment, you must return to Prolific by clicking the forward button at the bottom of your screen.
After this study has expired we will randomly select one of the 15 questions and check the responses made by other Prolific participants.
If your response is the same as the most frequent response by the other Prolific participants, you will receive a bonus of £4 for this study.
Otherwise, you only receive your participation fee of £1.
Any bonus you earn will be in addition to your guaranteed participation fee of £1.
Once you have returned to Prolific by clicking the forward button below, you can close this window.
Thank you for your participation!
<<
>>

B.3 Attention check Gender



B.4 Attention check Race



0% 100%

BONUS QUESTION: This question is a test of your memory. We will give you a bonus of £1 if you answer it correctly. (this is in addition to any other bonuses you earn from this study)

In the last situation that you have just evaluated, there was a cashier who is a registered medical marijuana user. This cashier was described as which of the following?

- African American
- Asian American
- Hispanic American
- White American

>>

B.5 Race pilot names

0% 100%

Welcome!

Thank you for participating in our study.
Including the time for reading these instructions, the study will take about **10 minutes** to complete.

During the study, please do not close this window or leave the HIT's web pages in any other way.
If you do close your browser or leave the HIT, you will not be able to re-enter, and we will not be able to pay you.

In addition of your guaranteed **participation fee of $0.50**, you will receive an additional **bonus payment of $1** upon completion of the HIT.
You will receive a code to collect your payment via Mturk at the end of this HIT.

→

0% 100%

Instructions

In this HIT we will show you two lists of names.
One list will contain **40 male names** and the other will contain **40 female names**.

For each list, we will ask you a question about the names contained in the list.
We ask you to answer to each question as honestly and as accurately as possible.

When you are ready to start the task, click forward.

→

0% 100%

Female names

For each of the 40 names listed below, imagine that this is an American woman living in Florida. We would like you to think about the race that a woman with this name would typically have.
How likely is she to be **African American**? You will indicate your response on a scale from 0 to 10, where 0 means "very unlikely" and 10 means "very likely"

Name: Carly

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Laurie

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Latonya

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Deja

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Molly

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Asia

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Katie

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Keisha

0 = Very unlikely African American — Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Shanice

0 = Very unlikely African American — Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Jenna

0 = Very unlikely African American — Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Tanisha

0 = Very unlikely African American — Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Aisha

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10

Name: Aliyah

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10

Name: Imani

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10

Name: Meredith

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10

Name: Tamika

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Sarah

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Heather

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Emma

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Kenya

0 = Very unlikely African American — Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Aaliyah

0 = Very unlikely African American — Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Ebony

0 = Very unlikely African American — Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Kristen

0 = Very unlikely African American — Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Tierra

0 = Very unlikely African American   Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Carrie

0 = Very unlikely African American   Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Precious

0 = Very unlikely African American   Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Allison

0 = Very unlikely African American   Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Jill

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Madeline

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Amy

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Anne

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Katelyn

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Lakisha

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Claire

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Emily

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Jada

0 = Very unlikely African American										Very likely African American = 10
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Nia

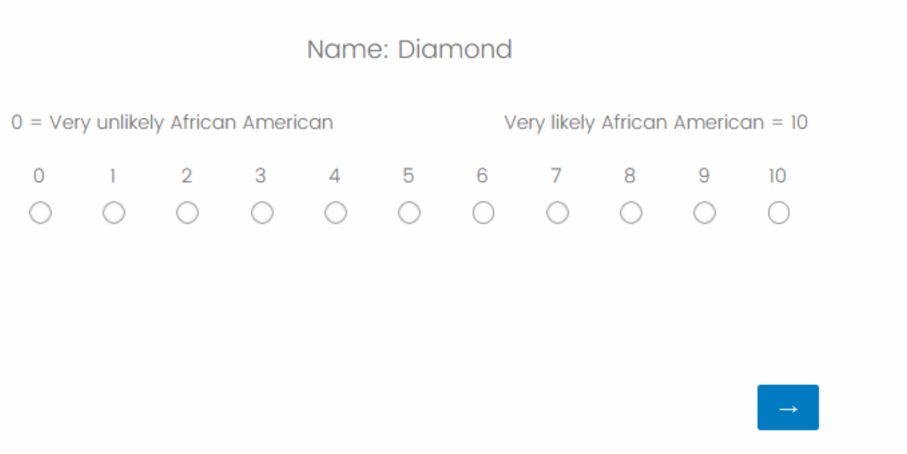
0 = Very unlikely African American										Very likely African American = 10
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Abigail

0 = Very unlikely African American										Very likely African American = 10
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Latoya

0 = Very unlikely African American										Very likely African American = 10
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○


Name: Diamond
0 = Very unlikely African American
Very likely African American = 10
0 1 2 3 4 5 6 7 8 9 10
→

0% 100%

Male names

For each of the 40 names listed below, imagine that this is an American man living in Florida. We would like you to think about the race that a man with this name would typically have. How likely is he to be **African American**? You will indicate your response on a scale from 0 to 10, where 0 means "very unlikely" and 10 means "very likely"

Name: DeShawn

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Hakim

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Greg

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Maurice

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Jack

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Trevon

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Jay

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Matthew

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Dustin

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Dominique

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Tanner

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Demetrius

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Willie

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Brad

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Leroy

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Tyrone

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Neil

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Logan

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Wyatt

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Jamal

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Brett

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Jake

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Geoffrey

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Todd

0 = Very unlikely African American Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Cody

0 = Very unlikely African American Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Malik

0 = Very unlikely African American Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Luke

0 = Very unlikely African American Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Marquis

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Connor

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Terrell

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: DeAndre

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Cole

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Tremayne

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Rasheed

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Reginald

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Scott

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Jermaine

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Brendan

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Kareem

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Darnell

0 = Very unlikely African American | Very likely African American = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

→

0% 100%

End of Study

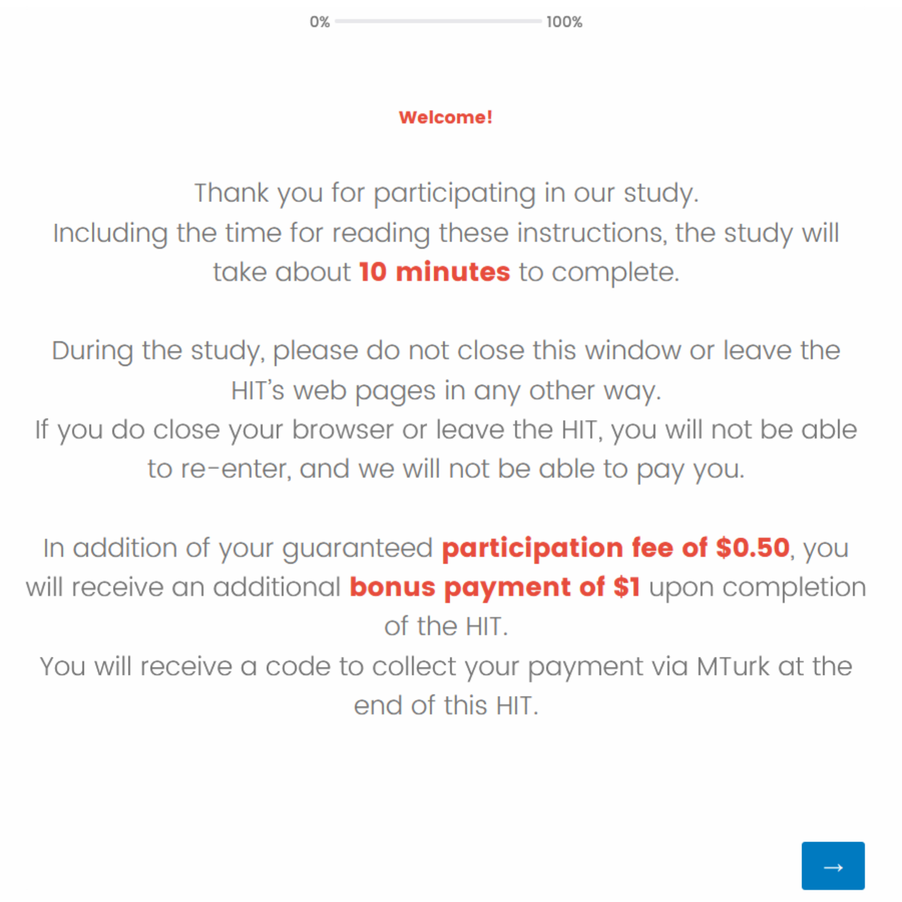
Thank you for participating in our study. This is the end of the HIT.
A bonus of $1 will be paid on top of your guaranteed participation fee of $0.50.

To receive your payment, please copy the following code and paste it into Mturk.

7096424

Once you have done that, you can close this window.

B.6 Race pilot socioeconomic status


0% 100%

Welcome!

Thank you for participating in our study.
Including the time for reading these instructions, the study will take about **10 minutes** to complete.

During the study, please do not close this window or leave the HIT's web pages in any other way.
If you do close your browser or leave the HIT, you will not be able to re-enter, and we will not be able to pay you.

In addition of your guaranteed **participation fee of $0.50**, you will receive an additional **bonus payment of $1** upon completion of the HIT.
You will receive a code to collect your payment via MTurk at the end of this HIT.

→

0% 100%

Instructions

In this HIT we will show you two lists of names.
One list will contain **10 male names** and the other will contain **10 female names**.

For each list, we will ask you about the most likely socio-economic status of people with the names contained in the list.
Think about **socio-economic status** in the following way: a person with high socio-economic status generally has higher income, higher wealth, and more years of education.
We ask you to answer to each question as honestly and as accurately as possible.

When you are ready to start the task, click forward.

→

0% 100%

Female names

Think about the typical **socio-economic status** of a **White American adult woman living in Florida**. (Remember: a person with high socio-economic status generally has higher income, higher wealth, and more years of education.)

For each of the 10 names listed below, we ask you to indicate whether a White American woman in Florida with that name is **likely to have a higher, lower or the same** socio-economic status **compared to a typical White American woman** in Florida.
You will answer on a scale from 0 to 10, where 0 means "definitely lower status", 5 means " definitely the same status" and 10 means "definitely higher status".

Name: Emma

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Carly

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Carrie

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Madeline

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Emily

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Abigail

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Katie

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Jill

0 = Definitely lower status　　　　Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Amy

0 = Definitely lower status　　　　Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Molly

0 = Definitely lower status　　　　Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

→

0% 100%

Male names

Think about the typical **socio-economic status** of a **White American adult man living in Florida**. (Remember: a person with high socio-economic status generally has higher income, higher wealth, and more years of education.)

For each of the 10 names listed below, we ask you to indicate whether a White American man in Florida with that name is **likely to have a higher, lower or the same** socio-economic status **compared to a typical White American man** in Florida.
You will answer on a scale from 0 to 10, where 0 means "definitely lower status", 5 means " definitely the same status" and 10 means "definitely higher status".

Name: Wyatt

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Todd

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Brad

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Scott

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Logan

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Tanner

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Connor

0 = Definitely lower status | Definitely higher status = 10

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Cody

0 = Definitely lower status										Definitely higher status = 10
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Dustin

0 = Definitely lower status										Definitely higher status = 10
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Name: Matthew

0 = Definitely lower status										Definitely higher status = 10
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

→

0% 100%

End of Study

Thank you for participating in our study. This is the end of the HIT.
A bonus of $1 will be paid on top of your guaranteed participation fee of $0.50.

To receive your payment, please copy the following code and paste it into Mturk.

1964557

Once you have done that, you can close this window.

C Experimental conditions / vignette versions

C.1 Gender Experiment

Below, we present the wordings in the male condition. For the female condition, 'man' is replaced by 'woman' and 'he' by 'she'.

Drink-Driving Vignette

A man works for a company in Houston which manufactures state-of-the-art breathalyzers, machines which can measure a person's blood alcohol content with extremely high accuracy. One day, the man has been drinking at home and decides to go to a bar. He remembers he has one of the breathalyzers at home, and wonders whether his blood alcohol content is below 0.08%, the maximum level at which a person can legally drive in Texas. He tests himself and discovers that his blood alcohol content is (BAC)*. The man then drives to the bar.

How socially appropriate would most (Subjects)** from Texas think it is for the man to drive to the bar?

***The possible values of BAC were 0.073%, 0.075%, 0.077%, 0.079%, 0.081%, 0.083%, 0.085% and 0.087%.**

****The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

Speeding Vignette

A man is driving between two cities in Texas in order to attend a meeting. He turns onto a highway and sees a sign stating that the speed limit is 75 miles per hour. The man drives for the next five minutes at (Speed)* miles per hour, before reaching his exit.

How socially appropriate would most (Subjects)** from Texas think it is for the man to drive at (Speed)* miles per hour?

***The possible values of Speed were 68, 70, 72, 74, 76, 78, 80 and 82.**

****The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

Alcohol to Youth Vignette

A man owns a store in a small town in Texas. One day, a young customer enters the store with the intention of buying some beer. The customer sees a sign in the store reminding customers that in the United States it is illegal for store owners to sell alcohol to people under the age of 21. The store owner is the father of a classmate of the customer and knows that the customer (Age)*. He also knows that the customer often gets drunk and vandalizes property in the neighborhood. The customer brings a 24-pack of alcoholic beer up to the counter. The man looks at the customer who appears sober. He then sells the beer to the customer.

How socially appropriate would most (Subjects)** from Texas think it is for the man to sell the beer to the customer?

***The possible wordings for Age were "will turn 21 in 7 days", "will turn 21 in 5 days", "will turn 21 in 3 days", "will turn 21 in 1 day", "turned 21 1 day ago", "turned 21 3 days ago", "turned 21 5 days ago" and "turned 21 7 days ago".**

****The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

<u>Casino Vignette</u>

A man works for a casino in Texas. His job is to check customers' ID and prevent those below the legal gambling age of 21 from entering the casino. One evening, he sees a young customer in the line and asks for ID. The ID shows the customer (Age)*. The man lets the customer enter the casino.

How socially appropriate would most (Subjects)** from Texas think it is for the man to let this customer enter the casino?

***The possible wordings for Age were "will turn 21 in 7 days", "will turn 21 in 5 days", "will turn 21 in 3 days", "will turn 21 in 1 day", "turned 21 1 day ago", "turned 21 3 days ago", "turned 21 5 days ago" and "turned 21 7 days ago".**

****The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

<u>Minimum Wage Vignette</u>

A 40 year old man, who lives in a small town in Texas, has been unemployed for six months. He applies for a custodial job at a local theater. The theater has recently been failing to turn a profit, and its manager has heard that other businesses in town are paying staff less than $7.25 per hour, the legal minimum wage in Texas. The theater manager discusses what to do with the deputy manager, who argues that the theater should pay above the minimum wage. Eventually, the theater manager offers the man the job at (Wage)* per hour. The man accepts the job.

How socially appropriate would most (Subjects)** from Texas think it is for the theater manager to employ the man at (Wage)* per hour?

***The possible values of Wage were $7.18, $7.20, $7.22, $7.24, $7.26, $7.28, $7.30 and $7.32.**

****The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

Parental Leave Vignette

A man works at a large warehouse in Texas. The man expects his first child and requests a period of unpaid leave from his manager. The manager consults with the company's legal department about the man's legal rights. The legal department tells the manager that the federal Family and Medical Leave Act requires the company to offer the man at least 12 weeks of unpaid, job-protected leave. The manager then phones the man, and offers him (Number)* weeks of unpaid leave, after which he must immediately return to work – if he refuses to do so, he will be dismissed.

How socially appropriate would most (Subjects)** from Texas think it is for the manager to make this demand to the man?

***The possible values of Number were 5, 7, 9, 11, 13, 15, 17 and 19.**

****The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

C.2 Race Experiment

Drink-Driving Vignette

(Name)* is a 35 year old (Race)** woman working for a company in Orlando which manufactures state-of-the-art breathalyzers, machines which can measure a person's blood alcohol content with extremely high accuracy. One day, (Name)* has been drinking at home and decides to go to a bar. She remembers she has one of the breathalyzers at home, and wonders whether her blood alcohol content is below 0.08%, the maximum level at which a person can legally drive in Florida. (Name)* tests herself and discovers that her blood alcohol content is (BAC)***. (Name)* then drives to the bar.

How socially appropriate would most (Subjects)**** from Florida think it is for (Name)* to drive to the bar?

***The possible wordings for Name were: Aisha, Deja, Ebony, Imani, Keisha, Kenya, Lakisha, Latonya, Tanisha and Tamika in the African American condition; Abigail, Amy, Carrie, Carly, Emily, Emma, Jill, Katie, Madeline and Molly in the White American condition.**

****The wording for Race was "African American" in the African American condition, and "White American" in the White American condition.**

*****The possible values of BAC were 0.073%, 0.075%, 0.077%, 0.079%, 0.081%, 0.083%, 0.085% and 0.087%.**

******The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

Speeding Vignette

(Name)* is a young, (Race)** sales representative from Tampa. She is driving between two cities in Florida in order to attend a meeting. She turns onto a highway and sees a sign stating that the speed limit is 70 miles per hour. (Name)* drives for the next five minutes at (Speed)*** miles per hour, before reaching her exit.

How socially appropriate would most (Subjects)**** from Florida think it is for (Name)* to drive at (Speed)*** miles per hour?

***The possible wordings for Name were: Aisha, Deja, Ebony, Imani, Keisha, Kenya, Lakisha, Latonya, Tanisha and Tamika in the African American condition; Abigail, Amy, Carrie, Carly, Emily, Emma, Jill, Katie, Madeline and Molly in the White American condition.**

****The wording for Race was "African American" in the African American condition, and "White American" in the White American condition.**

*****The possible values of Speed were 63, 65, 67, 69, 71, 73, 75 and 77.**

******The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

Minimum Wage Vignette

(Name)* is a 40 year old (Race)** woman, who lives in a small town in Florida and has been unemployed for six months. She applies for a custodial job at a local theater. The theater has recently been failing to turn a profit, and its manager has heard that other businesses in town are paying staff less than $8.56 per hour, the legal minimum wage in Florida. The theater manager discusses what to do with the deputy manager, who argues that the theater should pay above the minimum wage. Eventually, the theater manager offers (Name)* the job at (Wage)*** per hour. (Name)* accepts the job.

How socially appropriate would most (Subjects)**** from Florida think it is for the theater manager to employ (Name)* at (Wage)*** per hour?

***The possible wordings for Name were: Aisha, Deja, Ebony, Imani, Keisha, Kenya, Lakisha, Latonya, Tanisha and Tamika in the African American condition; Abigail, Amy, Carrie, Carly, Emily, Emma, Jill, Katie, Madeline and Molly in the White American condition.**

****The wording for Race was "African American" in the African American condition, and "White American" in the White American condition.**

*****The possible values of Wage were $8.49, $8.51, $8.53, $8.55, $8.57, $8.59, $8.61 and $8.63.**

******The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

Gun Possession Vignette

(Name)* is (Race)** carpenter, living in a dangerous district of Miami with his wife and one son. His son wants a gun for protection and asks (Name)* to buy one for the family. (Name)* agrees. When he buys the gun, the store owner draws his attention to a notice at the counter: "IT IS UNLAWFUL TO STORE OR LEAVE A FIREARM IN ANY PLACE WITHIN THE REACH OR EASY ACCESS OF A MINOR UNDER 18 YEARS OF AGE." (Name)* mentions that his son (Age)***. He returns home and gives his son the gun.

How socially appropriate would most (Subjects)**** from Florida think it is for (Name)* to give his son the gun?

***The possible wordings for Name were: DeAndre, DeShawn, Hakim, Jamal, Kareem, Malik, Marquis, Rasheed, Tremayne, Tyrone in the African American condition; Brad, Cody, Connor, Dustin, Logan, Matthew, Scott, Tanner, Todd, Wyatt in the White American condition.**

****The wording for Race was "an African American" in the African American condition, and "a White American" in the White American condition.**

*****The possible wordings for Age were "will turn 18 in 7 days", "will turn 18 in 5 days", "will turn 18 in 3 days", "will turn 18 in 1 day", "turned 18 1 day ago", "turned 18 3 days ago", "turned 18 5 days ago" and "turned 18 7 days ago".**

******The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

Marijuana Vignette

(Name)* is a young (Race)** cashier, living in Miami, who is a registered medical marijuana user. Today he has an appointment at his local Medical Marijuana Treatment Center. Before going, he weighs his remaining stock of marijuana at home, and finds that he has (Amount)*** ounces. At the treatment center, a staff member asks him how much marijuana he would like to buy. "As much as possible!" (Name)* replies. The staff member tells him: "The limit you can buy is 2.5 ounces. And the total limit you can legally possess is 4 ounces, so you should only buy 2.5 ounces if you don't already possess more than 1.5 ounces." (Name)* mentally computes that buying 2.5 ounces will leave him with (Difference)*** than the legal limit for possession of 4 ounces. (Name)* then buys 2.5 ounces.

How socially appropriate would most (Subjects)**** from Florida think it is for (Name)* to buy 2.5 ounces on this occasion?

***The possible wordings for Name were: DeAndre, DeShawn, Hakim, Jamal, Kareem, Malik, Marquis, Rasheed, Tremayne, Tyrone in the African American condition; Brad, Cody, Connor, Dustin, Logan, Matthew, Scott, Tanner, Todd, Wyatt in the White American condition.**

****The wording for Race was "African American" in the African American condition, and "White American" in the White American condition.**

*****The possible values for Amount were 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2 and 2.2. The possible wordings for Difference, therefore, were "0.7 ounces less", "0.5 ounces less", "0.3 ounces less", "0.1 ounces less", "0.1 ounces more", "0.3 ounces more", "0.5 ounces more" and "0.7 ounces more" (i.e. he always calculates this correctly).**

******The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

Age of Consent Vignette

Jermaine is a 24 year old African American student at Florida State University. One night, he meets (Name)*, (Race)** high school student at a party. Jermaine invites (Name)* to come to his home, and she agrees. At his home, Jermaine tells (Name)* he wants to have sex with her, but that she looks young, and asks whether she is above the legal age of consent of 18 years. (Name)* tells Jermaine that she (Age)***, and shows him an ID card which confirms this. She tells him that she wants to have sex with him. Jermaine then has sex with (Name)*.

How socially appropriate would most (Subjects)**** from Florida think it is for Jermaine to have sex with (Name)*?

***The possible wordings for Name were: Aisha, Deja, Ebony, Imani, Keisha, Kenya, Lakisha, Latonya, Tanisha and Tamika in the African American condition; Abigail, Amy, Carrie, Carly, Emily, Emma, Jill, Katie, Madeline and Molly in the White American condition.**

****The wording for Race was "an African American" in the African American condition, and "a White American" in the White American condition.**

*****The possible wordings for Age were "will turn 18 in 7 days", "will turn 18 in 5 days", "will turn 18 in 3 days", "will turn 18 in 1 day", "turned 18 1 day ago", "turned 18 3 days ago", "turned 18 5 days ago" and "turned 18 7 days ago".**

******The wording for Subjects was "MTurkers" in waves of data collection recruiting subjects via CloudResearch, and "Prolific participants" in waves of data collection recruiting subjects via Prolific.**

D Summary statistics

Table 1: Participant summary statistics main experiments

Variable	Gender	Race
female	0.53	0.55
age	34.05 (11.92)	35.85 (13.35)
US citizen	0.95	0.95
Texas resident	0.97	—
Texas born	0.63	—
Florida resident	—	0.98
Florida born	—	0.44
high income	0.22	0.21
middle income	0.30	0.30
low income	0.49	0.50
nonwhite	—	0.38
sexism index	13.79 (5.01)	—
sexist views	0.28	—
racism index	—	7.83 (2.48)
racist views	—	0.39
Cloud Research	0.60	0.57
both platforms	0.18	0.18
previous participant	0.05	0.04
four-point scale	0.41	—
N	2516	2447

Note: The table shows sample means for all participants in the Gender (left) and the Race (right) Experiment. Standard deviations are reported in parentheses for continuous variables. All other means denote the fraction of participants in the reported category. Potential repeat participants (those who explicitly indicated previous participation in the experiment and those who admitted to being registered on both platforms) were excluded from the final analyses presented in the main text, leaving N=2053 for the Gender Experiment and N=2012 for the Race Experiment.

Table 2: Distribution of appropriateness ratings in the GENDER experiment

Vignette	Gender	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Drink driving	**Female**	-7	0.528	0.106	0.244	0.098	0.008	0.016	123
		-5	0.557	0.155	0.155	0.093	0.021	0.021	97
		-3	0.429	0.206	0.23	0.087	0.024	0.024	126
		-1	0.397	0.176	0.301	0.066	0.044	0.015	136
		1	0.273	0.129	0.273	0.22	0.076	0.03	132
		3	0.216	0.144	0.309	0.137	0.094	0.101	139
		5	0.172	0.115	0.328	0.262	0.057	0.066	122
		7	0.147	0.132	0.25	0.279	0.081	0.11	136
	Male	-7	0.482	0.099	0.291	0.078	0.035	0.014	141
		-5	0.538	0.169	0.162	0.069	0.023	0.038	130
		-3	0.491	0.121	0.25	0.086	0.034	0.017	116
		-1	0.457	0.178	0.186	0.109	0.039	0.031	129
		1	0.203	0.113	0.376	0.211	0.068	0.03	133
		3	0.245	0.158	0.273	0.151	0.094	0.079	139
		5	0.181	0.118	0.283	0.26	0.087	0.071	127
		7	0.181	0.094	0.299	0.244	0.071	0.11	127

Table 2: Distribution of appropriateness ratings in the GENDER experiment

Vignette	Gender	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Speeding	**Female**	-7	0.081	0.054	0.215	0.289	0.141	0.221	149
		-5	0.05	0.025	0.207	0.207	0.198	0.314	121
		-3	0.008	0.034	0.176	0.21	0.168	0.403	119
		-1	0.033	0.017	0.167	0.158	0.108	0.517	120
		1	0.008	0.008	0.071	0.103	0.183	0.627	126
		3	0.036	0.036	0.045	0.152	0.205	0.527	112
		5	0.063	0.014	0.147	0.189	0.175	0.413	143
		7	0.064	0.079	0.186	0.179	0.171	0.321	140
	Male	-7	0.085	0.031	0.225	0.233	0.147	0.279	129
		-5	0.064	0.018	0.147	0.294	0.11	0.367	109
		-3	0.008	0.03	0.144	0.227	0.189	0.402	132
		-1	0.008	0.008	0.117	0.211	0.109	0.547	128
		1	0.027	0.007	0.073	0.173	0.14	0.58	150
		3	0.025	0.033	0.125	0.125	0.142	0.55	120
		5	0.059	0.037	0.125	0.213	0.103	0.463	136
		7	0.084	0.042	0.151	0.193	0.193	0.336	119

Table 2: Distribution of appropriateness ratings in the GENDER experiment

Vignette	Gender	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Casino	**Female**	-7	0.24	0.109	0.419	0.155	0.047	0.031	129
		-5	0.258	0.084	0.413	0.174	0.058	0.013	155
		-3	0.25	0.125	0.39	0.147	0.044	0.044	136
		-1	0.197	0.165	0.299	0.22	0.079	0.039	127
		1	0.008	0.008	0.025	0.05	0.176	0.731	119
		3	0.069	0.008	0.023	0.038	0.115	0.746	130
		5	0.031	0	0.031	0.063	0.071	0.803	127
		7	0.009	0.027	0.027	0.071	0.08	0.788	113
	Male	-7	0.307	0.197	0.321	0.139	0.029	0.007	137
		-5	0.333	0.135	0.297	0.162	0.054	0.018	111
		-3	0.198	0.149	0.314	0.231	0.05	0.058	121
		-1	0.118	0.143	0.328	0.244	0.143	0.025	119
		1	0.017	0	0.025	0.017	0.117	0.825	120
		3	0.029	0.007	0.065	0.065	0.116	0.717	138
		5	0.024	0.024	0.04	0.04	0.111	0.762	126
		7	0.014	0.021	0.069	0.062	0.076	0.759	145

Table 2: Distribution of appropriateness ratings in the GENDER experiment

Vignette	Gender	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Alcohol to youth	**Female**	-7	0.589	0.097	0.218	0.081	0.008	0.008	124
		-5	0.533	0.25	0.15	0.05	0	0.017	120
		-3	0.531	0.18	0.195	0.039	0.031	0.023	128
		-1	0.462	0.114	0.28	0.068	0.053	0.023	132
		1	0.092	0.021	0.155	0.254	0.141	0.338	142
		3	0.07	0.039	0.155	0.31	0.155	0.271	129
		5	0.062	0.021	0.221	0.241	0.186	0.269	145
		7	0.089	0.024	0.154	0.309	0.179	0.244	123
	Male	-7	0.592	0.223	0.146	0.023	0.008	0.008	130
		-5	0.643	0.143	0.103	0.095	0.008	0.008	126
		-3	0.538	0.171	0.205	0.06	0.017	0.009	117
		-1	0.47	0.187	0.216	0.075	0.037	0.015	134
		1	0.047	0.062	0.209	0.248	0.132	0.302	129
		3	0.059	0.059	0.144	0.229	0.203	0.305	118
		5	0.109	0.055	0.164	0.242	0.148	0.281	128
		7	0.07	0.047	0.164	0.227	0.203	0.289	128

Table 2: Distribution of appropriateness ratings in the GENDER experiment

Vignette	Gender	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Minimum wage	**Female**	-7	0.605	0.116	0.147	0.07	0.039	0.023	129
		-5	0.471	0.134	0.185	0.101	0.034	0.076	119
		-3	0.451	0.111	0.222	0.139	0.042	0.035	144
		-1	0.454	0.115	0.269	0.085	0.031	0.046	130
		1	0.064	0.083	0.119	0.339	0.138	0.257	109
		3	0.067	0.058	0.092	0.258	0.183	0.342	120
		5	0.046	0.019	0.102	0.25	0.204	0.38	108
		7	0.063	0.024	0.056	0.23	0.175	0.452	126
	Male	-7	0.532	0.173	0.158	0.094	0.029	0.014	139
		-5	0.402	0.205	0.165	0.134	0.039	0.055	127
		-3	0.466	0.127	0.246	0.085	0.025	0.051	118
		-1	0.383	0.177	0.234	0.106	0.043	0.057	141
		1	0.05	0.057	0.199	0.284	0.184	0.227	141
		3	0.103	0.016	0.079	0.262	0.206	0.333	126
		5	0.058	0.032	0.097	0.266	0.208	0.338	154
		7	0.041	0.066	0.09	0.189	0.221	0.393	122

Table 2: Distribution of appropriateness ratings in the GENDER experiment

Vignette	Gender	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Parental leave	**Female**	-7	0.774	0.071	0.103	0.032	0.006	0.013	155
		-5	0.816	0.066	0.044	0.037	0.015	0.022	136
		-3	0.732	0.077	0.12	0.042	0	0.028	142
		-1	0.72	0.112	0.064	0.048	0.008	0.048	125
		1	0.143	0.023	0.158	0.278	0.135	0.263	133
		3	0.124	0.022	0.146	0.212	0.19	0.307	137
		5	0.119	0.067	0.133	0.193	0.148	0.341	135
		7	0.091	0.036	0.136	0.227	0.118	0.391	110
	Male	-7	0.717	0.088	0.088	0.035	0.035	0.035	113
		-5	0.602	0.133	0.124	0.071	0.009	0.062	113
		-3	0.642	0.15	0.125	0.033	0.025	0.025	120
		-1	0.544	0.184	0.132	0.053	0.026	0.061	114
		1	0.079	0.064	0.121	0.214	0.179	0.343	140
		3	0.055	0.073	0.128	0.239	0.156	0.349	109
		5	0.113	0.04	0.079	0.179	0.179	0.411	151
		7	0.075	0.042	0.075	0.2	0.167	0.442	120

Note: The table displays the percentages of subjects, by sample and treatment, who chose each evaluation in a given vignette. Modal evaluations are shaded. To abbreviate column headers, the asterisk * is used as a placeholder for the word ‘socially’, which was used throughout.

Table 3: Distribution of appropriateness ratings in the RACE experiment

Vignette	Race	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Drink driving	**Black American**	-7	0.446	0.281	0.124	0.074	0.041	0.033	121
		-5	0.381	0.283	0.15	0.106	0.071	0.009	113
		-3	0.449	0.265	0.176	0.051	0.037	0.022	136
		-1	0.308	0.283	0.2	0.117	0.067	0.025	120
		1	0.196	0.224	0.29	0.14	0.112	0.037	107
		3	0.222	0.2	0.215	0.119	0.163	0.081	135
		5	0.183	0.217	0.217	0.208	0.133	0.042	120
		7	0.188	0.195	0.211	0.158	0.195	0.053	133
	White American	-7	0.364	0.306	0.231	0.05	0.033	0.017	121
		-5	0.388	0.364	0.085	0.124	0.023	0.016	129
		-3	0.363	0.339	0.161	0.089	0.032	0.016	124
		-1	0.303	0.333	0.197	0.083	0.053	0.03	132
		1	0.199	0.375	0.154	0.14	0.096	0.037	136
		3	0.137	0.23	0.281	0.209	0.101	0.043	139
		5	0.165	0.174	0.217	0.2	0.13	0.113	115
		7	0.113	0.188	0.233	0.226	0.158	0.083	133

Table 3: Distribution of appropriateness ratings in the RACE experiment

Vignette	Race	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Gun	**Black American**	-7	0.23	0.216	0.237	0.129	0.122	0.065	139
		-5	0.194	0.177	0.226	0.202	0.153	0.048	124
		-3	0.206	0.16	0.267	0.107	0.153	0.107	131
		-1	0.123	0.145	0.159	0.138	0.304	0.13	138
		1	0.085	0.14	0.155	0.132	0.302	0.186	129
		3	0.113	0.113	0.15	0.188	0.286	0.15	133
		5	0.134	0.063	0.165	0.173	0.339	0.126	127
		7	0.113	0.105	0.12	0.203	0.286	0.173	133
	White American	-7	0.238	0.131	0.262	0.172	0.139	0.057	122
		-5	0.222	0.262	0.206	0.159	0.119	0.032	126
		-3	0.185	0.213	0.241	0.148	0.139	0.074	108
		-1	0.161	0.153	0.258	0.202	0.137	0.089	124
		1	0.057	0.115	0.123	0.23	0.328	0.148	122
		3	0.203	0	0.148	0.18	0.289	0.18	128
		5	0.118	0.101	0.143	0.185	0.286	0.168	119
		7	0.162	0.135	0.135	0.189	0.27	0.108	111

Table 3: Distribution of appropriateness ratings in the RACE experiment

Vignette	Race	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Marijuana	**Black American**	-7	0.089	0.185	0.341	0.104	0.193	0.089	135
		-5	0.067	0.178	0.259	0.281	0.126	0.089	135
		-3	0.041	0.18	0.32	0.18	0.213	0.066	122
		-1	0.015	0.091	0.273	0.227	0.235	0.159	132
		1	0.033	0.025	0.041	0.066	0.311	0.525	122
		3	0.01	0.029	0.068	0.126	0.33	0.437	103
		5	0.014	0.047	0.054	0.081	0.331	0.473	148
		7	0.016	0.008	0.057	0.081	0.285	0.553	123
	White American	-7	0.097	0.195	0.221	0.221	0.204	0.062	113
		-5	0.066	0.161	0.299	0.226	0.168	0.08	137
		-3	0.048	0.097	0.234	0.298	0.21	0.113	124
		-1	0.063	0.117	0.279	0.144	0.306	0.09	111
		1	0.016	0.008	0.023	0.101	0.31	0.543	129
		3	0.016	0.039	0.023	0.085	0.341	0.496	129
		5	0.009	0.035	0.061	0.105	0.263	0.526	114
		7	0.029	0.015	0.088	0.066	0.299	0.504	137

Table 3: Distribution of appropriateness ratings in the RACE experiment

Vignette	Race	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Speeding	**Black American**	-7	0.076	0.076	0.136	0.182	0.303	0.227	132
		-5	0.047	0.062	0.14	0.132	0.318	0.302	129
		-3	0.009	0.019	0.093	0.13	0.407	0.343	108
		-1	0.007	0.03	0.059	0.156	0.274	0.474	135
		1	0.024	0.039	0.063	0.071	0.283	0.52	127
		3	0.041	0.041	0.103	0.069	0.283	0.462	145
		5	0.02	0.108	0.098	0.147	0.284	0.343	102
		7	0.078	0.112	0.138	0.155	0.25	0.267	116
	White American	-7	0.056	0.089	0.202	0.218	0.242	0.194	124
		-5	0.008	0.039	0.165	0.205	0.315	0.268	127
		-3	0.025	0.025	0.068	0.203	0.305	0.373	118
		-1	0.008	0.023	0.061	0.092	0.267	0.55	131
		1	0	0.066	0.049	0.123	0.238	0.525	122
		3	0.029	0.022	0.087	0.13	0.275	0.457	138
		5	0.03	0.097	0.097	0.142	0.284	0.351	134
		7	0.063	0.095	0.19	0.127	0.222	0.302	126

Table 3: Distribution of appropriateness ratings in the RACE experiment

Vignette	Race	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Age of consent	**Black American**	-7	0.244	0.275	0.26	0.107	0.084	0.031	131
		-5	0.268	0.26	0.228	0.106	0.081	0.057	123
		-3	0.203	0.266	0.227	0.148	0.109	0.047	128
		-1	0.216	0.269	0.157	0.142	0.164	0.052	134
		1	0.031	0.047	0.118	0.26	0.339	0.205	127
		3	0.057	0.075	0.16	0.094	0.349	0.264	106
		5	0.031	0.054	0.101	0.209	0.341	0.264	129
		7	0.049	0.041	0.041	0.197	0.377	0.295	122
	White American	-7	0.296	0.306	0.157	0.148	0.074	0.019	108
		-5	0.342	0.283	0.167	0.1	0.083	0.025	120
		-3	0.319	0.277	0.177	0.099	0.085	0.043	141
		-1	0.225	0.239	0.232	0.148	0.134	0.021	142
		1	0.018	0.083	0.138	0.193	0.349	0.22	109
		3	0.054	0.061	0.122	0.15	0.388	0.224	147
		5	0.015	0.083	0.121	0.114	0.371	0.295	132
		7	0.087	0.017	0.122	0.13	0.435	0.209	115

Table 3: Distribution of appropriateness ratings in the RACE experiment

Vignette	Race	Distance to threshold	Very* inappropriate	* inappropriate	Somewhat * inappropriate	Somewhat * appropriate	* appropriate	Very * appropriate	Total
Minimum wage	**Black American**	-7	0.468	0.23	0.127	0.071	0.087	0.016	126
		-5	0.408	0.246	0.108	0.108	0.085	0.046	130
		-3	0.36	0.28	0.153	0.093	0.093	0.02	150
		-1	0.402	0.174	0.174	0.121	0.083	0.045	132
		1	0.044	0.053	0.193	0.263	0.289	0.158	114
		3	0.045	0.114	0.129	0.212	0.265	0.235	132
		5	0.07	0.023	0.078	0.156	0.43	0.242	128
		7	0.031	0.047	0.055	0.195	0.359	0.312	128
	White American	-7	0.496	0.248	0.099	0.066	0.074	0.017	121
		-5	0.504	0.282	0.084	0.084	0.038	0.008	131
		-3	0.479	0.218	0.143	0.059	0.059	0.042	119
		-1	0.312	0.319	0.17	0.043	0.099	0.057	141
		1	0.071	0.089	0.125	0.223	0.33	0.161	112
		3	0.043	0.034	0.095	0.216	0.405	0.207	116
		5	0	0.035	0.062	0.257	0.398	0.248	113
		7	0.041	0.066	0.05	0.149	0.405	0.289	121

Note: The table displays the percentages of subjects, by sample and treatment, who chose each evaluation in a given vignette. Modal evaluations are shaded. To abbreviate column headers, the asterisk * is used as a placeholder for the word ‘socially,’ which was used throughout.

E Race Experiment Pilots

In the Race Experiment, we use first names to convey to participants the race/ethnicity of the person described in the vignette. We selected the names of the African American and White American protagonists by running two pilots in April 2020, prior to the launch of the Race Experiment. The pilots were distributed on MTurk via CloudResearch, with participation only eligible to respondents from Florida.

In a first pilot, we pretested 20 male and 20 female names that we expected to be associated with African American people and 20 male and 20 female names that we expected to be associated with White American people, based on previous literature (Bertrand and Mullainathan, 2004). We ran a survey in which 105 respondents were presented with all of these names in random order; half of the respondents were asked to rank from 1-10 the likelihood that a person with each name would be African American, while the other half were asked to rank in the same way the likelihood the person would be White American.

We deduced the most strongly African / White American-associated names for each gender by calculating the average score for the likelihood of the person being White American divided by the average score for the likelihood of the person being African American. Table 4 shows the results.

We then selected the ten most strongly African American-associated and White American-associated names for each gender for use in our experiment, with the exception that we excluded some high-scoring names with close phonetic similarities to others selected for use.

Because names might convey not only race but also socioeconomic status, we ran a second pilot to measure perceptions of socioeconomic status associated with each of the names chosen. Specifically, we presented 123 subjects with, at random, either the 20 African American or 20 White American names selected from the first pilot. For each name, subjects were asked to score the likely socioeconomic status of a person who had this name and was confirmed to be of the typical race and gender associated with the name, relative to an average person of the same race and gender. Responses were given on a scale of 0-10, with 0 indicating that a typical person with this name would definitely be of lower socioeconomic status than the average for their race and gender, 10 indicating that they would definitely be of higher status than average, and five indicating that they would be equal to the average.

The results, displayed in Table 4, showed that all of the White American names selected were associated with high socioeconomic status relative to an average White American, while most of the African American names were associated with low socioeconomic status relative to an average African American. This is an unsurprising consequence of using stereotypical names. More importantly, however, we found substantial variation in socioeconomic status within the sets of names for each race and gender, which allow us to account for the effect of socioeconomic status in our main experiment, as described in the main text.

Table 4: Perceived 'whiteness' and socioeconomic status of names

Name	Female	AfricanAmer	WhiteAmer	Whitename	SES
Rasheed	FALSE	7.84 (2.35)	1.74 (2.65)	0.22	4.80 (2.11)
Tremayne	FALSE	8.39 (1.65)	1.92 (2.83)	0.23	4.33 (2.22)
DeShawn	FALSE	8.55 (2.29)	2.04 (3.03)	0.24	4.78 (2.46)
Kareem	FALSE	8.06 (2.23)	1.94 (3.02)	0.24	4.85 (2.06)
Hakim	FALSE	7.67 (2.05)	1.89 (2.82)	0.25	4.70 (2.19)
DeAndre	FALSE	8.69 (1.71)	2.25 (2.97)	0.26	4.93 (2.16)
Jamal	FALSE	8.24 (2.15)	2.36 (3.08)	0.29	4.78 (2.03)
Marquis	FALSE	8.51 (1.75)	2.47 (3.00)	0.29	5.70 (2.21)
Tyrone	FALSE	8.73 (1.40)	2.64 (3.17)	0.30	4.43 (2.44)
Malik	FALSE	7.69 (2.17)	2.42 (3.15)	0.31	5.22 (2.04)
Jermaine	FALSE	8.31 (1.76)	2.64 (3.06)	0.32	–
Trevon	FALSE	8.16 (1.93)	2.60 (2.98)	0.32	–
Terrell	FALSE	7.92 (2.20)	2.72 (3.00)	0.34	–
Demetrius	FALSE	7.39 (2.62)	2.81 (3.16)	0.38	–
Darnell	FALSE	8.00 (1.92)	3.13 (3.17)	0.39	–
Dominique	FALSE	7.25 (2.36)	3.23 (3.29)	0.45	–
Leroy	FALSE	6.96 (2.35)	3.30 (3.15)	0.47	–
Maurice	FALSE	6.63 (2.29)	4.13 (3.52)	0.62	–
Reginald	FALSE	5.31 (2.60)	4.79 (3.11)	0.90	–
Willie	FALSE	5.22 (2.34)	5.85 (3.12)	1.12	–
Jay	FALSE	4.10 (2.59)	6.83 (2.61)	1.67	–
Cole	FALSE	3.82 (2.48)	7.23 (2.84)	1.89	–
Geoffrey	FALSE	3.76 (2.63)	7.62 (2.37)	2.03	–

Jack	FALSE	3.61 (2.71)	8.25 (2.38)	2.29	–
Luke	FALSE	3.55 (2.71)	8.17 (2.13)	2.30	–
Greg	FALSE	3.37 (2.35)	7.96 (2.37)	2.36	–
Brendan	FALSE	3.31 (2.53)	7.92 (2.25)	2.39	–
Neil	FALSE	3.24 (2.71)	7.77 (2.41)	2.40	–
Jake	FALSE	3.37 (2.55)	8.25 (2.21)	2.45	–
Wyatt	FALSE	3.18 (2.73)	7.92 (2.18)	2.49	6.11 (2.43)
Matthew	FALSE	3.20 (2.65)	8.00 (2.43)	2.50	6.60 (1.53)
Logan	FALSE	3.06 (2.49)	7.83 (2.61)	2.56	6.81 (1.88)
Todd	FALSE	3.25 (3.19)	8.45 (2.22)	2.60	6.05 (2.04)
Dustin	FALSE	2.94 (2.57)	7.72 (2.44)	2.63	5.86 (2.15)
Tanner	FALSE	2.94 (2.56)	8.09 (2.44)	2.75	6.47 (2.11)
Cody	FALSE	2.86 (2.71)	8.28 (2.16)	2.90	5.18 (2.03)
Scott	FALSE	2.84 (2.79)	8.34 (2.35)	2.94	6.68 (1.87)
Connor	FALSE	2.84 (2.48)	8.43 (2.08)	2.97	6.95 (1.81)
Brett	FALSE	2.55 (2.39)	8.17 (2.26)	3.20	–
Brad	FALSE	2.41 (2.67)	8.40 (2.28)	3.49	6.11 (1.93)
Lakisha	TRUE	9.00 (1.66)	1.87 (2.71)	0.21	4.50 (2.30)
Tamika	TRUE	8.76 (1.84)	1.91 (2.71)	0.22	4.61 (2.53)
Latonya	TRUE	8.57 (2.37)	2.11 (2.90)	0.25	4.57 (2.18)
Latoya	TRUE	8.49 (2.40)	2.11 (2.90)	0.25	–
Tanisha	TRUE	8.57 (1.71)	2.15 (2.99)	0.25	4.81 (2.43)
Ebony	TRUE	8.37 (2.51)	2.21 (3.14)	0.26	4.94 (2.26)
Kenya	TRUE	8.20 (2.03)	2.17 (2.97)	0.26	5.06 (2.06)
Deja	TRUE	7.69 (2.12)	2.21 (2.59)	0.29	4.76 (2.30)
Keisha	TRUE	8.45 (1.54)	2.57 (3.16)	0.30	5.07 (2.29)

Aisha	TRUE	7.63 (2.21)	2.43 (2.89)	0.32	5.11 (2.46)
Imani	TRUE	7.43 (2.38)	2.36 (2.82)	0.32	5.33 (2.54)
Shanice	TRUE	8.65 (1.72)	2.79 (3.36)	0.32	–
Aaliyah	TRUE	7.98 (2.39)	2.64 (2.86)	0.33	–
Precious	TRUE	7.86 (1.96)	3.09 (3.57)	0.39	–
Diamond	TRUE	7.33 (2.36)	2.94 (3.18)	0.40	–
Aliyah	TRUE	7.25 (2.35)	3.32 (2.99)	0.46	–
Asia	TRUE	6.16 (2.44)	2.87 (2.79)	0.47	–
Jada	TRUE	6.67 (2.61)	3.40 (3.01)	0.51	–
Nia	TRUE	6.35 (2.22)	3.26 (2.77)	0.51	–
Tierra	TRUE	7.24 (1.81)	3.91 (3.23)	0.54	–
Jenna	TRUE	3.59 (2.62)	7.81 (2.21)	2.18	–
Kristen	TRUE	3.18 (2.71)	8.04 (2.30)	2.53	–
Katelyn	TRUE	3.06 (2.46)	8.00 (2.41)	2.61	–
Anne	TRUE	2.92 (2.90)	7.70 (2.61)	2.64	–
Allison	TRUE	2.94 (2.60)	7.79 (2.75)	2.65	–
Sarah	TRUE	2.96 (2.43)	7.87 (2.60)	2.66	–
Claire	TRUE	3.08 (2.75)	8.21 (2.42)	2.67	–
Laurie	TRUE	2.96 (2.53)	7.89 (2.52)	2.67	–
Heather	TRUE	3.02 (2.82)	8.23 (2.33)	2.73	–
Abigail	TRUE	2.67 (2.36)	7.32 (2.94)	2.74	6.98 (2.02)
Meredith	TRUE	2.90 (2.70)	7.94 (2.48)	2.74	–
Katie	TRUE	2.90 (2.33)	8.25 (2.50)	2.84	6.09 (2.09)
Jill	TRUE	2.75 (2.67)	7.94 (2.66)	2.89	6.21 (2.19)
Carly	TRUE	2.69 (2.31)	7.81 (2.30)	2.90	5.93 (2.01)
Emily	TRUE	2.76 (2.42)	8.28 (2.26)	3.00	6.98 (1.68)

Carrie	TRUE	2.76 (2.44)	8.32 (2.15)	3.01	6.14 (2.12)
Amy	TRUE	2.67 (2.50)	8.08 (2.52)	3.03	5.91 (2.22)
Emma	TRUE	2.69 (2.62)	8.19 (2.30)	3.04	6.77 (1.70)
Madeline	TRUE	2.59 (2.41)	8.28 (2.25)	3.20	7.42 (1.58)
Molly	TRUE	2.31 (2.16)	8.15 (2.48)	3.53	5.82 (2.28)

Note: The table shows sample means from two pilots that we ran prior to the Race Experiment (as detailed in Section E), with standard deviations in parentheses. In Pilot 1, we asked half the participants (at random) to evaluate (on a scale from 0-10) the likelihood that a person of a given name is African American or, the other half, White American. Low (high) values indicate that, on average, a name was considered unlikely (likely) to belong to a person of that race. From this, we calculate the variable White Name as the ratio of African American and White American, a measure that summarizes the relative "Whiteness" of a name. In Pilot 2, we chose a subset of the most and least White names and asked subjects to score (on a scale from 0-10) the likely socioeconomic status of a person with this name relative to an average person of the same race and gender. Values lower (higher) than 5 indicate that a typical person with this name is, on average, considered of lower (higher) socioeconomic status than the average person for their race and gender, with five indicating that they would be equal to the average.

F Additional results and robustness

F.1 Gender Experiment

This subsection reports results from robustness checks and heterogeneity analyses for the Gender Experiment. Table 5 supplies the full output, including all controls for the regressions presented in the main text (Section 4.1). Table 6 shows that the results barely change if we drop all controls from the analyses. Coefficients obtained from an Ordered Logit Model rather than OLS are shown in Table 7; the results are qualitatively very similar. Finally, Table 8 shows OLS results but with flexible controls for distance to the threshold. Notably, the interaction between *male* and *illegal* remains insignificant throughout.

We conducted further tests to probe the robustness of our findings (results available upon request). We investigated whether there are any systematic differences between (i) the CloudResearch and Prolific sample, (ii) vignettes that had been randomly assigned earlier during the experiment or later (to check for order effects), or (iii) respondents that were presented with the four-point vs. the six-point scale. We ran separate regressions for each subgroup in each case and compared our coefficient of interest (β_3 on $male \times illegal$) across regressions using z-tests. We found no support for any statistical differences across subgroups.

We also checked whether our results were sensitive to how we coded appropriateness in our 4-point scale sample. We reran our estimations, assigning a value of $(-).33$ rather than $(-).2$ to the middle categories ("Somewhat socially (in-) appropriate") and compared our coefficient of interest to those obtained from our main analyses using z-tests. Again, we found no indication that the coding matters for our results. The same is true when we reran our analyses excluding respondents who self-report not currently residing in Texas ($N = 44$), despite the filter condition we set on CloudResearch and Prolific. Our results remain virtually unchanged if we exclude those respondents.

Finally, we performed several exploratory heterogeneity analyses using a similar approach: we split the sample by a variable of interest and ran separate regressions for these subgroups, then compared our coefficient of interest across subgroups using z-tests. In doing so, we can study whether the gender difference in the discontinuity of the norm functions at the threshold differs across different subgroups of respondents. In particular, we compared respondents born in Texas to those not, female to male respondents, and respondents who self-reported above-median levels of agreement with sexist statements to those who did not. None of these comparisons turned up statistically significant differences.

Table 5: Social appropriateness in the Gender experiment, full output

	Drink driving	Speeding	Casino	Alcohol to youth	Minimum wage	Parental leave
male==1 (M)	.05	−.09	.05	−.04	.02	.13*
	(.07)	(.07)	(.06)	(.07)	(.07)	(.07)
illegal==1 (I)	−.16**	−.18**	−1.03***	−.87***	−.67***	−.93***
	(.07)	(.07)	(.06)	(.07)	(.07)	(.07)
M x I	−.11	.15	.13	.00	−.01	.00
	(.10)	(.10)	(.09)	(.10)	(.10)	(.11)
absolute distance (AD)	.03***	−.07***	.00	−.01	.04***	.02*
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
M x AD	−.01	.02	−.01	.01	−.01	−.01
	(.02)	(.02)	(.01)	(.01)	(.02)	(.02)
I x AD	−.06***	.01	−.01	−.02	−.07***	−.03**
	(.02)	(.02)	(.01)	(.01)	(.02)	(.02)
M x I x AD	.03	−.02	−.03*	−.01	.02	.00
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Vignette position	.00	−.00	−.00	.01*	.00	.01**
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
4-pt scale	−.07*	.10***	−.00	−.11***	−.11***	−.02
	(.04)	(.04)	(.03)	(.03)	(.04)	(.04)
Cloudresearch == 1	.09**	−.06	−.06*	.07*	.07*	.01
	(.04)	(.04)	(.03)	(.04)	(.04)	(.04)
S age	−.00**	.00*	−.00***	−.00	.00***	.00
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
S female == 1	.03	.02	−.01	−.06**	−.00	−.01
	(.03)	(.03)	(.02)	(.02)	(.03)	(.03)
S US citizen == 1	−.13**	.19***	.08	.06	.04	−.09
	(.06)	(.06)	(.05)	(.06)	(.06)	(.06)
S Texas born == 1	.03	−.05**	−.01	.00	.01	.02
	(.03)	(.03)	(.02)	(.03)	(.03)	(.03)
S middle income == 1	.01	−.04	.03	.04	.03	.05*
	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)
S high income == 1	.05	.00	.03	−.03	.07**	−.01
	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)
S sexist views == 1	.13***	−.07***	−.04	−.05*	.16***	.07**
	(.03)	(.03)	(.02)	(.03)	(.03)	(.03)
Intercept	−.27***	.63***	.90***	.31***	−.04	.14
	(.08)	(.09)	(.08)	(.08)	(.09)	(.09)
Adj. R^2	.13	.07	.57	.44	.41	.47
Num. obs.	2053	2053	2053	2053	2053	2053

Note: Coefficients estimated from OLS models, including the full set of controls as reported. The prefix 'S' in variable names indicates that the variable controls for characteristics of the respondent, not the vignette. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 6: Social appropriateness in the Gender experiment, w/o controls

	Drink driving	Speeding	Casino	Alcohol to youth	Minimum wage	Parental leave
white==1 (W), β_1	.10	$-.01$	$-.09$	$-.09$	$-.05$	$-.04$
	(.07)	(.07)	(.08)	(.07)	(.07)	(.07)
illegal==1 (I), β_2	$-.12^{*}$	.05	$-.37^{***}$	$-.55^{***}$	$-.68^{***}$	$-.67^{***}$
	(.07)	(.07)	(.08)	(.07)	(.07)	(.07)
W x I, β_3	$-.08$	$-.06$	$.30^{***}$	.13	.07	.13
	(.10)	(.10)	(.11)	(.09)	(.10)	(.10)
absolute distance (AD), β_4	$.05^{***}$	$-.06^{***}$	$-.03^{**}$	$-.01$	$.04^{***}$	.00
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
W x AD, β_5	$-.03^{**}$	$-.00$	.03	.02	.00	.01
	(.02)	(.02)	(.02)	(.01)	(.02)	(.02)
I x AD, β_6	$-.06^{***}$	$-.02$	.01	$-.02$	$-.07^{***}$	$-.03^{*}$
	(.02)	(.02)	(.02)	(.01)	(.02)	(.02)
W x I x AD, β_7	.03	.02	$-.06^{**}$	$-.03$	.01	$-.01$
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Intercept	$-.35^{***}$	$.73^{***}$	$.27^{***}$	$.73^{***}$	$.26^{***}$	$.37^{***}$
	(.05)	(.05)	(.06)	(.05)	(.05)	(.05)
Adj. R^2	.10	.06	.07	.28	.41	.31
Num. obs.	2053	2053	2053	2053	2053	2053

Note: Coefficients estimated from OLS models without controls. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 7: Social appropriateness in the Gender experiment (Ordered Logit)

	Drink driving	Speeding	Casino	Alcohol to youth	Minimum wage	Parental leave
male==1 (M), β_1	.17	−.31	.37	−.16	.05	.32
	(.22)	(.25)	(.30)	(.23)	(.23)	(.22)
illegal==1 (I), β_2	$-.51^{**}$	$-.66^{***}$	-3.73^{***}	-2.73^{***}	-1.92^{***}	-2.83^{***}
	(.23)	(.25)	(.28)	(.24)	(.24)	(.26)
M x I, β_3	−.48	.45	.19	.08	−.03	.36
	(.32)	(.35)	(.38)	(.33)	(.33)	(.35)
absolute distance (AD), β_4	$.11^{***}$	$-.24^{***}$	.05	−.03	$.15^{***}$	$.06^{*}$
	(.03)	(.04)	(.05)	(.03)	(.04)	(.04)
M x AD, β_5	−.04	.07	−.10	.03	−.05	−.01
	(.05)	(.05)	(.07)	(.05)	(.05)	(.05)
I x AD, β_6	$-.20^{***}$	.05	−.08	−.07	$-.25^{***}$	$-.12^{**}$
	(.05)	(.05)	(.06)	(.05)	(.05)	(.06)
M x I x AD, β_7	.11	−.08	−.04	−.04	.07	−.01
	(.07)	(.07)	(.08)	(.07)	(.07)	(.08)
AIC	6245.95	5989.33	5073.40	5952.91	6154.40	5460.51
BIC	6369.74	6113.13	5197.19	6076.70	6278.20	5584.30
Log Likelihood	−3100.97	−2972.67	−2514.70	−2954.45	−3055.20	−2708.25
Deviance	6201.95	5945.33	5029.40	5908.91	6110.40	5416.51
Num. obs.	2053	2053	2053	2053	2053	2053

Note: Coefficients estimated from Ordered Logit models, including the full set of controls as reported in Table 5. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 8: Social appropriateness in the Gender experiment, OLS with flexible controls

	Drink driving	Speeding	Casino	Alcohol to youth	Minimum wage	Parental leave
male==1 (M)	.07	−.08	.05	−.04	.02	.14**
	(.07)	(.07)	(.06)	(.07)	(.07)	(.07)
illegal==1 (I)	−.20***	−.19***	−1.06***	−.86***	−.75***	−.96***
	(.07)	(.07)	(.06)	(.06)	(.07)	(.07)
absolute distance==3 (AD3)	.11	−.11	−.09	−.04	.13*	.08
	(.07)	(.07)	(.06)	(.07)	(.08)	(.07)
absolute distance==5 (AD5)	.13*	−.27***	−.01	−.02	.20***	.07
	(.07)	(.07)	(.06)	(.06)	(.08)	(.07)
absolute distance==7 (AD7)	.22***	−.41***	−.01	−.06	.28***	.15**
	(.07)	(.07)	(.06)	(.07)	(.07)	(.07)
M x I	−.09	.16	.07	−.01	.02	.01
	(.10)	(.10)	(.09)	(.09)	(.10)	(.10)
M x AD3	−.11	.06	−.04	.08	−.03	−.07
	(.09)	(.10)	(.09)	(.09)	(.10)	(.10)
M x AD5	−.03	.09	−.08	−.04	−.09	−.02
	(.10)	(.09)	(.09)	(.09)	(.10)	(.10)
M x AD7	−.10	.12	−.09	.08	−.07	−.05
	(.10)	(.10)	(.09)	(.09)	(.10)	(.10)
I x AD3	−.15	.06	.02	−.07	−.13	−.10
	(.09)	(.10)	(.09)	(.09)	(.10)	(.10)
I x AD5	−.27***	.10	−.08	−.16*	−.19*	−.14
	(.10)	(.10)	(.08)	(.09)	(.11)	(.10)
I x AD7	−.35***	.07	−.06	−.09	−.44***	−.22**
	(.10)	(.10)	(.09)	(.09)	(.10)	(.10)
M x I x AD3	.11	−.12	−.01	−.05	−.03	−.02
	(.14)	(.14)	(.12)	(.13)	(.14)	(.15)
M x I x AD5	.08	−.15	−.12	.07	.07	.07
	(.14)	(.14)	(.12)	(.13)	(.14)	(.14)
M x I x AD7	.19	−.14	−.18	−.10	.07	.00
	(.14)	(.14)	(.12)	(.13)	(.14)	(.15)
Intercept	−.25***	.56***	.93***	.31***	−.02	.16*
	(.08)	(.09)	(.08)	(.08)	(.09)	(.09)
Adj. R^2	.13	.07	.57	.44	.41	.47
Num. obs.	2053	2053	2053	2053	2053	2053

Note: Coefficients estimated from OLS models including flexible controls for the absolute distance to the threshold, including the full set of controls as reported in Table 5. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

F.2 Race Experiment

As for the Gender Experiment, we performed several robustness checks and heterogeneity analyses for the Race Experiment. We begin by supplying the full output in Table 9, including all controls for the regressions presented in the main text (Section 4.2). Next, we show in Table 10 that the results barely change when dropping all controls from the analyses. Estimated coefficients from an Ordered Logit Model rather than OLS are shown in Table 11, demonstrating that the main result is unaffected by the model choice. Finally, Table 12 shows OLS results but with flexible controls for distance to the threshold. Using this specification, the estimated coefficient on the interaction of $white \times illegal$ loses precision in the Gun Possession Vignette. The p-value decreases from $p = .010$ to $p = .039$ ($p = .174$ after Benjamini-Hochberg correction for multiple hypotheses testing). The coefficient becomes marginally significant at the 10-percent level in the Marijuana Vignette, with $p = .058$ ($p = .174$ after Benjamini-Hochberg correction).

As in the Gender Experiment, we conducted further tests to probe the robustness of our findings (results available upon request). We investigated whether there are any systematic differences in the results across (i) the CloudResearch and Prolific sample, (ii) vignettes that had been randomly assigned earlier during the experiment or later (to check for order effects), or (iii) the full sample or the restricted sample excluding respondents who self-report not currently residing in Florida ($N = 33$), despite the filter condition we set on CloudResearch and Prolific.

Lastly, we also performed several exploratory heterogeneity analyses by running separate regressions for subgroups, for whom we then compared our coefficient of interest using z-tests. For the Race Experiment, we compared (iv) respondents who were born in Florida to those who were not, (v) White to Non-white respondents, and (vi) respondents who self-reported above-median levels of agreement with racist statements from the questionnaire to those who did not. Finally, we examined (vii) the impact of socio-economic status associated with different names. This is somewhat challenging given that, in our sample, White American names are associated with a relative socio-economic status of at least 5.18 and African American names with a relative socio-economic status of at most 5.70. We constructed two samples in which we grouped together (1) below-median SES white above-median SES black names vs. (2) above-median white below-median black names. In sample (1), White and African American names are more similar than in the full sample, whereas in sample (2), the differences are more extreme. If the results differ markedly across these two samples, this would indicate that socioeconomic status is an important driver of our main results.

While we found no general pattern of systematic differences across these robustness and heterogeneity analyses, a small number of these numerous comparisons turned up a statistically significant z-test. The tests indicate a difference in the discontinuity at the threshold for (i) respondents born in Florida vs. respondents born elsewhere in the Drink-Driving Vignette ($p =$ .030); (ii) non-white vs. white respondents in the Gun Possession Vignette ($p = .048$); (iii) respondents evaluating the vignette at an earlier point in the experiment vs. later in the Speeding Vignette ($p = .030$). However, adjusting for 42 comparisons[3] using the Benjamini-Hochberg correction renders these differences insignificant, with (i) $p = .209$, (ii) $p = .338$, and (iii) $p = .207$.

[3]Comparing six regressions (drink driving, speeding, gun, marijuana, minimum wage, age of consent) across seven pairs of groups (CloudResearch vs. Prolific, earlier vs. later vignettes, including vs. excluding non-residents, born in Florida vs. not, White vs. Non-white, above-median racism vs. not, socioeconomic status samples (1) vs. (2).

Table 9: Social appropriateness in the RACE experiment, full output

	Drink driving	Speeding	Gun	Marijuana	Minimum wage	Age of consent
white==1 (W)	.29	−.15	−.04	.01	−.32	.06
	(.25)	(.25)	(.20)	(.16)	(.24)	(.25)
illegal==1 (I)	−.10	.05	−.37***	−.55***	−.69***	−.68***
	(.07)	(.07)	(.08)	(.07)	(.07)	(.07)
W x I	−.09	−.07	.29***	.14	.11	.14
	(.10)	(.10)	(.11)	(.09)	(.10)	(.10)
absolute_distance (AD)	.05***	−.06***	−.03**	−.01	.03***	.00
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
W x AD	−.03*	−.00	.03	.02	.01	.02
	(.02)	(.02)	(.02)	(.01)	(.02)	(.02)
I x AD	−.06***	−.02	.01	−.02	−.06***	−.03
	(.02)	(.02)	(.02)	(.01)	(.02)	(.02)
W x I x AD	.03	.02	−.06**	−.03	.00	−.02
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Vignette position	.01*	.00	.00	−.00	.00	.01**
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
SES	−.01	−.01	−.07**	−.05**	.04	−.05*
	(.03)	(.03)	(.03)	(.02)	(.03)	(.03)
Whiteness of name	−.06	.06	.02	−.01	.07	−.01
	(.09)	(.09)	(.07)	(.05)	(.08)	(.08)
Cloudresearch == 1	.03	.02	−.00	.02	−.03	.03
	(.03)	(.03)	(.03)	(.02)	(.03)	(.03)
S age	−.00	.00	−.00	−.00	.00***	−.00
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
S nonwhite == 1	.04	−.00	−.07**	−.03	.07**	−.08***
	(.03)	(.03)	(.03)	(.02)	(.03)	(.03)
S female == 1	−.10***	.01	−.05*	.02	.01	−.05**
	(.02)	(.02)	(.03)	(.02)	(.02)	(.03)
S US citizen == 1	−.03	−.01	.14**	.13***	.01	−.06
	(.05)	(.05)	(.06)	(.05)	(.05)	(.05)
S Florida born == 1	.03	−.05**	.07**	−.03	.00	.01
	(.03)	(.03)	(.03)	(.02)	(.03)	(.03)
S middle income == 1	.05	.03	.03	.00	.04	.03
	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)
S high income == 1	.11***	−.02	.03	.03	.14***	.04
	(.03)	(.03)	(.04)	(.03)	(.03)	(.03)
S racist views == 1	.09***	−.05**	−.00	−.06***	.07***	.02
	(.03)	(.03)	(.03)	(.02)	(.03)	(.03)
Intercept	−.34**	.76***	.49***	.89***	−.15	.69***
	(.17)	(.17)	(.18)	(.15)	(.17)	(.18)
Adj. R^2	.12	.06	.08	.28	.42	.32
Num. obs.	2014	2014	2014	2014	2014	2014

Note: Coefficients estimated from OLS models, including the full set of controls as reported. The prefix 'S' in variable names indicates that the variable controls for characteristics of the respondent, not the vignette. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 10: Social appropriateness in the RACE experiment, w/o controlls

	Drink driving	Speeding	Gun	Marijuana	Minimum wage	Age of consent
white==1 (W), β_1	.10	$-.01$	$-.09$	$-.09$	$-.05$	$-.04$
	(.07)	(.07)	(.08)	(.07)	(.07)	(.07)
illegal==1 (I), β_2	$-.12^{*}$	.05	$-.37^{***}$	$-.55^{***}$	$-.68^{***}$	$-.67^{***}$
	(.07)	(.07)	(.08)	(.07)	(.07)	(.07)
W x I, β_3	$-.08$	$-.06$	$.30^{***}$	.13	.07	.13
	(.10)	(.10)	(.11)	(.09)	(.10)	(.10)
absolute$_distance(AD)$,β_4	$.05^{***}$	$-.06^{***}$	$-.03^{**}$	$-.01$	$.04^{***}$	.00
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
W x AD, β_5	$-.03^{**}$	$-.00$	.03	.02	.00	.01
	(.02)	(.02)	(.02)	(.01)	(.02)	(.02)
I x AD, β_6	$-.06^{***}$	$-.02$	.01	$-.02$	$-.07^{***}$	$-.03^{*}$
	(.02)	(.02)	(.02)	(.01)	(.02)	(.02)
W x I x AD, β_7	.03	.02	$-.06^{**}$	$-.03$	.01	$-.01$
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Intercept	$-.35^{***}$	$.73^{***}$	$.27^{***}$	$.73^{***}$	$.26^{***}$	$.37^{***}$
	(.05)	(.05)	(.06)	(.05)	(.05)	(.05)
Adj. R^2	.10	.06	.07	.28	.41	.31
Num. obs.	2014	2014	2014	2014	2014	2014

Note: Coefficients estimated from OLS models without controls. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 11: Social appropriateness in the RACE experiment (Ordered Logit)

	Drink driving	Speeding	Gun	Marijuana	Minimum wage	Age of consent
white==1 (W), β_1	.46	$-.66$	$-.05$	.24	-1.02	.15
	(.83)	(.85)	(.55)	(.57)	(.79)	(.79)
illegal==1 (I), β_2	$-.38^{*}$	.12	-1.07^{***}	-2.07^{***}	-2.05^{***}	-2.03^{***}
	(.22)	(.24)	(.23)	(.25)	(.24)	(.23)
W x I, β_3	$-.25$	$-.38$	$.85^{***}$	.42	.36	.39
	(.33)	(.34)	(.32)	(.34)	(.33)	(.32)
absolute distance (AD), β_4	$.14^{***}$	$-.20^{***}$	$-.08^{**}$	$-.03$	$.12^{***}$	.02
	(.03)	(.04)	(.04)	(.04)	(.04)	(.04)
W x AD, β_5	$-.09^{*}$	$-.02$	.07	.06	.04	.05
	(.05)	(.05)	(.05)	(.05)	(.05)	(.05)
I x AD, β_6	$-.20^{***}$	$-.06$	.02	$-.06$	$-.23^{***}$	$-.10^{*}$
	(.05)	(.05)	(.05)	(.05)	(.05)	(.05)
W x I x AD, β_7	.06	.10	$-.17^{**}$	$-.10$	$-.01$	$-.04$
	(.07)	(.07)	(.07)	(.07)	(.07)	(.07)
AIC	6397.93	6022.20	6985.51	5911.16	6112.56	6466.96
BIC	6532.52	6156.79	7120.10	6045.75	6247.15	6601.55
Log Likelihood	-3174.97	-2987.10	-3468.76	-2931.58	-3032.28	-3209.48
Deviance	6349.93	5974.20	6937.51	5863.16	6064.56	6418.96
Num. obs.	2014	2014	2014	2014	2014	2014

Note: Coefficients estimated from Ordered Logit models, including the full set of controls as reported in Table 9. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 12: Social appropriateness in the RACE experiment, OLS with flexible controls for the absolute distance to the threshold

	Drink driving	Speeding	Gun	Marijuana	Minimum wage	Age of consent
white==1 (W)	.27	−.13	.02	.05	−.24	.09
	(.26)	(.25)	(.19)	(.16)	(.24)	(.25)
illegal==1 (I)	−.12*	.06	−.33***	−.61***	−.68***	−.67***
	(.07)	(.07)	(.08)	(.06)	(.07)	(.07)
absolute distance==3 (AD3)	.15**	−.05	−.07	−.05	.15**	−.01
	(.07)	(.07)	(.08)	(.06)	(.07)	(.07)
absolute distance==5 (AD5)	.24***	−.20***	−.07	−.07	.24***	.06
	(.07)	(.07)	(.08)	(.06)	(.07)	(.07)
absolute distance==7 (AD7)	.28***	−.34***	−.20**	−.07	.20***	−.01
	(.07)	(.07)	(.08)	(.06)	(.07)	(.07)
W x I	−.05	−.07	.23**	.17*	.01	.07
	(.10)	(.10)	(.11)	(.09)	(.10)	(.10)
W x AD3	−.07	−.04	−.00	−.01	−.12	−.00
	(.10)	(.09)	(.11)	(.09)	(.10)	(.10)
W x AD5	−.16	−.02	.02	.03	−.08	−.01
	(.10)	(.10)	(.11)	(.09)	(.10)	(.10)
W x AD7	−.17*	−.04	.17	.11	.04	.12
	(.10)	(.10)	(.11)	(.09)	(.10)	(.10)
I x AD3	−.23**	−.11	−.00	.08	−.26***	−.11
	(.09)	(.10)	(.11)	(.09)	(.10)	(.10)
I x AD5	−.35***	−.06	−.12	−.03	−.45***	−.22**
	(.10)	(.09)	(.11)	(.09)	(.10)	(.10)
I x AD7	−.38***	−.13	.10	−.08	−.36***	−.12
	(.10)	(.10)	(.11)	(.09)	(.10)	(.10)
W x I x AD3	−.00	.14	−.17	−.24*	.18	.08
	(.14)	(.14)	(.16)	(.13)	(.14)	(.14)
W x I x AD5	.18	.06	−.09	−.18	.24*	.04
	(.14)	(.14)	(.16)	(.13)	(.14)	(.14)
W x I x AD7	.13	.16	−.42***	−.23*	−.02	−.11
	(.14)	(.14)	(.16)	(.13)	(.14)	(.14)
Intercept	−.33*	.68***	.46**	.91***	−.15	.68***
	(.17)	(.17)	(.18)	(.15)	(.17)	(.18)
Adj. R^2	.12	.06	.08	.28	.42	.31
Num. obs.	2014	2014	2014	2014	2014	2014

Note: Coefficients estimated from OLS models, including the full set of controls as reported in Table 9. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

F.3 Potential Repeat Observations

In this subsection, we redo our main analyses without dropping potential repeat observations. That is, we keep subjects who self-reported being registered on both platforms, CloudResearch and Prolific, in the sample. For the Gender Experiment, this increases the sample size by about 23%, from $2,053$ to $2,516$ observations. For the Race Experiment, the increase is about 22%, from $2,014$ to $2,447$ observations. Tables 13 and 14 show the results for the Gender and Race Experiment, respectively. The tables are structured such that for each vignette, the first column shows the results from the main analyses (as presented in Section 4), and the second column shows those from the sample that includes potential repeats.

First, we note that, across the board, estimated coefficients and corresponding standard errors are quite similar, with the result that, for the most part, statistical (in-) significance as reported in the main results is unaffected. With respect to our coefficient of interest, β_3, on the interaction $male \times illegal$ / $white \times illegal$, we see little change in the Race Experiment. Across all vignettes, the estimated coefficients are slightly smaller (in absolute terms) in the sample that includes potential repeats and remain insignificant in all cases where they were previously insignificant. In the Gun Possession Vignette, the only case where we detected a significant interaction effect, the estimate becomes slightly less precise (p-value increases from $p = .010$ to $p = .028$, and from $p = .057$ to $p = .170$ when applying the Benjamini-Hochberg correction for six hypotheses).

In the Gender Experiment, the coefficients estimated for β_3 are either almost the same or slightly larger (in absolute terms) in the sample that includes potential repeats. In the case of the Speeding and the Casino Vignette, they also become (marginally) significant. For Speeding, the p-value goes from $p = .144$ to $p = .027$ (but increases to $p = .164$ once we apply the Benjamini-Hochberg correction for six hypotheses). For Casino, the p-value reduces from $p = .149$ to $p = .081$ (and rises to $p = .488$ after Benjamini-Hochberg correction).

Overall, we conclude that the exclusion of potential repeat participants does not alter our results in a meaningful way.

Table 13: Social appropriateness of behaviors in the GENDER experiment, excluding vs. including potential repeat observations

	Drink driving		Speeding		Casino		Alcohol to youth		Minimum wage		Parental leave	
male==1 (M), β_1	.05	.05	−.09	−.13**	.05	.01	−.04	.01	.02	.04	.13*	.14**
	(.07)	(.06)	(.07)	(.06)	(.06)	(.06)	(.07)	(.06)	(.07)	(.07)	(.07)	(.07)
illegal==1 (I), β_2	−.16**	−.16***	−.18**	−.20***	−1.03***	−.98***	−.87***	−.82***	−.67***	−.60***	−.93***	−.90***
	(.07)	(.06)	(.07)	(.07)	(.06)	(.06)	(.07)	(.06)	(.07)	(.07)	(.07)	(.07)
M x I, β_3	−.11	−.10	.15	.20**	.13	.14*	.00	−.01	−.01	−.07	.00	−.05
	(.10)	(.09)	(.10)	(.09)	(.09)	(.08)	(.10)	(.09)	(.10)	(.09)	(.11)	(.10)
absolute distance (AD), β_4	.03***	.03***	−.07***	−.07***	.00	−.00	−.01	−.01	.04***	.04***	.02*	.02
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
M x AD, β_5	−.01	−.01	.02	.02*	−.01	−.00	.01	−.00	−.01	−.01	−.01	−.01
	(.02)	(.01)	(.02)	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)	(.01)	(.02)	(.02)
I x AD, β_6	−.06***	−.05***	.01	.02	−.01	−.02	−.02	−.02	−.07***	−.07***	−.03**	−.03**
	(.02)	(.01)	(.02)	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)	(.01)	(.02)	(.01)
M x I x AD, β_7	.03	.02	−.02	−.03	−.03*	−.04**	−.01	−.00	.02	.02	.00	.02
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Intercept	−.27***	−.33***	.63***	.70***	.90***	.86***	.31***	.28***	−.04	−.06	.14	.12
	(.08)	(.08)	(.09)	(.08)	(.08)	(.07)	(.08)	(.08)	(.09)	(.08)	(.09)	(.08)
Sample: Incl. repeats	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R^2	.13	.15	.07	.07	.57	.53	.44	.41	.41	.39	.47	.44
Num. obs.	2053	2516	2053	2516	2053	2516	2053	2516	2053	2516	2053	2516

Note: Coefficients estimated from OLS models, including the full set of controls as reported. The prefix 'S' in variable names indicates that the variable controls for characteristics of the respondent, not the vignette. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 14: Social appropriateness of behaviors in the RACE experiment, excluding vs. including potential repeat observations

	Drink driving		Speeding		Gun		Marijuana		Minimum wage		Age of consent	
white==1 (W), β_1	.29	.10	$-.15$	$-.02$	$-.04$	$-.14$	.01	.07	$-.32$	$-.27$	.06	$-.04$
	(.25)	(.23)	(.25)	(.22)	(.20)	(.18)	(.16)	(.14)	(.24)	(.22)	(.25)	(.23)
illegal==1 (I), β_2	$-.10$	$-.16^{**}$	.05	$-.01$	$-.37^{***}$	$-.33^{***}$	$-.55^{***}$	$-.51^{***}$	$-.69^{***}$	$-.69^{***}$	$-.68^{***}$	$-.62^{***}$
	(.07)	(.06)	(.07)	(.06)	(.08)	(.07)	(.07)	(.06)	(.07)	(.07)	(.07)	(.07)
W x I, β_3	$-.09$	$-.03$	$-.07$	$-.00$	$.29^{***}$	$.22^{**}$	.14	.06	.11	.11	.14	.06
	(.10)	(.09)	(.10)	(.09)	(.11)	(.10)	(.09)	(.08)	(.10)	(.09)	(.10)	(.09)
absolute distance (AD), β_4	$.05^{***}$	$.04^{***}$	$-.06^{***}$	$-.06^{***}$	$-.03^{**}$	$-.03^{**}$	$-.01$	$-.01$	$.03^{***}$	$.03^{***}$	.00	.00
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
W x AD, β_5	$-.03^{*}$	$-.03^{*}$	$-.00$	$-.01$	.03	.02	.02	.01	.01	.01	.02	.01
	(.02)	(.01)	(.02)	(.01)	(.02)	(.02)	(.01)	(.01)	(.02)	(.01)	(.02)	(.01)
I x AD, β_6	$-.06^{***}$	$-.05^{***}$	$-.02$	$-.00$	.01	$-.00$	$-.02$	$-.02^{*}$	$-.06^{***}$	$-.06^{***}$	$-.03$	$-.03^{**}$
	(.02)	(.01)	(.02)	(.01)	(.02)	(.02)	(.01)	(.01)	(.02)	(.01)	(.02)	(.01)
W x I x AD, β_7	.03	.02	.02	.01	$-.06^{**}$	$-.04^{**}$	$-.03$	$-.02$	.00	$-.01$	$-.02$	$-.01$
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Intercept	$-.34^{**}$	$-.35^{**}$	$.76^{***}$	$.88^{***}$	$.49^{***}$	$.39^{**}$	$.89^{***}$	$.88^{***}$	$-.15$	$-.15$	$.69^{***}$	$.60^{***}$
	(.17)	(.16)	(.17)	(.15)	(.18)	(.17)	(.15)	(.13)	(.17)	(.16)	(.18)	(.16)
Sample: Incl. repeats	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R^2	.12	.11	.06	.06	.08	.08	.28	.28	.42	.41	.32	.31
Num. obs.	2014	2447	2014	2447	2014	2447	2014	2447	2014	2447	2014	2447

Note: Coefficients estimated from OLS models, including the full set of controls as reported. The prefix 'S' in variable names indicates that the variable controls for characteristics of the respondent, not the vignette. Standard errors are reported in parentheses. Stars indicate statistical significance at $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

G Simulations for power analyses

To assess the meaningfulness of our null results, we ran Monte-Carlo simulations as an ex-post power analysis exercise. We aimed to determine the minimum group differences in the discontinuities in appropriateness ratings at the threshold that we could detect with a reasonably high likelihood, given our number of observations. To this aim, we used one group in each experiment to determine a benchmark distribution of appropriateness ratings.[4] We used the benchmark distributions to generate counterfactual distributions (one for each experiment) by synthetically shifting the ratings of actions falling on the illegal side of the threshold. More precisely, for each illegal action studied in each of our vignettes, we shifted a mass s of the ratings distribution from the least appropriate category and distributed it evenly over the remaining categories.[5] This generates a rightward shift (i.e., towards higher appropriateness) in the distribution of ratings of illegal actions in the counterfactual distributions compared to the benchmark distributions.

We then drew (with replacement) from the two distributions to generate two samples (benchmark and counterfactual), each containing, on average, 128 ratings for each of the eight actions in each vignette. Thus, for each vignette, we drew samples of approximately 1000 benchmark and 1000 counterfactual observations, comparable with the sample sizes used in our experiments. We repeated this process 1000 times. Figures 1 and 2 show the norm

[4]The results we report here are based on the female group in the Gender Experiment and the African American group in the Race Experiment as benchmarks. Results are very similar when we use the male and White American groups as benchmarks.

[5]For instance, if the benchmark distribution of ratings for a given action was 50% 'Very socially inappropriate,' 30% 'Somewhat socially inappropriate' and 20% 'Somewhat socially appropriate,' and s=*.12* we generated the counterfactual distribution: 40% 'Very socially inappropriate,' 34% 'Somewhat socially inappropriate,' 24% 'Somewhat socially appropriate' and 4% 'Very socially appropriate.' If s was larger than the mass in the lowest appropriateness category, we drew from the adjacent category until we rightward-shifted a total of s percentage points.

functions obtained by averaging over the 1000 simulations for the Gender and Race Experiment, respectively, for a shift of s=.22. This is the smallest shift that allowed us, on average, to detect a statistically significant effect (i.e., a p-value associated with β_3 lower than $p < .1$) in about 81% of the regressions. This calculation excludes the Speeding Vignette regressions, for which we detect significant differences in substantially fewer regressions compared to all other vignettes due to the high appropriateness ratings of illegal behavior. The red (black) lines show the appropriateness ratings for the samples drawn from the benchmark (counterfactual) distribution. In the legal domain, the norm functions for the two groups are identical by construction. In the illegal domain, appropriateness ratings for the counterfactual group are shifted upwards (higher appropriateness). This produces synthetic differences in the magnitude of the discontinuities at the threshold for the benchmark and counterfactual groups (generally smaller for the latter).

With each of these samples, we ran OLS regressions as we did for our main analyses, including indicators for the counterfactual group and its interaction with the illegal dummy and the absolute distance of the action from the threshold (*absdist*), keeping the benchmark group as the reference:

$$rating_v = \beta_0 + \beta_1 counterfactual_v + \beta_2 illegal_v + \beta_3 counterfactual_v \times illegal_v + \beta_4 absdist_v + \beta_5 counterfactual_v \times absdist_v + \beta_6 illegal_v \times absdist_v + \beta_7 counterfactual_v \times illegal_v \times absdist_v + \epsilon.$$

As in our main analyses, β_3 is our coefficient of interest, as it measures the difference in the discontinuity at the threshold for observations drawn from counterfactual distribution relative to those drawn from the benchmark distribution. We determine the Benjamini-Hochberg corrected p-values in each regression (corrected for the six hypotheses we test in each experiment) and calculate the average estimate of β_3, the average associated p-value, and the fraction of regressions that return a p-value of $p < .1$.

Table 15 shows the results of this exercise. The average estimated coefficients range from .140 to .299, meaning that the decrease in appropriateness ratings when moving from just legal to just illegal is less pronounced for the counterfactual group than for the benchmark group. The average Benjamini-Hochberg corrected p-values range between .020 and .245; the fraction of p-values below .01 ranges from .387 to .960.

How "large" are these effects that we are powered to detect? To gain some insights, we computed how an average shift in the ratings distributions of $s=.22$ translates into differences between benchmark and counterfactual groups in terms of Hedges' g, a standard measure of effect size. For each vignette and each action on the illegal side of the threshold, we calculated the average Hedges' g between benchmark and counterfactual groups across the 1000 simulations. We then averaged these measures across actions and vignettes to obtain an aggregate measure of effect size for each of our Gender and Race experiments. We obtain an aggregate Hedges' g of -0.47 in each experiment.

To put this number in perspective, consider the average effect size of the discontinuities in the norm functions observed in our experiments (measured as the standardized mean difference between the two actions closest to the threshold). For the Gender Experiment, the average Hedges' g for the observed discontinuities at the threshold is -1.14 for women and -1.10 for men. For the Race Experiment, the average Hedges' g is -0.63 for African Americans and -0.69 for White Americans. Thus, in the case of Gender, we are powered to detect differences between groups that are between 41% and 43% of the size of the discontinuities observed in that experiment. In the case of Race, we are powered to detect differences between groups that are between

68% and 75% of the size of the discontinuities observed in that experiment.

As another benchmark, consider the original Krupka-Weber experiment manipulating the frame of dictator game actions (give or take). The paper reports differences in the distribution of ratings across the give and take frames of a magnitude of -0.40 in terms of Hedges' g for dictator actions that allocate more resources to the dictator than to the recipient.[6] Thus, the minimum detectable effects in our study are roughly equivalent to the observed magnitude of the effect of give/take framing on norms in the original Krupka-Weber paper.

[6]The difference in ratings for actions that allocate equal resources to the two players, or more resources to the recipient, are smaller (Hedges' g = -0.12), which is not surprising since the actual framing of these actions did not differ across treatments.

Table 15: Simulation results

Base	Vignette	$\overline{\beta}_3$	$\overline{p}$	σ
Women	Drink driving	0.259	0.074	0.795
	Speeding	0.180	0.178	0.542
	Casino	0.299	0.020	0.960
	Alcohol to youth	0.267	0.076	0.797
	Minimum wage	0.264	0.095	0.760
	Parental leave	0.266	0.098	0.751
African American	Drink driving	0.269	0.063	0.829
	Speeding	0.140	0.245	0.387
	Gun	0.290	0.077	0.791
	Marijuana	0.207	0.101	0.722
	Minimum wage	0.264	0.084	0.770
	Age of consent	0.268	0.070	0.810

Note: The table shows, based on 1,000 regressions with 2,000 simulated observations, the mean point estimate for the coefficient of interest ($\overline{\beta}_3$), the mean Benjamini-Hochberg corrected p-value corresponding to the coefficient ($\overline{p}$), and the fraction of regressions (σ) for which β_3 was statistically significant at $p \leq .1$ after applying Benjamini-Hochberg correction.

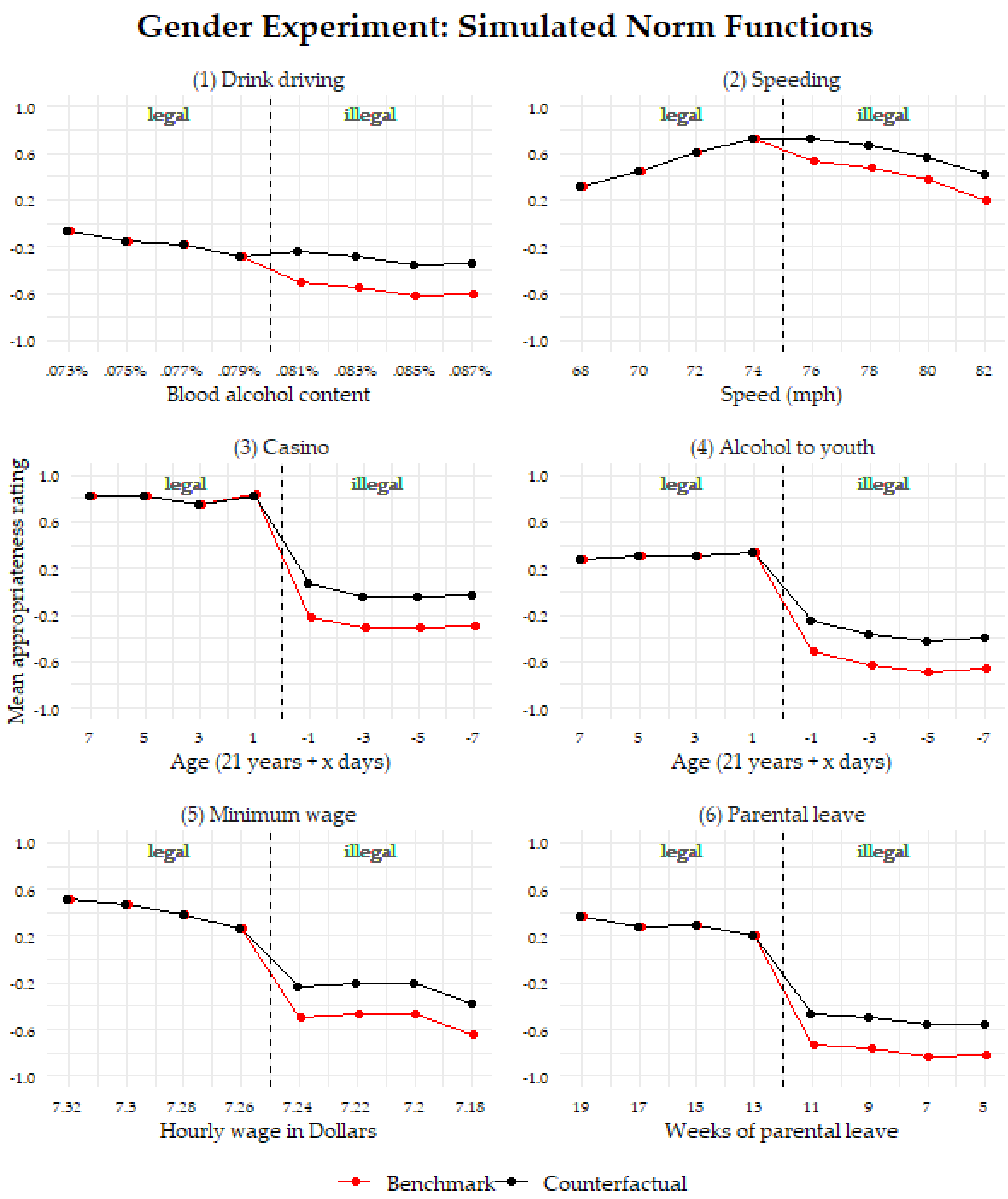


Figure 1: Simulated data, based on Women distribution

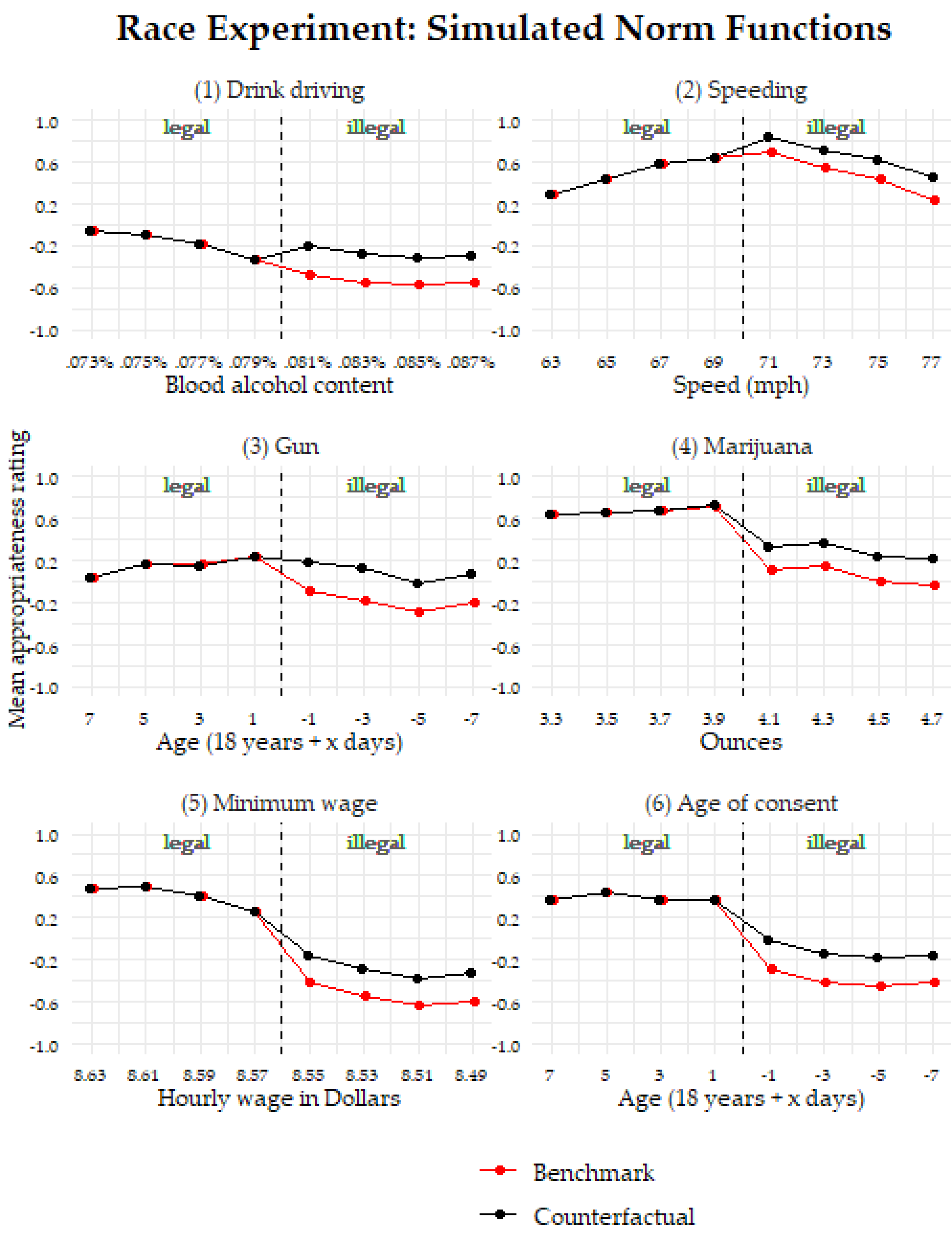


Figure 2: Simulated data, based on African American distribution