

Disciplining Punishment (Attitudes)*

Sanford C. Gordon

Gregory A. Huber

Wilf Family Department of Politics

Political Science and ISPS

New York University

Yale University

sanford.gordon@nyu.edu

gregory.huber@yale.edu

August 12, 2026

Abstract

Attitudes toward crime and punishment are hypothesized to heavily influence criminal justice politics and reform. But commonly used survey items often conflate analytically distinct concepts, obscuring the roots of policy disagreements. To disentangle these, we present a theoretical model clarifying what traditional survey measures can and cannot reveal about citizen punitiveness. Then, in a series of novel surveys, we introduce new items designed to isolate three specific considerations: beliefs about system fallibility, sentencing preferences for guilty (as distinguished from convicted) defendants, and appropriate burdens of proof. We find that (1) Black respondents perceive the system as less trustworthy but report lower burdens of proof to convict; (2) respondents across all races perceive institutional discrimination against Black defendants; (3) in experimental vignettes, participants do not condition ideal sentences or burdens of proof on defendant race; and (4) the association between our measures and standard survey items largely comports with theoretical predictions.

*We thank John Cho for programming assistance and Yarden Morad for research assistance. Earlier versions of this paper were presented at the Yale University Justice and Injustice Conference, the New York University Political Economy Workshop, the Wallis Institute Political Economy Workshop at the University of Rochester, and the London Politics, Philosophy, and Economics Conference, where the authors received valuable feedback.

1 Introduction

How punitive is the American public, and how much do Americans disagree about crime and punishment? Existing research suggests that U.S. citizens are, on average, especially punitive compared to citizens in other countries (Roberts 2003, ch. 2, but see Kugler et al. 2013), and that substantial differences persist between racial groups in support for harsh punishments such as the death penalty (e.g., Pew Research Center, 2021).

In this paper, we articulate and implement a strategy for sharpening the inferences we can make from survey evidence about the distribution of underlying punitiveness and beliefs about criminal justice institutions in the United States. We argue that conceptual ambiguity is intrinsic to the most commonly used survey items about crime and punishment. This ambiguity arises because those questions, particularly ones designed to measure overall “top-level” evaluations of the criminal justice system, evoke multiple considerations that a microfounded theory of preferences and beliefs would properly treat as analytically distinct. These considerations include beliefs about system *fallibility*, preferences regarding appropriate punishment for defendants who are *certainly guilty* (as distinguished from defendants who have been *charged* or *convicted*), and preferences concerning the *burden of proof* for convicting defendants. Isolating them is critical for understanding whether criminal justice policy is or is not democratically responsive to public opinion, and how group differences inform our understanding of what kinds of coalitions are feasible for pursuing (or stymieing) systemic reforms. Absent informative and unbiased measures, we cannot be certain of either the mapping from attitudes to policies or what is implied by the existence (and persistence) of apparent group differences.

To further develop this intuition, we describe a heuristic decision-theoretic model in which a “finder of fact” must decide whether to convict or acquit a specific defendant charged with a particular crime given uncertainty about that defendant’s guilt and the punishment for that defendant if convicted. The model extends naturally to a survey setting in which respondents are asked to contemplate these decisions with respect to populations of defendants. Our framework provides a structure for understanding what considerations ought to be measured, how survey

questions should be posed to elicit them, and how existing survey items likely miss the mark by conflating them. It also illustrates the dependence of responses to these questions on, *inter alia*, the severity of the crime, defendant characteristics (which may invoke legitimate classifications, e.g., prior history, or suspect ones, e.g. defendant race), and anticipated punishments.

In light of these limitations, we describe a sequence of three surveys aimed at cleanly measuring our theoretically-derived building blocks of variation in top-level criminal justice attitudes. After laying out our novel survey instruments, we assess averages and racial-group differences, provide experimental estimates of the effects of defendant- and case-level attributes on different outcomes, and demonstrate that the conclusions that are warranted from this microfounded account are distinct from those that might be arrived at using traditional measures.

Four principle sets of findings emerge. First, we demonstrate that intrinsic punitiveness and beliefs about the fallibility of the criminal justice system are meaningfully distinct concepts in the minds of respondents. Second, we demonstrate that well-documented racial differences in top-level attitudes about criminal justice (for example, support for the death penalty) are misleading. We find that to the extent that Black respondents can be categorized as “less punitive” than White ones, this difference is due almost entirely to the fact that Black respondents perceive higher levels of systemic fallibility than White respondents, *even holding the race of the posited defendant constant*, rather than exhibiting a higher burden of proof to convict given uncertainty about defendant guilt. Third, subjects of all races perceive institutional discrimination against Black defendants. At the same time, in our analysis of responses to questions about detailed vignettes, we find no evidence of racial discrimination by the subjects in either preferred sentences for certainly guilty defendants or their respective burdens of proof. Finally, we demonstrate using factor analysis that our novel questions map to three separate dimensions, and show that several existing and commonly-used survey items are each predicted by two or more of these factors.

We conclude by discussing the implications of our findings for understanding the representativeness of criminal justice policy, across-group differences in perceptions of the criminal justice system, and the likelihood that specific criminal justice reforms will attract public support and reduce racial group differences in evaluations of American criminal justice policy. More broadly, we

argue that researchers interested in assessing the relationship between public attitudes and public policy in a broad array of policy domains can take a similar approach to the one advanced here by developing survey questions more closely linked to fundamental considerations. Such an approach should also inform efforts to model individual-level differences in attitudes in this broader range of domains.

2 Related Research

Our research draws on and seeks to contribute to several strands of related research about attitudes toward crime and punishment. The first of these concerns methodology and measurement. In political science, a number of researchers starting with Stimson (1999, 2004) have sought to assess the contribution of public attitudes about crime and punishment to macro-level “public mood” (i.e., average public opinion on a liberal-conservative dimension) over time, and the effects of that mood on public policy outcomes. Stimson develops a “dyad ratios algorithm” to knit together numerous survey instruments and impute a dynamic, two-dimensional measure of mood. Questions about punishment map strongly onto his second dimension. Other scholars deploying this or related measurement strategies include Nicholson-Crotty et al. (2009), Enns (2014, 2016), and Ramirez (2013), often with the goal of understanding how and when public opinion shapes policy.

Scholars taking a micro-level perspective on measurement of punitive attitudes have stressed the mutability of those attitudes in the face of public ignorance about the actual incidence of crime (Roberts and Stalans, 2000; Frost, 2010). Others have argued that reducing criminal justice-related attitudes to a single-dimensional measure is reductive. Pickett and Thomas (2014), for example, note that the seemingly contradictory views of American survey respondents, who tend to express support for both harsh sentences and rehabilitation, are reconcilable in a multidimensional framework (see also Pickett, 2019). Matthews (2005) likewise stresses the importance of disaggregating attitudes about punishment and rehabilitation. Adriaenssen and Aertsen (2015) suggest distinguishing among attitudes toward sentencing severity, the goals of sentencing, and specific policies (e.g., three strikes rules). Employing a survey vignette approach similar to the one we adopt below, Simonson (2011) documents differences in attitudes about the adequacy of penal laws and their

implementation.

Our approach, by contrast, seeks to ground survey measurement of criminal-justice related attitudes in a microfounded decision-theoretic account of preferences and beliefs. This approach draws on an older, parallel line of theory-development concerning criminal justice errors and their effects on citizens and jury members (Kaplan 1968; Nagel et al. 1981; Milanich 1981; Connolly 1987; for an early critique see Tribe 1971); and the effect of criminal justice errors on optimal deterrence (Kaplrow and Shavell, 1994; Polinsky and Shavell, 2000). Our theoretical strategy is a generalization of this approach and differs in its application to developing improved survey measures for understanding public opinion and sharpening the interpretation of extant survey data. Closest in spirit to the current paper’s theoretical contribution is Andreoni (1991), who notes that excessive punishment might induce juror reluctance to convict, and that this effect informs optimal levels of punishment for specific crimes (see also Robinson and Darley, 2003; Bierschbach and Stein, 2005).

The key advantage of our approach, we argue, is to clarify how analytically distinct considerations may influence various top-level evaluations of the criminal justice system, making interpretation of the latter without strong (and likely untenable) assumptions difficult. Additionally, our model helps us to identify which sets of preferences and beliefs underlie those more general assessments and should therefore be measured directly. Importantly, the model guides how to measure these outcomes in a way that is minimally affected by other preferences and beliefs, preserving conceptual clarity and facilitating comparisons across individuals and groups.

Our study also is relevant for research concerning the antecedents of attitudes about criminal justice. Two prominent research questions are especially pertinent: one assesses whether the historical increase in public demand for more punitive measures in the last decades of the 20th century was more of an elite-driven process (as argued by Weaver 2007 and Gottschalk 2014) or a bottom-up one (Garland, 2001; Miller, 2016; Enns, 2016). Another considers the effects of emphasizing the possibility of criminal justice system errors on support for punitive sanctions, finding in the experimental context that emphasizing these errors decreases support for the death penalty (e.g., Dardis et al., 2008).

We are particularly interested in antecedents because variation in public preferences and assess-

ments of the criminal justice system inform, to a great degree, how we assess democratic responsiveness. For example, suppose respondents report a belief that the criminal justice system is too harsh. Does this reflect an assessment that punishments don't fit crimes or that the accused may not actually be guilty of those crimes? If the former, then the policy change that would satisfy public preferences concerns decriminalization or sentencing reform; if the latter, it is about reducing errors in who is punished (for example, by investing in surveillance technology or changing the incentives of police and prosecutors). Looking forward, should public support for more punitive criminal justice practices continue to evolve, our work suggests that assessing what policies are entailed by that change requires understanding in what ways the system is actually perceived as deficient.

As noted above, the study of racial disparity is bound up in the study of criminal justice in the United States. A related strain of research concerns the relationship between race and criminal justice *attitudes*: this includes documenting differences in the attitudes of White and Black respondents (e.g., Secret and Johnson, 1989; Cochran and Chamlin, 2006; Unnever and Cullen, 2007) as well as research on predictors of punitive attitudes among White respondents, such as racial resentment and prejudice (e.g., Peffley and Hurwitz, 2010; Unnever and Cullen, 2010; Unnever, 2013). Our model reveals the difficulty of diagnosing the causes of group-level differences in top-level criminal justice evaluations, because such evaluations could originate in different primitive preferences (e.g., the appropriate punishment for someone who is truly guilty of a crime) or different beliefs (e.g., about the error-proneness of the criminal justice system). And in our descriptive analysis using our novel survey items, we assess whether and where racial differences in preferences and beliefs arise.

An additional research agenda related to the current paper concerns policy responsiveness and crime in the United States. Arnold and Carnes (2012), for example, find a negative relationship between the public's approval of New York City mayors and the city's murder rate. Enns (2014,2016) finds that a critical driving force behind the incarceration rate in the United States at both the national level and in the states has been the public's increasing punitiveness. And moderate Democrats attributed electoral losses in 2021 to Republican success in linking Democratic

candidates with unpopular proposals to “defund” the police (Sklansky, 2025), which is interpreted by survey respondents as seeking to reduce the number of police on the street and as leading to increases in crime (Vaughn et al., 2022). Isolating the sources of public attitudes about crime and punishment is critical for assessing how elected officials respond to political pressures in democratic settings and whether those responses are linked to underlying changes in public beliefs and preferences.

Finally, our research contributes indirectly to the literature on legitimate authority. Error-proneness is closely related to trust in, and perceived legitimacy of, governing institutions, particularly those endowed with enforcement responsibilities (Easton, 1965; French and Raven, 1959; Leventhal, 1980; Gordon and Huber, 2019; Dickson et al., 2022). Our analysis sheds light on how beliefs about the fallibility of criminal justice institutions are related to revealed preferences about punishment, and we show those preferences will depart from the sanctions one would prefer if one did not otherwise believe the system was error-prone.

3 A Model for Distinguishing Criminal Justice Attitudes

In this section we present a heuristic model of criminal justice attitudes focusing on three interlocking elements: (1) system fallibility, instantiated specifically in an individual’s suspicion that defendants convicted or on trial may not be guilty; (2) the burden of proof, meaning how sure someone would have to be to convict a defendant on trial; and (3) the individual’s views regarding preferred punishments for guilty defendants. Our model demonstrates how these considerations interact, and how in light of these interactions, commonly-used survey questions seeking to elicit top-level evaluations of the criminal justice system are conceptually ambiguous. In particular, those questions threaten to conflate multiple, distinct aspects of underlying beliefs and preferences. While a model cannot capture every set of attitudes relevant for beliefs about the criminal justice system, this exercise demonstrates the disciplining value of sharply articulating the mapping between fundamental considerations and elicited survey responses. This, in turn, informs the survey and survey-experimental empirical approach we take below.

Figure 1: Preferences over Criminal Justice Outcomes: Restrictive and General Payoffs

(1) Restrictive Payoffs			(2) General Payoffs		
	Convict	Acquit		Convict	Acquit
Guilty	0	$-a$	Guilty	$b_i(v, d, s)$	$-a_i(v, d, s = \hat{s}_i(v, d))$
Innocent	$-c$	0	Innocent	$-c_i(v, d, s)$	0

Note: in the restrictive model shown in the left panel, a and c are both positive. In the general model, $a_i(v, s, d)$ and $c_i(v, s, d)$ are both positive, while $b_i(v, s, d)$ can be either positive or negative.

3.1 A Restrictive Baseline Model

We first present a restrictive baseline model, drawing on decision-theoretic models of trial fact-finding, which at least since Kaplan (1968) have expressed outcomes and their associated payoffs in the form of a 2×2 matrix, with conviction/acquittal on one axis and true guilt/innocence on the other. The simplest formulation of this approach, displayed in the left panel of Figure 1, normalizes the payoffs of both kinds of accurate decisions – true positives (convicting the guilty) and true negatives (acquitting the innocent) – to zero (e.g., Shotts and Wiseman 2010; Gordon and Huber 2002), while allowing the costs of Type I errors (convicting the innocent, given the null hypothesis of “not guilty”) and Type II errors (acquitting the guilty) to vary.

Let $p_{i,d}$ denote an individual “finder of fact” (e.g., a juror) i ’s belief that defendant d is guilty. The expected utility of acquittal is affected by the cost of acquitting the guilty (a), which happens with probability $p_{i,d}$. The expected utility of conviction is affected by the harm from punishing the innocent (c), which happens with probability $(1 - p_{i,d})$. Comparing these, an individual finder of fact will prefer convicting if and only if

$$p_{i,d} > \frac{c}{c + a} \equiv \hat{p}_{i,d}^r,$$

where $\hat{p}_{i,d}^r$ is individual i ’s *burden of proof to convict* in the current (restrictive, denoted by the superscript r) setting. The burden of proof refers to how certain a person must be about a defendant’s guilt to prefer convicting to acquitting—that is, how high $p_{i,d}$ must be to justify a conviction.

It is sometimes useful to express the burden of proofs in terms of odds:

$$\frac{p_{i,d}}{1 - p_{i,d}} > \frac{c}{a} \equiv \hat{R}_{i,d}^r.$$

The value of the odds formulation is made plain when one considers Sir William Blackstone’s famous dictum that it is “[b]etter that ten guilty persons escape than that one innocent suffer” (Blackstone, 1765). Viewed through the lens of the payoff matrix above, this statement is equivalent to the following inequality: $c > 10a$ or, substituting, $\hat{R}_{i,d}^r > 10$. Given the foregoing, Blackstone’s rule implies that the odds of guilt must be higher than 10:1 (alternatively, the probability must be higher than 90.9%) to justify convicting. 10:1, of course, reflects a particular set of normative commitments, which might differ across individuals or circumstances (Volokh, 1997).

We label this baseline specification “restrictive” for several reasons. First, imposing equality in the payoffs associated with “correct” outcomes lacks strong substantive justification: people may not view appropriately punishing the guilty and acquitting the innocent as equally beneficial.

Second, it ignores the fact that what is at stake in these tradeoffs is not absolute, but contingent, depending on, among other things, the severity of the crime, the anticipated punishment for the crime, and characteristics of the defendant.¹ How these considerations are weighted and interact in the minds of individuals is likely to be complex and idiosyncratic, so the notion that individuals will maintain a constant burden of proof across all of these is implausible.

Third and related, it does not account for the fact that individuals may have a sense of justice that informs their belief about the appropriate penalty for a guilty individual given a particular crime (Andreoni, 1991; Robinson and Darley, 2003). This is a person’s preferred punishment for a defendant when their guilt is certain. Suppose this appropriate penalty is significantly lower than the one the individual expects will be assigned by the state. Then the perceived cost to the individual of convicting a *guilty* defendant may be larger than the perceived cost of acquitting that person. In other words, these considerations induce variation not just in the magnitude of the payoffs in each cell of the 2×2 , but also potentially their ordering.

¹Of course, with respect to defendant characteristics, we may view some as legitimate to consider (e.g., prior criminal history) and others as illegitimate (e.g., defendant race).

At the same time, the restrictive model incorporates two quantities that, as we discuss below, are invoked in commonly-used survey questions: trust in the criminal justice system (as proxied by the individual’s belief that the defendant on trial is in fact guilty of the crime) and the burden of proof to convict. We tend to regard higher trust in the system and/or a lower burden of proof as indicative of greater punitiveness in general, but they are analytically distinct. Additionally, the restrictive model does not include provision for a quantity fundamental to how punitive an individual is: that individual’s preferred punishment for a person certainly guilty of a particular crime.

3.2 A More General Approach

Setup. In light of these limitations, we shift to the more general payoff matrix depicted in the right panel of Figure 1. To capture individual-level differences and contingency, let $c_i(v, d, s) \geq 0$ denote individual i ’s subjective harm from convicting the innocent as a function of crime severity v , defendant characteristics d , and sentence s ; and $a_i(v, d, s) \geq 0$ the individual’s subjective harm from acquitting the guilty given those considerations. As with prior work, and reflecting the assumption that errors are costly, each of these functions is assumed to be weakly positive. Additionally, as in prior work, we normalize the benefit of acquitting the innocent to zero. To allow the payoffs from “correct” outcomes to diverge, let $b_i(v, d, s)$ denote i ’s subjective benefit from convicting the guilty (which could be positive or negative).

We assume that if the defendant is convicted and actually guilty, individual i ’s preferences (as captured by the $b_i(\cdot)$ function) are single-peaked in s and reach their maximum at an *ideal sentence given certain guilt*, $\hat{s}_i(v, d)$, which may be zero or positive. Second, if the defendant is convicted and actually innocent, the associated harm perceived by individual i is strictly increasing in the magnitude of the sentence.² Finally, (and as depicted in Figure 1), if the defendant is acquitted and actually guilty, the loss to the finder of fact is a function of the sentence that individual would have preferred to have seen implemented, $\hat{s}_i(v, d)$, and not the one that would have, counterfactually, been implemented had the guilty person instead been convicted.³

²By implication, individual i ’s ideal sentence given actual innocence is zero.

³This is in contrast to the benefit of convicting the guilty, which, as just discussed, does account for the expected

This more general model allows us to formally analyze interactions among burdens of proof, ideal sentences, and beliefs about system trustworthiness. Understanding these interactions will, in turn, inform our subsequent discussion of the ambiguity of existing survey items and our proposals for new ones.

The Burden of Proof Revisited. From the perspective of individual i , the expected utilities to convicting and acquitting a defendant are, respectively,

$$\begin{aligned} E[u_i(\textit{acquit}|v, s, d)] &= -p_{i,d}a_i(v, d, s = \hat{s}_i(v, d)) + (1 - p_{i,d}) \times 0 \\ E[u_i(\textit{convict}|v, s, d)] &= p_{i,d}b_i(v, d, s) - (1 - p_{i,d})c_i(v, d, s). \end{aligned} \tag{1}$$

An individual would prefer to convict if and only if

$$p_{i,d} > \frac{c_i(v, d, s)}{c_i(v, d, s) + b_i(v, d, s) + a_i(v, d, s = \hat{s}_i(v, d))} \equiv \hat{p}_i^g(v, d, s),$$

where the g superscript denotes the burden of proof in the general framework. As before, we can also express the burden of proof in terms of odds, à la Blackstone:

$$\frac{p_{i,d}}{1 - p_{i,d}} > \frac{c_i(v, d, s)}{a_i(v, d, s = \hat{s}_i(v, d)) + b_i(v, d, s)} \equiv \hat{R}_i^g(v, d, s).$$

The individual's burden of proof in the general framework responds in predictable ways to changes in the values of the $a_i(\cdot)$, $b_i(\cdot)$, and $c_i(\cdot)$ functions. When the cost of convicting the innocent increases, so does the burden of proof. When the cost of acquitting the guilty or the benefit of convicting the guilty increases, the individual's burden of proof decreases. Each of these payoffs may vary depending on the individual contemplating conviction, and how that person weighs the severity of the crime, the sentence at hand, and defendant characteristics. Accordingly, there is no reason to anticipate that the burden of proof would be constant either across different individuals

sentence. Hence, trading off between the two already accounts for what will happen if the person chooses to instead convict.

or even for the same individual considering different circumstances.⁴

Next, focus in particular on the relationship between the burden of proof and the anticipated sentence, s . At first glance, one might expect that as the anticipated sentence increases, the burden of proof should likewise increase owing to the potential cost of a wrongful conviction. But this intuition is an artifact of the restrictive model, and need not be the case once we allow the benefit of a conviction to vary. As we demonstrate in Appendix A, an indeterminacy arises because the cost of wrongful conviction associated with an increased sentence must be weighed against the possibility that, from the individual’s perspective, a guilty defendant will receive a more appropriate punishment. This may occur if the anticipated sentence initially lies below the individual’s ideal sentence given certain guilt. If the latter effect dominates, an increase in the sentence might actually push the burden of proof *downward* – in effect, the possibility that the “desserts” might be more “just” increases the fact finder’s willingness to risk convicting the innocent. By contrast, if the anticipated sentence already exceeds what a respondent prefers given certain guilt, then the tradeoff disappears, and further increases will lead to an increase in the burden of proof.

For the same reasons, the relationship between the burden of proof and i ’s preferred sentence given certain guilt, $\hat{s}_i(v, d)$, is indeterminate: depending on which side of that preferred sentence the anticipated sentence falls, an increase in punitiveness in terms of $\hat{s}_i$ can lead a finder of fact to increase *or* decrease her burden of proof to convict.

From a survey implementation perspective, formalizing the optimal burden of proof and recognizing its contingency imply subtleties that may arise when trying to measure the burden of proof that individuals prefer. Because the underlying costs and benefits depend on the crime’s severity, the defendant, the anticipated sentence, and the respondent’s preferred sentence given certain guilt, leaving these unspecified in a survey question about the optimal burden of proof makes an elicited burden less clearly informative. We demonstrate in Appendix A that under plausible assumptions, this issue is dramatically mitigated by eliciting a subject’s burden of proof when the anticipated sentence is exactly equal to the respondent’s ideal punishment for certainly guilty defendants. Un-

⁴We are not the first to make this observation. Kaplan (1968), for example, writes, “Probably the most important reason why we do not attempt to express reasonable doubt in terms of quantitative odds, however, is that in any rational system the utilities (or disabilities) that determine the necessary probability of guilt will vary with the crime for which the defendant is being tried, and indeed with the particular defendant.”

der those conditions, the elicited burden of proof has an intuitive “ratio of Type I to Type II error costs” interpretation akin to the one in the restricted model.

Preferred Punishment Given Conviction. Above, we defined a term $\hat{s}_i(v, d)$: the ideal sentence given defendant d ’s *certain guilt* of a crime with severity v . Guilt, of course, may not be certain. Uncertainty – encapsulated in the belief term $p_{i,d}$ – will inform individual i ’s preferred sentence given a *conviction*. Conditional on a conviction, the expected utility to i of sentence s is given by the second line of (1). In Appendix A, we show that under standard assumptions regarding the shape of the functions $b_i(\cdot)$ and $c_i(\cdot)$, i ’s optimal sentence given conviction, $s^*(p_{i,d}, v, d)$ is (1) unique; (2) strictly increasing in i ’s subjective belief that a defendant is guilty and her preferred sentence given actual guilt; and (3) strictly less than her preferred sentence given actual guilt for any $p_{i,d} < 1$. The effect of uncertainty about guilt is intuitive: as the individual becomes less certain that a defendant is guilty, she hedges by reducing the magnitude of her ideal sentence to balance out the increased threat of harm against the reduced benefits of appropriate punishment.

3.3 Implications for Survey Instruments

Resolving Conceptual Ambiguity. Our formal analysis identifies three analytically distinct quantities that, together, provide a rich account of variation in an individual’s attitudes about criminal justice:

1. *Beliefs about the fallibility of the criminal justice system.* How likely is a defendant *actually guilty* of the crime that defendant was accused of?⁵
2. *Preferences over appropriate punishment conditional on guilt.* If I am certain that a defendant charged with a crime is guilty, what is the appropriate punishment?
3. *Preferences over the burden of proof.* Given an anticipated sentence s , how certain do I have to be that a defendant is guilty of a crime to prefer convicting him to acquitting him?

⁵A different dimension of fallibility, which we do not consider in the current paper, concerns Type II errors: how likely a person guilty of a crime is apprehended, or apprehended but wrongly acquitted.

Note that while our model is of an individual “finder of fact” tasked with assessing a specific defendant, it is readily extended to individuals’ more general assessments of criminal justice of interest to the survey researcher. Thus, whereas in the model, the key question is whether a *specific defendant* on trial is guilty, in thinking about the system as a whole, we might be interested in a respondent’s belief about the incidence of guilt in the population of indicted or convicted criminals (or perhaps a subset, for example if primed to think about defendants of a particular race). Likewise, we might be interested in whether subjects would prefer the death penalty as a punishment for those guilty of murder *in general*.

Importantly, our analysis also demonstrates how these three apparently distinct factors interact. Briefly, beliefs about system fallibility affect preferred punishments for those found guilty of a crime. They also incorporate assessments of the system because error-prone systems wrongly punish innocent people (and acquit the guilty). How far an anticipated punishment departs from a preferred punishment affects the burden of proof because an individual might think that even a guilty defendant does not deserve the punishment the state is likely to carry out.

By comparison, consider some commonly used survey questions about crime and punishment from the General Social Survey that have been asked regularly over the survey’s history. Two are intended to capture comprehensive evaluations of the criminal justice system:

1. “How much confidence do you have in courts and the legal system?”
2. “In general, do you think the courts in this area deal too harshly or not harshly enough with criminals?”

Note the inherent conceptual ambiguity in these items. Low confidence (1) or beliefs that the system is too punitive (2) could arise for multiple reasons, including beliefs that those on trial are innocent, that the system’s burden of proof is wrong (i.e., from the respondent’s perspective, too many innocent people are convicted/guilty people are acquitted), and/or that the punishment for those who are guilty is excessive or inadequate.⁶

⁶This is in addition to the general problem of uncertainty about what confidence measures, ambiguity about what component of the legal system (e.g., the police, courts, juries, etc.) is being evaluated, and uncertainty about which criminals and for which crimes these evaluations are being solicited. Note also that low confidence could reflect beliefs that the burden of proof is too high (too many guilty people are not punished) or that assigned punishments are too lenient.

A third question is about preferred punishment for a given crime:

3. “Do you favor or oppose the death penalty for persons convicted of murder?”

Asking about the preferred punishment for those found guilty of murder and interpreting this as a direct measure of punitiveness conflates the respondent’s ideal sentence conditional on guilt ($\hat{s}_i$) and the preferred sentence conditional on being found guilty (s_i^*). The latter is affected by potential mistrust in the ability of the criminal justice system to differentiate between the guilty and the innocent ($p_{i,d}$) given the burden of proof that *others* impose, as well as the perceived cost of punishing the innocent. In other words, opposition to the death penalty (or any general preference for less severe sentences conditional on conviction) could stem from a belief that it is never appropriate for the state to take a life under any circumstances (or that this sanction is simply too harsh), or from a fear that the chances of punishing an innocent person are too high given the costs. To separate the former from the latter, it is instead necessary to measure what the appropriate penalty would be for an individual if that person were certainly guilty.⁷

A fourth question implicates preferences about the appropriate burden of proof:

4. “All systems of justice make mistakes, but which do you think is worse, to convict an innocent person or to let a guilty person go free?”

This question does not specify for which crime or for what potential sanction this assessment should be offered. Aside from the fact that being vague introduces the possibility that respondents are contemplating different crimes and/or defendants – a source of measurement error – it also leaves unspecified the sentence under consideration. As we noted above, when the sentence that would be assigned differs from what a respondent would prefer if a defendant was guilty, it affects the calculation of this error tradeoff directly. A respondent who anticipates a sentence harsher than they would prefer for a guilty person (i.e., over punishment of the actually guilty) has less to lose from acquitting the guilty, which might lead them to state that avoiding convicting the innocent is more important. On the other end of the spectrum, if the respondent thinks that sentences are

⁷A recent study by Mullinix et al. (2026) is noteworthy for employing survey vignettes to elicit preferences over punishment for hypothetical individuals who *committed* particular crimes instead of ones who were convicted of those crimes – in line with the prescription offered here.

very low, for example, it could lead them to respond that it is worse to acquit the guilty than convict the innocent, even if that same respondent would answer differently given more severe punishments. Consequently, answers to this question cannot be considered direct measures of the preferred burden of proof.⁸

Implications for Analysis of Across-Respondent Differences. As we discussed above, researchers have examined both average public opinion and group-level differences in those attitudes. The measurement challenges we diagnose in the prior subsection are general, but it is important to think about how they likely also affect our understandings of group-level differences.

Suppose, for example, that one finds Black Americans are more likely than White Americans to state that the criminal justice system “deals too harshly” with criminals. Assuming that this question is interpreted as eliciting an evaluation of whether the sentences assigned to those found guilty are appropriate, one might be tempted to interpret this as evidence that Black rather than White respondents are intrinsically less punitive, meaning that they prefer different sentences for the same truly guilty defendants. But, as we discuss above, assessments of harshness will also be affected by beliefs about the likelihood that the criminal justice system is erroneously punishing the innocent and acquitting the guilty, and Black and White Americans might differ in their beliefs about these error rates.⁹

Setting this measurement challenge aside, assume for the moment that beliefs about error-proneness are the same across racial groups. To reach the conclusion that Black respondents are less punitive than White ones, one would *also* need to assume that they are contemplating the same (distribution of) defendant(s) who had committed the same crime(s) and would receive the same sentence(s). If, for example, Black Americans tend to think about low-level drug offenders receiving long prison sentence while White Americans tend to imagine violent criminals being given short ones, the questions would capture assessments of different criminals rather than beliefs

⁸Additionally, this item does not assess how much worse it is to convict the innocent than to let the guilty go free, forcing all respondents who believe it is worse to convict the innocent to offer the same response.

⁹Another possibility is that Black and White respondents have beliefs about the relative error-proneness of the system for defendants of different races, i.e., differences in beliefs about racial bias in the criminal justice system. We examine this empirically below. Such beliefs about racial bias might also affect overall evaluations of the criminal justice system.

about a common object. The ambiguity created by vague survey questions suggests substantial value in being specific, in order to reduce differences across respondents (and subgroups) in what considerations they bring to their answers.¹⁰

4 Empirical Analysis

We conducted three studies of attitudes toward criminal justice incorporating both standard questions about punishment as well as novel questions and vignettes that aim to measure more directly the concepts discussed above. We implemented these studies using the Qualtrics web-based survey platform and fielded them to samples of US residents aged 18 and over provided by Bovitz/Forthright Access.¹¹

Study 1 was fielded between March 13 and 26, 2024 ($N = 2,650$); Study 2 between May 21 and 30 ($N = 1,734$); and Study 3 from September 27-October 14 ($N = 2,748$). Details about obtaining informed consent, sample demographics, and pre-treatment screening to identify inattentive respondents are discussed in the Appendix. All three studies were deemed exempt by the IRB at [REDACTED].

Rather than report the results of each study sequentially, we proceed by dividing initial discussion into three topics: (1) beliefs about the fallibility of the criminal justice system; (2) ideal sentence conditional on guilt; and (3) the burden of proof. Then, we demonstrate how our survey questions measure analytically distinct aspects of attitudes toward criminal justice in a way that traditional questions do not, and how these distinct considerations relate to those traditional survey items empirically.

¹⁰This would be a form of correlated measurement error, in which race would be correlated with other beliefs (apart from underlying punitiveness) that affect measured punitiveness, biasing across group comparisons.

¹¹Bovitz builds and maintains a proprietary panel of online respondents from which they purposively sample in an effort to provide a nationally-representative sample of respondents. Nonetheless, like all opt-in panels, those choosing to participate may not be representative of the larger population. For our experimental estimates, these difference are unlikely to affect inferences, but for our across-racial group analysis, we also control for other demographic factors (see below) to account for the possibility that racial subpopulations also differ on those dimensions. (For a discussion of how Bovitz builds its panel, removes low-quality respondents, and invites and compensates respondents, see <https://www.forthrightaccess.com/methods>.) See Appendix B for complete details on survey implementation, filtering, and sample demographics. In line with the recommendations and findings of Stagnaro et al. (2026), which discusses the quality and representativeness of Bovitz samples, we filter all of our samples on unrelated attention checks at the beginning of our surveys, which is shown to improve response quality without negatively affecting representativeness.

In our description of survey questions, we use **bold** to signify questions posed multiple ways to subjects, and *italics* to signify randomization.

4.1 Beliefs about Guilt and Innocence

We posed questions aimed at eliciting beliefs about system fallibility, with higher responses associated with greater trust that the system is not charging innocent people.

Survey Instruments. Study 1 was a pilot whose primary function was to determine question wording for our novel survey items and test different approaches to capturing responses (sliders vs. text boxes).¹² We asked respondents to estimate the proportion of people who were charged with murder who are either guilty or innocent, specifically: “Think about people who have been charged with murder in the United States in the last year. If found guilty, the sentence is life in prison. Out of every 100 people who were charged with murder, how many do you think were actually *[guilty/innocent]* of murder?”¹³ Study 2 posed this question separately for defendants **charged** and **found guilty** and used only the text entry method.¹⁴

In Study 3, which also used only the text entry method, we posed the question separately for armed robbery and burglary. To account for differences in understanding, each question contained a preamble defining the crime. Armed robbery was defined as “the taking, or attempted taking, of something of value from someone else by threat of force or violence while carrying a gun or other lethal weapon;” while burglary was defined as “the act of illegally entering a building with the intention of committing a theft or any other felony.” Subjects were then asked, “Out of every 100 *[Black/White/[BLANK]]* people **[charged with/convicted of]** of **[burglary/armed robbery]** in the last year, how many do you think are guilty?”

¹²Related research (Orr et al., 2026) has found that using sliders to gather numerical responses on the 0-100 scale produces larger groupings at 50, with smaller groupings at 0 and 100, suggesting some combination of measurement error and satisficing in survey response. While we did not find large differences by survey mode in clustering at 0, 50, and 100, we chose to implement open-ended text responses in the subsequent studies because it was easier to request a specific response from respondents who wanted to skip a question (see below). We restrict our analysis here to respondents assigned to the text-entry mode to make it consistent with studies 2 and 3.

¹³In the text-entry mode, we requested responses and validated that any entered response was numerical and between 0 and 100.

¹⁴In both studies 2 and 3, we also implemented validation to require respondents to enter a number between 0 and 100 and specified “Please enter -99 if you would like to skip this question.” We report rates of missing responses for each survey in Appendix B. These rates are higher in Studies 2 and 3.

Aggregate Findings. We begin our analysis by examining respondent beliefs about the guilt (or innocence) of defendants at various stages of the criminal justice process for three crimes: murder, robbery, and burglary. In Study 1, the mean response under the “guilty” frame was 78 percent with a median of 90, while the corresponding figures for “innocent” were 18 and 10. If wording the question in terms of guilt or innocence is immaterial, then the sum of the average responses in the two treatment groups should equal 100, a null hypothesis we can evaluate statistically. The medians add to 100 precisely, while the averages do not – we can reject the null hypothesis that they do, though only at the 10 percent level. This is likely due to the non-trivial grouping for both items at 50 (a common feature of 0-100 scales). This implies that the guilt versus innocence framing does not have a large influence on how respondents consider these items, apart from both being subject to bias toward 50. In light of this finding, in studies 2 and 3 we implemented only the guilty framing.

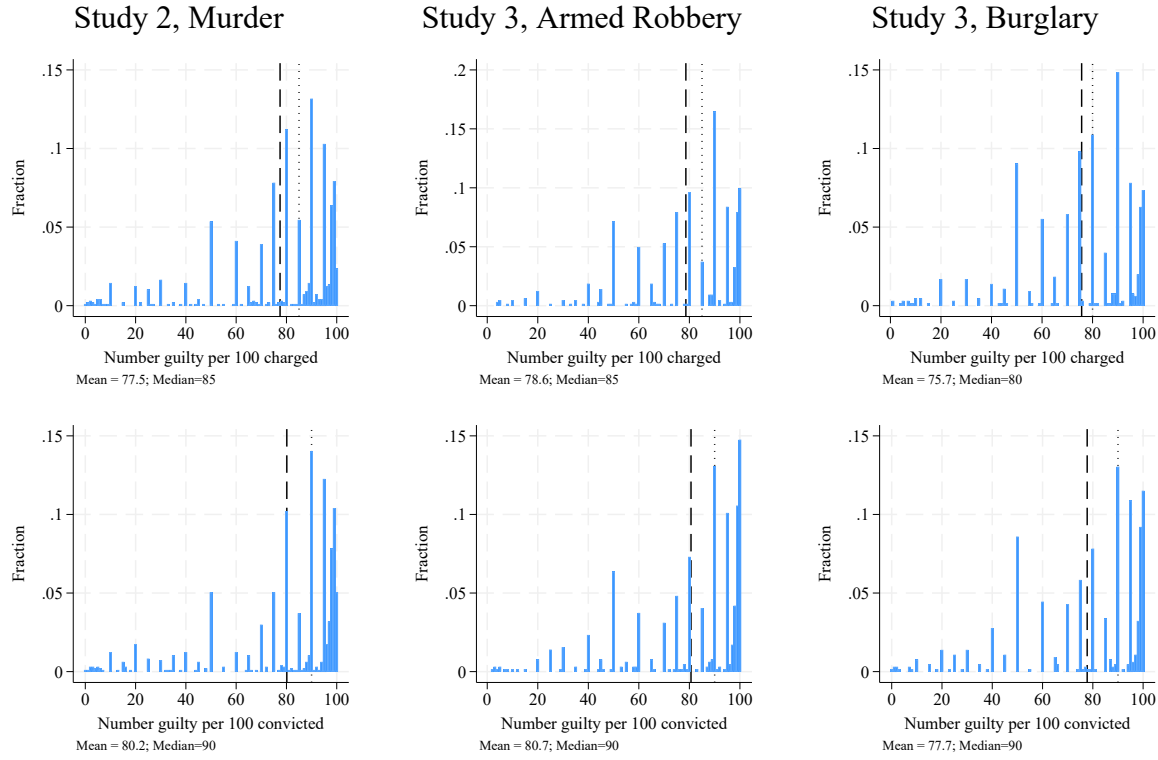
Taking the responses for number guilty at face value, these figures imply that at least half the sample believes 9 out of every 10 individuals charged with murder are guilty. There is wide variation in responses, however: The 75th percentile of responses is 95 (implying 19 out of every 20 individuals charged with murder are guilty) while the 25th is 70 (7 out of 10 guilty). 5% of responses are exactly 50, and 11% are less than 50.

Building on this analysis, in Studies 2 and 3, we distinguished between phases of the criminal justice process by gathering estimates both for those charged with a crime and those found guilty. In Study 2 we again focused on Murder, while in Study 3 respondents were asked about both Armed Robbery and Burglary (in a random order), randomizing whether defendant race was provided, and if so, if it was for Black or White defendants.¹⁵

The distribution of responses in Studies 2 and 3 are displayed in Figure 2, with the top row showing histograms for those charged with a given crime and the bottom row showing corresponding estimates among those found guilty. (For Study 3 we use only estimates from the condition in which the defendant’s race is unspecified.) When presenting distribution plots, we overlay the mean (dashed line) and median (dotted line).

¹⁵Unlike in Studies 1 and 2, in Study 3 we did not specify the punishment that would be given if found guilty, which should not affect those beliefs.

Figure 2: Beliefs about Guilt and Innocence of those Charged and Convicted by Crime: Studies 2 and 3



Race unspecified condition for Study 3

Looking across the top row, we see patterns similar to those from Study 1 for murder: More than half of respondents believe that at least 8 out of every 10 defendants are guilty. Averages range from 76 percent guilty (for Burglary) to 79 (for Armed Robbery). The 25th percentile of responses is 70 for all three crimes, implying that at least three-quarters or more of respondents believe that at least 7 in 10 of those charged with a crime are guilty.

The responses also suggest that respondents believe that trials reduce error rates at least somewhat. In all cases, the mean and median estimates of proportion guilty are larger for convicted (bottom row) than charged (top row) defendants. The median estimates imply that most respondents believe at least 9 out of every 10 persons convicted are actually guilty, an increase from between 8.5 and 9 out of every 10 among people charged. At the 75th percentile, among those

found guilty the estimates are 97, 99, and 96.5 percent, respectively, for Murder, Armed Robbery, and Burglary. We provide further analysis of changes in individual-level beliefs about guilt rates between those charging and found guilty in the Appendix C.

Racial Differences in Beliefs about Systemic Fallibility. We next examine differences across respondents in perceptions of the guilt of defendants in the criminal justice system. In light of the aforementioned prior scholarship finding stark racial differences in evaluations of the criminal justice system, we focus on difference between Black and White respondents in their average evaluations of errors in the criminal justice system, and all analysis of Black-White differences is restricted to respondents who identify as Black or White-alone.

Our first step is to regress respondent beliefs about the proportion of defendants who are actually guilty on an indicator for whether the respondent is Black, controlling for other demographic characteristics: Hispanic identity (non-exclusive), age (in years), educational attainment (categorical: baseline is high school or less, some college, college degree, post-BA), gender (categorical: baseline male, female, non-binary), and income (scale from 1-11, 6=refused, separate indicator for income refused). Our estimates therefore account for many average differences between Black and White individuals. Coefficient estimates appear in columns (1), (2), and (4) of Table 1 (for beliefs about guilt given charged) and Table 2 (for beliefs about guilt given convicted). Note that for purposes of comparability, this part of our analysis of responses from Study 3, presented in columns (2) and (4), uses only estimates from the defendant race-unspecified condition (defendant race was not presented in Study 2).

Across all crimes and stages of the criminal justice system, Black respondents systematically estimate that (race-unspecified) defendants are less likely to be guilty than do White respondents. In all cases, these estimates are negative and statistically significant ($p < .01$). The magnitudes of these effects are also large. For example, Black respondents report that those *convicted* of murder are, on average, guilty only about 67% (2/3) of the time compared to a sample average of 80% (4/5) for White respondents. Estimated racial differences are similar for conviction on armed robbery, but smaller for burglary. Comparing between Tables 1 and 2, we estimate larger (more negative) effects of respondent race for convicted than for charged defendants – approaching nearly twice as

Table 1: Defendant Race and Perceptions of Guilt: Charged Defendants

	(1) Murder	(2) Robbery	(3) Robbery	(4) Burglary	(5) Burglary
Respondent Black	-11.757 [2.496]***	-5.521 [2.581]**	-6.643 [2.567]***	-5.753 [2.560]**	-6.471 [2.526]**
Randomized defendant Black			-12.037 [1.522]***		-11.034 [1.581]***
Randomized defendant White			-4.128 [1.463]***		-2.644 [1.476]*
Resp. Black $\times$ Rand. def. Black			1.060 [3.942]		-2.844 [3.942]
Resp. Black $\times$ Rand. def. White			-2.630 [4.391]		-2.720 [4.483]
Respondent Hispanic (any race)	1.481 [2.717]	-4.475 [2.933]	0.129 [1.821]	-5.337 [2.968]*	-0.420 [1.872]
Age in years	0.231 [0.050]***	0.310 [0.051]***	0.250 [0.038]***	0.254 [0.056]***	0.238 [0.038]***
Educ. = some college	-2.262 [1.945]	0.447 [2.328]	0.404 [1.560]	2.428 [2.558]	0.723 [1.617]
Educ. = college	-2.405 [2.393]	4.863 [2.733]*	1.073 [1.776]	5.903 [2.875]**	1.780 [1.817]
Educ. = post-BA	-2.403 [2.987]	6.581 [2.929]**	-1.804 [2.306]	4.525 [3.572]	-2.789 [2.387]
Gender=Female	1.138 [1.678]	0.976 [1.707]	-1.480 [1.198]	0.232 [1.816]	-1.651 [1.219]
Gender=non-binary	-4.476 [5.195]	-11.513 [6.000]*	-7.801 [5.183]	-11.687 [8.182]	-9.539 [5.406]*
Income (1-11; 6=refused)	0.963 [0.387]**	0.055 [0.441]	0.760 [0.294]***	0.400 [0.467]	0.810 [0.291]***
Income refused	-9.191 [8.049]	10.702 [4.313]**	2.177 [3.597]	5.591 [5.212]	-0.760 [3.606]
Saw robbery first (before burglary)		2.585 [1.689]	1.571 [1.184]	0.117 [1.813]	-1.727 [1.194]
Constant	66.110 [3.575]***	60.974 [3.975]***	64.634 [2.881]***	60.677 [4.021]***	63.943 [2.897]***
Observations	810	554	1578	564	1586
R-squared	0.089	0.127	0.099	0.087	0.100
Race Unspecified Only	Yes	Yes	No	Yes	No
Sample mean	77.760	78.890	73.728	79.020	71.467
B-W difference among Whites			-7.909 [1.674]***		-8.390 [1.668]***
B-W difference among Blacks			-4.219 [4.385]		-8.513 [4.512]*

OLS coefficients with robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2: Defendant Race and Perceptions of Guilt: Convicted Defendants

	(1) Murder	(2) Robbery	(3) Robbery	(4) Burglary	(5) Burglary
Respondent Black	-13.301 [2.626]***	-10.434 [2.788]***	-11.233 [2.776]***	-9.771 [2.750]***	-10.632 [2.701]***
Randomized defendant Black			-11.041 [1.592]***		-9.155 [1.655]***
Randomized defendant White			-2.987 [1.503]**		-2.283 [1.569]
Resp. Black $\times$ Rand. def. Black			5.601 [4.096]		-1.311 [4.121]
Resp. Black $\times$ Rand. def. White			0.128 [4.365]		-1.045 [4.611]
Respondent Hispanic (any race)	1.367 [2.803]	-7.694 [3.193]**	-1.756 [1.875]	-10.173 [3.193]***	-3.993 [2.056]*
Age in years	0.233 [0.052]***	0.242 [0.054]***	0.230 [0.039]***	0.226 [0.059]***	0.218 [0.040]***
Educ. = some college	1.317 [2.033]	1.136 [2.561]	1.020 [1.612]	2.243 [2.758]	1.651 [1.694]
Educ. = college	0.980 [2.524]	5.215 [2.884]*	2.976 [1.790]*	6.160 [3.085]**	3.842 [1.914]**
Educ. = post-BA	-3.000 [3.407]	4.689 [3.215]	-0.509 [2.384]	7.259 [3.550]**	0.209 [2.424]
Gender=Female	0.779 [1.730]	1.019 [1.818]	-2.873 [1.229]**	-0.725 [1.943]	-2.453 [1.282]*
Gender=non-binary	3.074 [4.947]	-20.254 [6.686]***	-10.578 [5.756]*	-16.749 [7.461]**	-9.744 [5.371]*
Income (1-11; 6=refused)	1.000 [0.401]**	0.255 [0.470]	0.793 [0.303]***	0.301 [0.496]	0.651 [0.312]**
Income refused	0.941 [5.930]	3.007 [5.452]	1.530 [3.623]	2.836 [5.625]	-0.351 [3.702]
Saw robbery first (before burglary)		2.441 [1.832]	1.268 [1.208]	2.083 [1.939]	-0.716 [1.252]
Constant	66.228 [3.843]***	66.928 [3.957]***	68.438 [2.974]***	65.020 [4.418]***	67.299 [3.097]***
Observations	808	554	1583	564	1588
R-squared	0.096	0.133	0.099	0.121	0.096
Race Unspecified Only	Yes	Yes	No	Yes	No
Sample mean	80.260	80.640	76.661	78.010	74.286
B-W difference among Whites			-8.054 [1.701]***		-6.873 [1.741]***
B-W difference among Blacks			-2.582 [4.249]		-7.139 [4.563]

OLS coefficients with robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

large (in percentage point terms) for robbery and burglary.

Overall, these estimates imply much greater pessimism about the accuracy of the criminal justice system for Black relative to White respondents when a defendant’s race is unspecified.¹⁶

A concern that arises immediately in this analysis is that these observed group differences may arise because Black and White respondents do not have the same populations of defendants in mind when answering the survey. For example, perhaps Black respondents think about Black defendants when answering these questions while White respondents think about White defendants. To hold constant perceptions of the race of the defendants a respondent might consider, in Study 3 we randomized whether the race of the defendants in the pool of cases the respondent was asked to evaluate was provided and, if so, whether it was for Black or White defendants. We can therefore estimate how differences in defendant race affect beliefs about systemic fallibility and whether the effect of defendant race (Black versus White) varies with respondent race. To do so, we predict average perceptions of guilt using indicators for the race specified in the question (with race unspecified as the baseline category), and their interaction with the race of the respondent. Estimates employing this specification appear in columns (3) and (5) of Tables 1 and 2.

The coefficient estimates imply that when defendant race is specified, respondents tend to believe that defendants are more likely to be innocent than when race is unspecified. As we noted above, though, it could be that “race unspecified” summons different populations to the minds of respondents of different races. With this in mind, the two comparisons of greatest relevance are how White and Black respondents evaluate the likely guilt of pools of Black and White defendants, respectively. The Black/White defendant difference for White respondents is calculated by taking the difference between the “randomized defendant White” and “randomized defendant Black” coefficients. We report these estimates in at the bottom of each table in the row labeled “B-W difference among Whites.” Per column (3) in Table 1, the highly statistically significant difference of -7.909 implies that White respondents believe that Black defendants charged with robbery are nearly 8 percentage points *less* likely to be guilty than comparable White defendants. This finding—that White respondents believe Black defendants are more likely to be innocent than

¹⁶ Among the other demographic covariates we include in our analysis, only age consistently predicts perceptions, with older people reporting greater faith that defendants are actually guilty.

White defendants—replicates, both statistically and substantively, for those “convicted of robbery” and both burglary conditions.

To estimate the Black/White defendant difference for Black respondents, we calculate (randomized defendant Black + respondent Black $\times$ randomized defendant Black) - (randomized defendant White + respondent Black $\times$ randomized defendant White). We report these estimates at the bottom of each table in the row labeled “B-W difference among Blacks.” As with White respondents, Black respondents on average are more likely to believe a Black defendant is innocent than a White one. However, only one of four estimates is significant at conventional levels ($p = 0.06$). The noisier estimates for the Black/White difference for Black respondents likely result from smaller sample sizes. At the same time, we cannot reject the null hypothesis that Black and White respondents react similarly to information about defendant race (i.e., difference-in-differences estimates are never statistically significant). Substantively, for Burglary the estimates are nearly the same for Black and White respondents.

We note also that in the specifications that account for defendant race and the interaction of defendant and respondent race, we continue to estimate a negative and significant effect of respondent race. That is, apart from the effect of the defendant’s race and its interaction with the respondent’s own race, Black respondents remain significantly more pessimistic about defendant guilt across all defendants.¹⁷

Taken together, the results in this section suggest that compared to White respondents, Black respondents have much greater skepticism about the criminal justice system’s ability to correctly identify the guilty. But there is, simultaneously, a shared belief among Black and White respondents that, on average, criminal justice errors are more common for Black defendants than White ones. In fact, when White respondents are asked specifically to contemplate the guilt of Black defendants, their responses approximate those of Black subjects when race is unspecified. Comparatively speaking, Black respondents’ pessimism about criminal justice errors applies more comprehensively to defendants of different races, and somewhat more so when asked to focus on Black defendants.

¹⁷More formally, for any given defendant race condition (unspecified, White, or Black), the estimated effect of a respondent being Black rather than White is always negative and statistically significant.

4.2 Preferences about Just Punishment

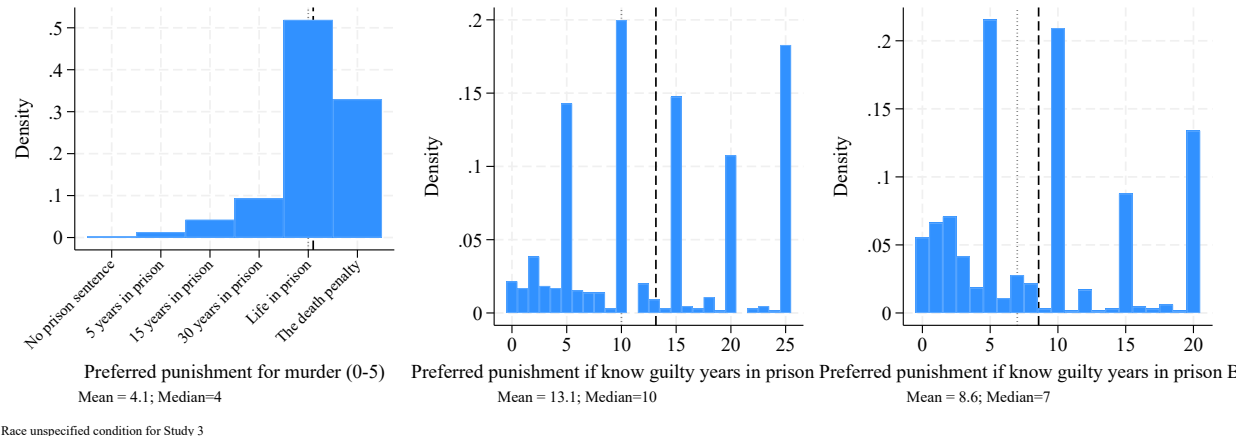
Next we examine preferences about appropriate punishments for individuals *certain to be guilty* of the three crimes introduce above: murder, robbery, and burglary. By fixing defendant guilt, we seek to address the concern that preferred sanctions for the convicted might reflect uncertainty about the actual guilt or innocence of defendants – uncertainty that we documented in the previous section and show is correlated with respondent race.

Survey Instruments. We focus here on data from Studies 2 and 3. In Study 2 we asked respondents, “Suppose that you know for certain that someone is guilty of premeditated murder. Which of the following do you think is the most appropriate punishment?” with response options listed as “No prison sentence”, “5 years in prison”, “15 years in prison”, “30 years in prison”, “Life in prison”, and “The death penalty”. In Study 3, we elicited preferred punishments for armed robbery and burglary (having provided the definitions for the crimes earlier in the survey – see above), including a randomization of defendant race, with questions worded as follows: “Suppose a [Black/White/“[BLANK]’] person is definitely guilty of [armed robbery/burglary]. By law, the maximum sentence is [25/20] years in prison. If they have no prior convictions, how many years in prison is the most appropriate punishment? Enter your answer as a whole number from 0 to [25/20]. If you think probation (with no prison term) is the appropriate punishment, please enter 0. (If you’d like to skip this question, please enter -99.)”. In both cases, we used text entry validation to require responses in the valid numerical range.

Aggregate Findings. We plot histograms of preferred punishments for each crime in Figure 3. As before, for Study 3 we use observations from the “race unspecified” condition to make it comparable to the estimates for Study 2. As the leftmost panel shows, there is relatively little disagreement about the appropriate punishment for someone guilty of murder: The modal response is life in prison, selected by about half of respondents. Approximately 30% of respondents prefer the death penalty, and 10% prefer a sentence of 30 years.¹⁸ By contrast, the two right panels show much

¹⁸Note that our question asks for the “appropriate punishment” for someone guilty of murder, whereas the GSS asks whether the subject “favors or opposes” the death penalty for those convicted of murder. Given the difference

Figure 3: Preferred Punishment for those certainly guilty by Crime: Studies 2 and 3



greater differences in preferred sanctions for first-time convictions for armed robbery and burglary. For example, for robbery, almost as many people prefer a sentence of 25 years (18%) as prefer a sentence of 10 years (20%), and about 15% of people each prefer a sentence of 5 and 15 years. For burglary, there are more people who prefer sentences of less than 5 years (about 25% of the sample rather than 11% for robbery), but 25% of respondents also prefer a sentence of at least 15 years in prison.

Racial Differences in Preferred Punishments We next examine heterogeneity by respondent race in preferred punishments for certainly guilty defendants, as well as by (randomized) defendant race in Study 3. Our modeling approach parallels our above statistical specification. The dependent variable for murder is categorical and ranges from zero (no prison sentence) to 5 (death penalty). The dependent variable for the robbery regressions ranges from 0 to 25 years, and from 0 to 10 years for burglary.

We focus first on the specifications in columns (1), (2), and (4), corresponding to conditions in which the race of the defendant was unspecified. Our analysis reveals that Black respondents have modestly less punitive punishment preferences for murder – approximately 0.2 scale units less

in wording, it would be perfectly coherent for a subject to simultaneously express support for the death penalty on the GSS item and a most preferred sentence less severe on ours (even setting aside the guilty/convicted distinction discussed above).

Table 3: Predicting Preferred Punishments Given Certain Guilt

	(1) Murder (0-5)	(2) Robbery (0-25yrs)	(3) Robbery (0-25yrs)	(4) Burglary (0-20yrs)	(5) Burglary (0-20yrs)
Respondent Black	-0.189 [0.077]**	-1.247 [0.838]	-0.998 [0.803]	0.201 [0.686]	0.321 [0.648]
Randomized defendant Black			-0.370 [0.518]		0.060 [0.407]
Randomized defendant White			-0.271 [0.508]		0.241 [0.402]
Resp. Black $\times$ Rand. def. Black			-0.905 [1.142]		-0.909 [0.917]
Resp. Black $\times$ Rand. def. White			1.634 [1.212]		0.122 [0.962]
Respondent Hispanic (any race)	-0.154 [0.107]	-0.087 [1.007]	0.375 [0.616]	0.115 [0.819]	0.048 [0.484]
Age in years	0.011 [0.002]***	0.016 [0.021]	0.034 [0.012]***	0.013 [0.015]	0.016 [0.009]*
Educ. = some college	-0.040 [0.070]	0.638 [0.860]	-0.236 [0.496]	-0.459 [0.671]	-0.546 [0.394]
Educ. = college	-0.156 [0.088]*	0.191 [0.985]	-0.567 [0.572]	-0.895 [0.794]	-1.133 [0.463]**
Educ. = post-BA	-0.176 [0.108]	-0.463 [1.291]	-1.439 [0.734]**	-0.438 [1.047]	-1.095 [0.602]*
Gender=Female	0.008 [0.059]	1.757 [0.664]***	0.992 [0.386]**	1.398 [0.521]***	1.239 [0.306]***
Gender=non-binary	-0.422 [0.223]*	-5.413 [2.027]***	-4.105 [1.429]***	-3.183 [1.522]**	-3.619 [1.032]***
Income (1-11; 6=refused)	0.010 [0.014]	0.167 [0.167]	0.268 [0.097]***	0.000 [0.144]	0.146 [0.079]*
Income refused	0.055 [0.189]	2.997 [1.912]	1.256 [1.131]	2.460 [1.480]*	1.672 [0.954]*
Saw robbery first (before burglary)		-0.939 [0.653]	-0.498 [0.379]	-3.656 [0.509]***	-3.087 [0.301]***
Constant	3.670 [0.119]***	11.208 [1.553]***	10.683 [0.934]***	9.441 [1.218]***	8.724 [0.748]***
Observations	859	564	1628	564	1617
R-squared	0.080	0.034	0.027	0.103	0.083
Race Unspecified Only	Yes	Yes	No	Yes	No
Sample mean	4.120	13.080	12.929	8.590	8.595
B-W difference among Whites			-0.099 [0.511]		-0.182 [0.412]
B-W difference among Blacks			-2.637 [1.121]**		-1.213 [0.889]

OLS coefficients with robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

punitive ($p < 0.01$). In an analysis presented in Appendix C, we demonstrate this racial difference is driven by opposition to the death penalty among Black respondents.¹⁹ For robbery and burglary, the sign of the coefficient for respondent race fluctuates and it is nowhere close to statistically significant at conventional levels.²⁰

When we examine the effect of the randomized defendant race in isolation in Study 3 (columns (3) and (5), respectively, for robbery and burglary), we find very little evidence that defendant race on average affects preferred punishments for certainly guilty defendants. All of the coefficients in the main specification are statistically insignificant, suggesting that specifying defendant race does not affect judgments of preferred sentences relative to the baseline in which defendant race was unspecified. Turning to the Black/White defendant differences by respondent race, we observe no difference among White respondents in preferred sentences for Black versus White defendants. Black respondents are estimated to prefer sentences 2.6 years lower for Black than White defendants definitely guilty of robbery ($p < .05$). The analogous Black/White difference for robbery is small and statistically insignificant.

Overall, this portion of our analysis shows that, unlike with perceptions of the reliability of the criminal justice system, there are not stark and persistent racial differences in punitiveness conditional on knowing a defendant is guilty. The exception is lower Black support for the death penalty. Moreover, the race of a defendant appears to have little effect on preferred sanctions for guilty defendants, and those effects are not consistently different across individuals of different races.

4.3 The Burden of Proof

The last set of novel questions that we posed elicited preferences regarding the burden of proof to convict.

¹⁹When we replicate our analysis using a dichotomous indicator equal to one if the respondent favors the death penalty for a defendant certainly guilty of murder, and zero otherwise, we find that Black respondents are 14 percentage points less likely to report a preference for the death penalty ($p < 0.01$). By contrast, if we top-code preferred punishments by combining the life in prison and death penalty cases, we no longer estimate a significant Black respondent effect.

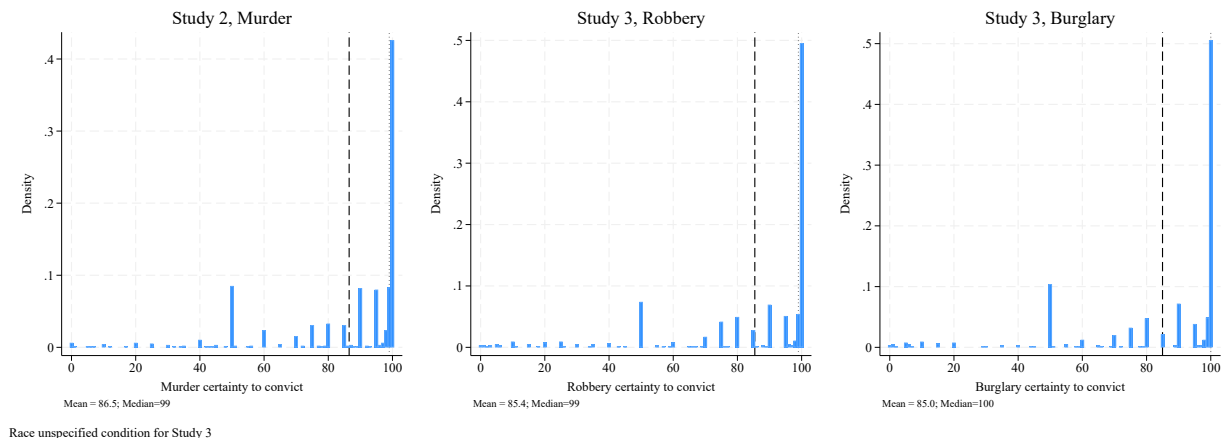
²⁰We also found that female respondents supported harsher sentences for defendants certainly guilty of robbery or burglary. Non-binary respondents reported favoring more lenient sentences, but given the small sample (20 subjects), we are hesitant to attach too much weight to that result.

Survey Instruments. Here we draw on Studies 2 and 3.²¹ In Study 2, we specified that if a person on trial for murder is convicted, they would receive a sentence of life in prison. Half of subjects received this question: “Think about 100 people on trial for murder. Suppose some of these 100 people are guilty and the remainder are innocent. You are the juror in this case and are given a choice: either convict all 100 people or let them all go free. What is the smallest number of people who would have to be GUILTY for you to convict all 100 people?” The other half received a question couched in terms of beliefs about the guilt of a single suspect: “Think about someone on trial for murder. You are the juror in this case and are given a choice: either convict or let them go free. How certain would you have to be of their guilt to convict them? Enter your answer as a whole number from 0 to 100, where 100 means you would only convict if you were certain he was guilty (100% chance of guilt), 50 means you would convict if he was equally likely to be guilty or innocent (50% chance of guilt), and 0 means that you would you would convict even if you were certain he was innocent (0% chance of guilt).”

In Study 3, we used the “belief about a single individual” wording embedded in two survey vignettes that varied the race of the suspect and the sentence at stake. The first vignette described an alleged armed robbery, beginning with “[*Name*], a 22-year old [*Black/White/[BLANK]*] man, is charged with [**burglary/armed robbery**]. He has no prior criminal convictions. He is accused of ...” The armed robbery vignette continues, “[c]onfronting another man from his neighborhood on the street at 8pm on a Tuesday, pointing a gun at him, and stealing about \$2,500 in jewelry and cash. The victim was not injured.” The burglary vignette continues, “[b]reaking into an apartment in his neighborhood at 2am while carrying a knife and stealing about \$2,000 in electronics and cash. The resident of the apartment, a 45-year old man, was asleep in the bedroom and did not awaken during the break in.” The subject was told that “If convicted, [*Name*] will receive a sentence of [s^- , $\hat{s}$, s^+],” where the assigned sentences were slightly below, at, or slightly above the subject’s

²¹In Study 1, we fielded a question about the burden of proof posed as follows: “Think about 100 people on trial for murder, that is unlawfully killing a person. If found guilty, the sentence is life in prison. How many of the 100 people would have to be [*guilty/innocent*] rather than [*innocent/guilty*] before you would [*convict/release*] all 100 people?” Based on internal analysis of these data, we determined that the “Innocent/Guilty/Release” wording confused subjects relative to the more intuitive “Guilty/Innocent/Convict” framing, so we used the latter framing in subsequent studies.

Figure 4: Elicited Thresholds to Convict: Studies 2 and 3



elicited preferred punishment conditional on guilt.²²

Aggregate Findings. Figure 4 displays histograms for three of the conditions that are directly comparable: the “single individual” framing for murder in Study 2 and the “race unspecified” conditions for robbery and burglary in Study 3. The mean burden of proof to convict across crimes lies between 85 and 87, but there is substantial bunching at the top end of the range: a majority of respondents reported a burden of proof of 99 percent certainty for murder and robbery, while a majority reported that 100 percent certainty was necessary to convict for burglary.²³

Racial Differences in the Burden of Proof. We present regression analysis of the willingness to convict for murder using data from Study 2 and robbery and burglary from the Study 3 vignettes in Table 4. For answers to the murder question, we separate out the “per 100” framing in column (1). (The remaining columns are all using data from the “single defendant” framing was employed.) For comparability the outcome variable is always expressed out of 100. Of interest, on average

²²Specifically, if the subject’s preferred punishment $\hat{s}$ was neither zero nor the maximum sentence m , the randomization was over $\hat{s}$, $\hat{s} - d$, and $\hat{s} + d$, where $d = \min\{5, \hat{s}, m - \hat{s}\}$. If $\hat{s}$ was zero, the randomization was between 0 and 1. If $\hat{s}$ was the maximum sentence m , the randomization was between $m - 5$ and m .

²³We are reluctant to overinterpret the latter finding: subjects could simply be using 100 to quantify an informal notion of “really, really sure” or just rounding up. It need not imply that they would need metaphysical certainty to convict (an impossible standard for any jury). With that said, because the Blackstone ratio formulation of the burden of proof is undefined for subjects who answer 100, we retain our focus on the reported probability in what follows.

Table 4: Predicting Thresholds to Convict: Studies 2 and 3

	(1) Murder (/100)	(2) Murder (/1)	(3) Robbery (/1)	(4) Robbery (/1)	(5) Burglary (/1)	(6) Burglary (/1)
Respondent Black	-7.190 [2.872]**	-7.702 [2.304]***	-8.367 [3.344]**	-8.382 [3.255]**	-4.697 [3.253]	-6.101 [3.141]*
Vignette defendant Black				-0.312 [1.498]		-1.047 [1.592]
Vignette defendant White				-0.815 [1.522]		-0.969 [1.542]
Respondent Black x Vignette Black				0.471 [4.532]		-0.796 [4.466]
Respondent Black x Vignette White				6.337 [4.455]		2.532 [4.521]
Robbery sentence above preferred			4.580 [2.412]*	-0.623 [1.466]		
Robbery sentence below preferred			0.335 [2.421]	-2.061 [1.419]		
Burglary sentence above preferred					-0.797 [2.454]	1.331 [1.504]
Burglary sentence below preferred					-0.289 [2.463]	0.452 [1.472]
Respondent Hispanic (any race)	-0.947 [3.620]	-1.951 [2.651]	-4.746 [3.366]	-0.345 [1.872]	-4.092 [3.326]	-0.416 [1.849]
Age in years	0.119 [0.069]*	0.180 [0.046]***	0.193 [0.058]***	0.243 [0.038]***	0.180 [0.064]***	0.177 [0.041]***
Educ. = some college	0.104 [2.813]	1.159 [1.971]	8.168 [2.984]***	5.706 [1.681]***	1.225 [2.722]	4.817 [1.707]***
Educ. = college	0.203 [3.455]	0.907 [2.481]	11.047 [3.262]***	7.910 [1.815]***	3.950 [2.948]	5.636 [1.857]***
Educ. = post-BA	-5.241 [4.669]	2.782 [2.860]	9.286 [3.902]**	3.349 [2.554]	4.441 [3.647]	3.177 [2.490]
Gender=Female	0.236 [2.261]	3.304 [1.585]**	-2.347 [1.991]	-0.593 [1.231]	-0.507 [2.069]	-0.315 [1.253]
Gender=non-binary	7.778 [6.849]	1.420 [4.929]	7.054 [5.298]	10.610 [3.781]***	11.035 [5.669]*	8.112 [4.255]*
Income (1-11; 6=refused)	1.104 [0.557]**	-0.058 [0.399]	-0.073 [0.495]	-0.226 [0.295]	0.665 [0.469]	-0.022 [0.298]
Income refused	-4.564 [7.332]	-6.651 [6.525]	1.322 [6.134]	-1.309 [4.417]	-5.781 [5.792]	-0.622 [3.775]
Asked certainty about 100 before about 1	-5.131 [2.234]**	-1.581 [1.528]				
Saw robbery first (before burglary)			3.495 [2.002]*	0.871 [1.201]	5.976 [2.077]***	0.654 [1.233]
Constant	68.314 [4.733]***	77.331 [3.245]***	69.237 [4.740]***	71.854 [3.117]***	70.999 [5.388]***	73.354 [3.371]***
Observations	765	818	552	1581	552	1573
R-squared	0.030	0.057	0.093	0.064	0.063	0.037
Race Unspecified Only	Yes	Yes	Yes	No	Yes	No
Sample mean	74.050	86.310	85.430	85.473	85.380	84.835
B-W difference among Whites				0.503 [1.558]		-0.077 [1.628]
B-W difference among Blacks				-5.363 [4.172]		-3.406 [4.308]

OLS coefficients with robust standard errors in brackets
 * significant at 10%; ** significant at 5%; *** significant at 1%

respondents require a higher burden of proof to convict a single defendant than a pool of 100 persons: The average number guilty to convict 100 is 74 out of 100 (.74), but it is 86 out of 100 (.86) for the individual framing. We speculate this may be due to the greater salience of releasing large numbers of guilty people.

Beyond this intriguing framing difference, the results are remarkably consistent across model specifications and measures: Relative to White respondents, Black respondents expressed a lower threshold to convict a defendant than White respondents. Focusing on the “per 100” framing for murder (and with an assumed life sentence), Black respondents expressed a threshold to convict

7.2 points lower than White respondents. The effect persists when the prompt concerns certainty about convicting a single defendant for murder (column (2)), and for the robbery vignettes (column (3)). By contrast, we do not observe a significant relationship between respondent Race and burden of proof for burglary (column (5)), although the sign of the effect is in the same direction as for murder and robbery. On average then, Black respondents are more tolerant of criminal justice errors than are White respondents.

Note that our theoretical model implies that, *ceteris paribus*, respondents whose preferred sentences conditional on guilt fall below the anticipated sentence will impose higher thresholds to convict. This relationship frames our interpretation of the observed racial difference. Recall the finding above that Black respondents preferred lower sentences for murder than their White counterparts, closer on average to the life sentence presented here. The conjunction of that finding and the current one implies that the observed difference in thresholds is not an artifact of preferring less severe punishments, but rather on the relative weights respondents implicitly assign to Type I and Type II errors.

We evaluated the *ceteris paribus* claim by randomizing the hypothetical anticipated sentence in the Study 3 vignettes to be at, slightly above, or slightly below what the respondent preferred. In the robbery vignette, we detect some evidence that an assigned sentence above the one preferred by the subject increased the threshold to convict (by 4.6 percentage points), but this was only significant at the 10 percent level and did not replicate in the burglary vignette. By contrast, we observe no significant difference in reported thresholds when the subject was told the defendant would be receiving a sentence below the one the subject most preferred. Taken together, these results suggest that, at least for these crimes and ranges of sentences, the burden of proof is not systematically affected by (moderate) deviations from a respondent’s own preferred sentence.²⁴

Next, we turn to the results on randomized defendant race from Study 3. The relevant regression coefficients appear in columns (4) and (6) of Table 4. The coefficients on the first order and interaction terms are never significant and are inconsistently signed, providing initial evidence against

²⁴Two other demographic findings of note in the regression are (1) a robustly significant effect of age, with older respondents reporting higher burdens of proof to convict across crimes; and (2) a positive relationship between educational attainment and the threshold to convict.

the hypothesis that defendant race affects the applied burden of proof compared to the unassigned race baseline. Direct tests of racial differentiation across defendants in assigned threshold by race of respondent appear at the bottom of the table. For White respondents, the observed difference is small in magnitude and statistically insignificant. For Black respondents, we see a surprisingly large *negative* value for the difference, which would imply that on average, these respondents assigned a *lower* burden of proof to convict Black than White defendants. However, these estimates are imprecisely estimated and come nowhere close to statistical significance at conventional levels.

Overall, what emerges from this examination of the burden of proof is a surprising pattern in which Black respondents are, on average, more tolerant of errors in the criminal justice system than are White Americans, *despite* having lower trust in the reliability of the system and lower levels of punitiveness (mostly concerning the death penalty for murder). Defendant race, on average, does not have large effects on this burden of proof, nor do we see either Black or White respondents varying their burden of proof by the race of the defendant. To the best of our knowledge, this is the first study to directly interrogate survey respondents about their preferred burden of proof and measure how this burden changes with features of crimes, potential punishments, and defendant race.

4.4 The Dimensionality of Punitiveness

Earlier in this paper, we presented a model that shows how the three considerations pertaining to beliefs and preferences about criminal justice that we have focused on are analytically distinct and argue that they may be conflated in standard survey items. In this section, we test whether what we established theoretically is borne out empirically. To do so, we draw on data from Study 3, where prior to asking our novel questions we asked several of the traditional criminal-justice related questions discussed above.

As a first step, we conduct a principal factor analysis using our questions about beliefs about guilt among the charged and convicted for armed robbery and burglary; preferred punishments for those crimes; and thresholds to convict for those crimes.²⁵ Table 5 displays the orthogonally

²⁵These scores were estimated in Stata 18.0 using the factor command.

Table 5: Rotated Factor Loadings for Study 3 Novel Survey Measures

	Component 1 Beliefs about Prob. Guilty	Component 2 Preferred Sentences if Guilty	Component 3 Conviction Threshold
Prob. guilty given charged, robbery	0.8434	0.1027	0.2047
Prob. guilty given charged, burglary	0.8449	0.1033	0.1747
Prob. guilty given convicted, robbery	0.8548	0.0597	0.2184
Prob. guilty given convicted, robbery	0.8515	0.0395	0.2040
Preferred punishment, robbery	0.1094	0.7741	-0.0247
Preferred punishment, burglary	0.0445	0.7718	-0.0406
Certainty to convict, robbery	0.1948	-0.0441	0.7093
Certainty to convict, burglary	0.2215	-0.0283	0.7117

rotated factor loadings for the first three dimensions. (A Scree plot appears in Appendix C, and depicts a substantial dropoff in eigenvalues after the first three retained factors.) As is evident from the table, all the fallibility measures load strongly onto the first component; the preferred punishments onto the second; and conviction thresholds onto the third. For simplicity, we label each factor along these lines. Importantly, these factors also measure distinct variance, meaning responses to one factor do not entirely predict responses to the others.

Next, we used regression analysis to predict answers to a series of traditional crime and punishment opinion measures using these three retained factor scores. The traditional questions we asked are the four discussed in Section 3, as well as a question about confidence in police and law enforcement and a question about prison terms for selling marijuana. If two or more factors predict answers to a standard measure, we can empirically corroborate our theoretical claim that the measure is conceptually ambiguous.

Results of our regression analysis appear in Table 6. *At least* two out of three factors are statistically significant predictors of all six traditional measures. The first factor, a composite measure of beliefs about defendant guilt, proxies $p_{i,d}$ in the theoretical model. It is a strong positive predictor of all of the traditional measures except the general belief that it is worse to convict the innocent than acquit the guilty.

The second factor is a composite of responses to preferred punishment given guilt questions,

Table 6: Regression Results: Our Items vs. Traditional Items

	(1)	(2)	(3)	(4)	(5)	(6)
	Confidence courts/legal system (0-3)	Confidence police/law enforce. (0-3)	System tough (2=Not tough, 0=too tough)	Prison selling marij. (2=yes, (1=dep.,0=no)	Death Penalty Murder (2=yes, (1=dep., 0=no)	Worse to Convict Innocent (2=yes, 1=not sure,0=no)
Factor 1: Beliefs about Prob. Guilty	0.074 [0.021]***	0.143 [0.023]***	0.094 [0.019]***	0.050 [0.018]***	0.122 [0.018]***	0.013 [0.020]
Factor 2: Preferred Sentences if Guilty	0.013 [0.024]	0.135 [0.026]***	0.247 [0.020]***	0.139 [0.020]***	0.231 [0.019]***	-0.148 [0.023]***
Factor 3: Conviction Threshold	0.048 [0.027]*	0.067 [0.028]**	0.011 [0.025]	-0.024 [0.022]	-0.044 [0.021]**	0.141 [0.026]***
Constant	1.489 [0.019]***	1.674 [0.021]***	1.439 [0.017]***	0.534 [0.017]***	1.177 [0.017]***	1.540 [0.019]***
Observations	1733	1732	1733	1732	1732	1733
R-squared	0.011	0.046	0.093	0.034	0.101	0.047
Sample mean	1.489	1.674	1.439	0.535	1.177	1.540

OLS coefficients with robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

and corresponds to $\hat{s}_i$ in the theoretical model. Increases in this measure are strongly predictive of confidence in law enforcement, a belief that the system is not tough enough, a belief that marijuana distribution justifies imprisonment, and support for the death penalty. It is a strong negative predictor of the belief that it is worse to convict the innocent than acquit the guilty, but is not a significant predictor of confidence in courts and the legal system.

The third factor is our measure of the burden of proof, $\hat{p}_i$ or $\hat{R}_i$ in the model. This factor is strongly and positively associated with confidence in the police and law enforcement and the traditional “worse to convict the innocent” measure. A higher threshold to convict is also associated with decreased support for the death penalty. It does not predict either a belief that the criminal justice system is insufficiently tough, nor is it a significant predictor of attitudes about punishments for marijuana distribution. And it is only a weak predictor of confidence in the courts and the legal system generally.

In Appendix C, we report the results of three robustness checks of these results. First, in lieu of factor analysis, we replicated our regression results using standardized means for the fallibility, preferred punishment, and conviction threshold questions. Second, to guard against the possibility that subjects who reflexively enter “50” (the scale midpoint) into their probability and threshold questions are driving the observed correlations, we replicated our analysis omitting all respondents who entered more than one 50 in their responses. Finally, one might object that the historical legacy of racism in criminal justice policy will lead Black respondents to have substantially different mental schema of the connections between systemic fallibility, appropriate punishment, and the burden of proof. To guard against the possibility that this might bias the results reported here, we replicated our analysis restricting the sample to non-Black respondents. For all three robustness checks, the results were very similar to the ones reported here.

All told, these results empirically confirm our underlying argument about the ambiguity of existing survey items and demonstrate the value of asking questions about the more basic considerations as we have done above.

5 Discussion

In this paper, we have sought to make a conceptual contribution, a methodological contribution, and an empirical contribution. The conjunction of these has important implications for how we understand responsiveness and disagreement in the domain of criminal justice politics.

The conceptual contribution is to offer a microfounded account of three analytically distinct dimensions of criminal justice attitudes: beliefs about system fallibility (as encapsulated in beliefs that defendants and those convicted may be innocent), sentencing preferences conditional on certain guilt, and preferred burdens of proof. We demonstrate how these considerations enter into a model of decision making by a finder of fact in a criminal proceeding, detail their contingency and how they interact, and describe how they translate to a survey environment in which the aim is to elicit attitudes about crime and punishment generally.

The methodological contribution is to describe how traditional survey questions suffer from conceptual ambiguity because they conflate these three underlying dimensions, and as a consequence may obscure evidence of responsiveness and the underlying sources of group differences. Additionally, in light of our results, we offer a framework for asking questions to elicit more fundamental beliefs and attitudes about crime and punishment.

The combination of these conceptual and methodological contributions leads to a prescription for survey researchers: a necessary condition for drawing clear-cut inferences about public preferences and disagreement, and for assessing the extent of democratic responsiveness in light of those, is to field survey items that assess respondents' underlying fundamental values and beliefs regarding the subject matter of interest, rather than items for which responses might reflect multiple distinct considerations (Lacy, 2001). We show that theory can be a valuable tool for identifying the former and assessing its relationship to the latter. While we have explicated this specifically with reference to attitudes and beliefs about criminal justice, this is a general point that applies to opinion research in other policy areas as well. For example, attitudes about social welfare programs in general may reflect beliefs about the true merits of those who seek benefits (i.e., deservingness, which is similar to error rates in the criminal justice system) as well as preferences about appropriate support even

conditional on having a meritorious claim.

The empirical contribution is twofold. First, we demonstrate that well-documented gaps in “top-level” punitiveness between Black and White Americans might be misleading. With the exception of attitudes about the death penalty for murder, the apparent lower punitiveness of Black respondents to traditional survey items appears to be largely driven by greater pessimism about the fallibility of the criminal justice system generally, rather than an intrinsic aversion to harsh sentences (except the death penalty for murder) or a demand for higher burdens of proof. We observe that this pessimism extends to White respondents when it comes to assessing systemic treatment of Black defendants. Our analysis suggests, further, that Black respondents have lower burdens of proof on average, and that neither Black nor White respondents appear to condition either their preferred sentences or burdens of proof on defendant race.

Second, our analysis corroborates the theoretical claim that common survey items related to criminal justice may be misleading. Using factor analysis, we demonstrate that our novel survey items map to three distinct dimensions corresponding to systemic trust, preferred sentencing for the guilty, and burden of proof. Subsequent regression analyses corroborate the ambiguity issue: responses to each of five commonly used survey questions are predicted by at least two of the three underlying factors, meaning none of the existing items map cleanly to a single theoretically distinct dimension.

Substantively, our findings have implications for our interpretation of group differences in criminal justice attitudes. Our null findings about defendant race on preferred punishments and thresholds for conviction contradict prominent claims that Black-White differences in criminal justice attitudes are primarily driven by greater racial resentment among Whites toward Black defendants. What we find is at stake is not beliefs about defendants of different races, but instead the belief among Black Americans that criminal justice errors are more common and a shared, cross-racial view that Black defendants are more likely to be innocent than White ones. One need not take a position on the veracity or inaccuracy of these beliefs to identify their political implications, particularly with respect to the perceived legitimacy (or lack thereof) of governing institutions.²⁶

²⁶Note that we do not test whether White respondents who vary in their resentment have different beliefs and preferences across the dimensions we measure directly, but this is a promising avenue for subsequent work. So too

It would be highly useful as a next step to consider extending these measures of beliefs to other aspects of the criminal justice system, for example to understand average and group-level differences in beliefs about the rate at which the criminal justice system fails to apprehend the guilty.

Finally, by disaggregating the sources of disagreement about criminal justice policy to their more elemental components, we are better positioned to map the relationship between attitudes and policy – a necessary step for assessing the responsiveness or congruence of the latter to the former. Consider some prominent proposed and enacted reforms – past and present – in the criminal justice space:

- Three-strikes laws
- Sentencing reform and decriminalization
- Changes in charging standards for low-level crimes and first-time offenses (e.g., by progressive prosecutors)
- Moving from a preponderance of evidence standard to a reasonable doubt standard for asset forfeiture
- Civilian police review boards
- Body worn cameras

Our analysis suggests caution in describing these moves as occurring in the direction of more or less punitiveness, and of benchmarking them against aggregate changes in public preferences. Instead, we counsel consideration of the relationship between policy change and the more closely associated underlying value(s). For example, the relevant attitudes for understanding support for three-strikes laws, sentencing reform, and changes in charging may include beliefs about both culpability and appropriate sentences, which implies complications for understand how representative policy is along at least two dimensions at once. For changes in rules about asset forfeiture standards, the

is a consideration of what other factors, like partisanship, ideology, personality, and past experience with crime and the criminal justice system shape these beliefs and preferences. Crucially, it is important to assess the correlation between these measures and beliefs about the different factors we measure (beliefs about guilt, preferred sanctions if guilty, and burdens of proof) both on average and for defendants of different races.

burden of proof is most clearly implicated. Finally, efforts to reduce criminal justice “errors” such as civilian review boards and body cameras are most likely related to beliefs about culpability, and thus systemic fallibility.²⁷

If policy outcomes are more clearly linked to relevant beliefs and attitudes, the result will be an understanding of the subtleties of democratic accountability and criminal justice far more comprehensive than what has been established to date.²⁸ And while our substantive focus is criminal justice, this points extend to a far broader range of policy domains.

²⁷By contrast, evaluations of surveillance technology such as Shot-Spotter may respond to beliefs about the incidence of different kinds of criminal justice errors – the failure to apprehend guilty individuals rather than the failure to exculpate innocent ones.

²⁸Importantly, without measurement of opinions relevant for specific policy reforms, establishing a (lack of a) correlation between the two does not constitute an informative test of theories of representation.

References

- Adriaenssen, A. and I. Aertsen (2015). Punitive attitudes: Towards an operationalization to measure individual punitivity in a multidimensional way. *European Journal of Criminology* 12(1), 92–112.
- Andreoni, J. (1991). Reasonable doubt and the optimal magnitude of fines: Should the penalty fit the crime? *The RAND Journal of Economics* 22(3), 385–395.
- Arnold, R. D. and N. Carnes (2012). Holding mayors accountable: New york’s executives from koch to bloomberg. *American Journal of Political Science* 56(4), 949–963.
- Bierschbach, R. A. and A. Stein (2005). Overenforcement. *Georgetown Law Journal* 93(6), 1743–1781.
- Blackstone, W. (1979 [1765]). *Commentaries on the Laws of England*. Chicago, IL: University of Chicago Press.
- Cochran, J. K. and M. B. Chamlin (2006). The enduring racial divide in death penalty support. *Journal of Criminal Justice* 34(1), 85–99.
- Connolly, T. (1987). Decision theory, reasonable doubt, and the utility of erroneous acquittals. *Law and Human Behavior* 11(2), 101–112.
- Dardis, F. E., F. R. Baumgartner, A. E. Boydstun, S. De Boef, and F. Shen (2008). Media framing of capital punishment and its impact on individuals’ cognitive responses. *Mass Communication and Society* 11(2), 115–140.
- Dickson, E. S., S. C. Gordon, and G. A. Huber (2022). Identifying legitimacy: Experimental evidence on compliance with authority. *Science Advances* 8(7), eabj7377.
- Easton, D. (1965). *A Systems Analysis of Political Life*. New York: Wiley.
- Enns, P. K. (2014). The public’s increasing punitiveness and its influence on mass incarceration in the united states. *American Journal of Political Science* 58(4), 857–872.
- Enns, P. K. (2016). *Incarceration Nation: How the United States Became the Most Punitive Democracy in the World*. Cambridge University Press.
- French, J. R. and B. Raven (1959). The bases of social power. In D. Cartwright and A. Zander (Eds.), *Studies in Social Power*, pp. 150–167. Ann Arbor, MI: University of Michigan Press.
- Frost, N. A. (2010). Beyond public opinion polls: Punitive public sentiment & criminal justice policy. *Sociology Compass* 4(3), 156–168.
- Garland, D. (2001). *The Culture of Control: Crime and Social Order in Contemporary Society*. Chicago: University of Chicago Press.
- Gordon, S. C. and G. A. Huber (2002). Citizen oversight and the electoral incentives of criminal prosecutors. *American Journal of Political Science* 46(2), 334–351.

- Gordon, S. C. and G. A. Huber (2019, 08). The empirical study of legitimate authority normative guidance for positive analysis. In J. Knight, M. Schwartzberg, J. Knight, and M. Schwartzberg (Eds.), *Political Legitimacy: NOMOS LXI*. NYU Press.
- Gottschalk, M. (2014). *Caught: The Prison State and the Lockdown of American Politics*. Princeton, NJ: Princeton University Press.
- Kaplan, J. (1968). Decision theory and the factfinding process. *Stanford Law Review* 20, 1065.
- Kaplow, L. and S. Shavell (1994). Accuracy in the determination of liability. *The Journal of Law and Economics* 37(1), 1–15.
- Kugler, M. B., F. Funk, J. Braun, M. c. Gollwitzer, A. C. Kay, and J. M. Darley (2013). Differences in punitiveness across three cultures: A test of american exceptionalism in justice attitudes. *Journal of Criminal Law and Criminology* 103(4), 1071–1113.
- Lacy, D. (2001). A theory of nonseparable preferences in survey responses. *American Journal of Political Science*, 239–258.
- Leventhal, G. S. (1980). What should be done with equity theory? In K. J. Gergen, M. S. Greenberg, and R. H. Willis (Eds.), *Social Exchange: Advances in Theory and Research*, pp. 27–55. Boston, MA: Springer.
- Matthews, R. (2005). The myth of punitiveness. *Theoretical Criminology* 9(2), 175–201.
- Milanich, P. G. (1981). Decision theory and standards of proof. *Law and Human Behavior* 5(1), 87–96.
- Miller, L. L. (2016). *The Myth of Mob Rule: Violent Crime and Democratic Politics*. New York: Oxford University Press.
- Mullinix, K. J., N. Brophy, R. J. Norris, and K. Cornell (2026). The punitiveness of the mass public: Lessons from exploratory studies. *Social Science Quarterly* 107(4), e70177.
- Nagel, S., D. Lamm, and M. Neef (1981). Decision theory and juror decision-making. In B. D. Sales (Ed.), *Perspectives in Law and Psychology, Volume 2: The Trial Process*, pp. 353–386. New York, NY: Plenum Press.
- Nicholson-Crotty, S., D. A. Peterson, and M. D. Ramirez (2009). Dynamic representation(s): Federal criminal justice policy and an alternative dimension of public mood. *Political Behavior* 31(4), 629–655.
- Orr, L. V., J. J. Cho, and G. A. Huber (2026). The challenge of estimating percentages: Reassessing the origins and meaning of apparent partisan stereotypes. Typescript: Yale University.
- Peffley, M. and J. Hurwitz (2010). *Justice in America: The Separate Realities of Blacks and Whites*. New York: Cambridge University Press.
- Pew Research Center (2021, June). Most americans favor the death penalty despite concerns about its administration. Technical report, Pew Research Center.

- Pickett, J. T. (2019). Public opinion and criminal justice policy: Theory and research. *Annual Review of Criminology* 2(Volume 2, 2019), 405–428.
- Pickett, J. T. and B. Thomas (2014). The pragmatic american: Empirical reality or methodological artifact? *Criminology* 52(2), 195–222.
- Polinsky, A. M. and S. Shavell (2000, March). The economic theory of public enforcement of law. *Journal of Economic Literature* 38(1), 45–76.
- Ramirez, M. D. (2013). Punitive sentiment. *Criminology* 51(2), 329–364.
- Roberts, J. V. (2003). *Penal Populism and Public Opinion: Lessons from Five Countries*. Oxford: Oxford University Press.
- Roberts, J. V. and L. J. Stalans (2000). Crime, criminal justice, and public opinion. In M. Tonry (Ed.), *The Handbook of Crime & Punishment*, pp. 31–57. New York: Oxford University Press.
- Robinson, P. H. and J. M. Darley (2003). The utility of desert. *Northwestern University Law Review* 91(2), 453–499.
- Secret, P. E. and J. B. Johnson (1989). Racial differences in attitudes toward crime control. *Journal of Criminal Justice* 17(5), 361–375.
- Shotts, K. W. and A. E. Wiseman (2010). The politics of investigations and regulatory enforcement by independent agents and cabinet appointees. *The Journal of Politics* 72(1), 209–226.
- Simonson, J. (2011). Problems in measuring punitiveness – results from a german study. In H. Kury and E. Shea (Eds.), *Punitivity. International Developments. Vol. 1: Punitiveness – A Global Phenomenon?*, pp. 277–302. Bochum: Universitätsverlag Brockmeyer.
- Sklansky, D. A. (2025). *Criminal Justice in Divided America*. Cambridge, MA: Harvard University Press.
- Stagnaro, M. N., J. N. Druckman, A. J. Berinsky, A. A. Arechar, R. Willer, and D. G. Rand (2026). Representativeness and response validity across nine opt-in online samples. *Nature Human Behaviour*.
- Stimson, J. A. (1999). *Public Opinion in America: Moods, Cycles, and Swings*. Boulder, CO: Westview Press.
- Stimson, J. A. (2004). *Tides of Consent: How Public Opinion Shapes American Politics*. New York, NY: Cambridge University Press.
- Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review* 84(6), 1329–93.
- Unnever, J. D. (2013). Race, crime, and public opinion. In S. M. Bucerius and M. Tonry (Eds.), *The Oxford Handbook of Ethnicity, Crime, and Immigration*, pp. 70–106. New York, NY: Oxford University Press.

- Unnever, J. D. and F. T. Cullen (2007). Reassessing the racial divide in support for capital punishment: The continuing significance of race. *Journal of Research in Crime and Delinquency* 44(1), 124–158.
- Unnever, J. D. and F. T. Cullen (2010). The social sources of americans’ punitiveness: A test of three competing models. *Criminology* 48(1), 99–129.
- Vaughn, P. E., K. Peyton, and G. A. Huber (2022). Mass support for proposals to reshape policing depends on the implications for crime and safety. *Criminology & Public Policy* 21(1), 125–146.
- Volokh, A. (1997). n guilty men. *University of Pennsylvania Law Review* 146(1), 173–216.
- Weaver, V. M. (2007). Frontlash: Race and the development of punitive crime policy. *Studies in American Political Development* 21(2), 230–265.

Appendix A. Formal Results

To reduce notational clutter, we suppress the i , d and v arguments and express $b_i(v, d, s)$ in terms of the anticipated sentence and the individual finder of fact's ideal sentence given certain guilt: $b(s, \hat{s})$. In addition to the assumptions discussed in the text, we assume that $c(s)$ is increasing and convex in s ; and that $b(s, \hat{s})$ is concave in s and supermodular.²⁹

The Burden of Proof

We begin by establishing the indeterminacy of the relationship between the sentence s and the burden of proof in the general framework.

Proposition 1 *An individual's burden of proof may be (1) increasing or decreasing in the anticipated sentence s , and (2) increasing or decreasing in ideal sentence given certain guilt, $\hat{s}$.*

Proof. (1) From the text,

$$\hat{R}^g(s) = \frac{c(s)}{a(s = \hat{s}) + b(s, \hat{s})}. \quad (\text{A.1})$$

Differentiating with respect to s yields

$$\frac{\partial \hat{R}^g(s)}{\partial s} = \frac{(a(s = \hat{s}) + b(s, \hat{s})) \frac{\partial c(s)}{\partial s} - c(s) \frac{\partial b(s, \hat{s})}{\partial s}}{(a(s = \hat{s}) + b(s, \hat{s}))^2}.$$

This quantity is strictly positive if and only if

$$\frac{a(s = \hat{s}) + b(s, \hat{s})}{c(s)} > \frac{\partial b(s, \hat{s}) / \partial s}{\partial c(s) / \partial s}. \quad (\text{A.2})$$

To establish the indeterminacy of the sign of the derivative, it is sufficient to delineate cases in which the inequality holds and cases in which it doesn't. Cases of the former exist if $b(s, \hat{s}) + a(s = \hat{s}) \geq 0$ and $s > \hat{s}$: the left side of (A.2) will be weakly positive and the right side strictly negative, so the

²⁹Supermodularity has an intuitive interpretation: it states that given certain guilt, the marginal benefit of an increase in the anticipated sentence (toward an individual's own preferred sentence) is larger for individuals with higher ideal sentences.

inequality holds. Suppose instead that $b(s, \hat{s}) + a(s = \hat{s}) < 0$ and $s < \hat{s}$. Then the left side of (A.2) will be weakly negative and the right side strictly positive, thus reversing the inequality.

(2) Differentiating (A.1) with respect to $\hat{s}$ and rearranging, the sign of $\frac{\partial \hat{R}^g}{\partial \hat{s}}$ depends on that of

$$-c(s) \left(a(s = \hat{s}) + b(s, \hat{s}) \right)^{-2} \frac{\partial b(s, \hat{s})}{\partial \hat{s}},$$

which is negative if $s < \hat{s}$ and positive if $s > \hat{s}$. ■

Next, we establish sufficient conditions under which both s and $\hat{s}$ may be purged from our measure of the burden of proof. We do so by describing a specific functional form for each quantity in the generalized payoff matrix. Let $\kappa \in \mathbb{R}^{++}$ scale harm to the finder of fact from perceived overpunishment and $\beta \in \mathbb{R}^{++}$ harm from perceived underpunishment. We assume both kinds of harms are increasing in the quadratic distance between the anticipated sentence s and the conditional ideal sentence $\hat{s}$ if the defendant is truly guilty and 0 otherwise. We also assume that $b(s = \hat{s}, \hat{s}) = 0$; this imposes the constraint that convicting the guilty and acquitting the innocent yield the same payoff to the individual (zero) if and only if the guilty defendant receives that individual's ideal sentence. The expected utility of an acquittal, then, is

$$\begin{aligned} E[u(acquit|p, \hat{s})] &= -p\beta(\hat{s} - 0)^2 + (1 - p) \times 0 \\ &= -p\beta\hat{s}^2. \end{aligned} \tag{A.3}$$

The expected utility of conviction depends on whether the anticipated sentence falls below or above $\hat{s}$; in the former case a convicted, guilty defendant would be underpunished from the perspective of the individual, while in the latter he would be overpunished.

$$E[u(convict|p, \hat{s}, s)] = \begin{cases} -p\beta(\hat{s} - s)^2 - (1 - p)\kappa s^2 & \text{if } s < \hat{s} \\ -p\kappa(\hat{s} - s)^2 - (1 - p)\kappa s^2 & \text{otherwise.} \end{cases} \tag{A.4}$$

Comparing the expected utility of acquittal and conviction, the individual's burden of proof depends on what side of $\hat{s}$ the anticipated sentence s falls:

$$\hat{p}(s, \hat{s}, \kappa, \beta) \equiv \begin{cases} \frac{\kappa s}{\kappa s + \beta(2\hat{s} - s)} & \text{if } s < \hat{s} \\ \frac{\kappa s^2}{\beta \hat{s}^2 + \kappa \hat{s}(2s - \hat{s})} & \text{otherwise.} \end{cases} \quad (\text{A.5})$$

Proposition 2 *If $s = \hat{s}$, then $\hat{p}(s = \hat{s}, \hat{s}, \kappa, \beta) = \frac{\kappa}{\kappa + \beta}$ and $\hat{R}(s = \hat{s}) = \frac{\kappa}{\beta}$. For all other values of s , $\hat{p}(s, \hat{s}, \kappa, \beta)$ and $\hat{R}(s, \hat{s}, \kappa, \beta)$ depend on both s and $\hat{s}$.*

Proof. Follows immediately from inspection of (A.5). ■

Optimal Sentence Given Conviction

Proposition 3 *Suppose $p \in (0, 1)$ and $\hat{s} > 0$. Then the finder of fact's optimal sentence given conviction, s^* , is (1) unique, (2) strictly less than $\hat{s}$, and (3) increasing in both p and $\hat{s}$.*

Proof. The expected utility from conviction is $pb(s, \hat{s}) - (1 - p)c(s)$. s^* is defined implicitly by the first order condition

$$pb_s(s^*, \hat{s}) = (1 - p)c_s(s^*),$$

where $b_s(\cdot)$ and $c_s(\cdot)$ are first partial derivatives of $b(\cdot)$ and $c(\cdot)$ with respect to s . Suppose $s^* \geq \hat{s}$. Then $b_s(s^*) \leq 0$, and the left side of the first order condition is weakly negative while the right side is strictly positive, a contradiction. Therefore $s^* < \hat{s}$. Next, note that the expected utility is the sum of two concave functions and thus itself concave. Having established concavity and (given $s \geq 0$) $s^* \in [0, \hat{s}]$, s^* is a unique maximum.

Implicitly differentiating the first order condition with respect to p and rearranging yields

$$(pb_{ss}(s^*, \hat{s}) - (1 - p)c_{ss}(s^*)) \frac{\partial s^*}{\partial p} = -(c_s(s^*) + b_s(s^*, \hat{s})).$$

By concavity of $b(\cdot)$ and convexity of $c(\cdot)$, the term multiplying $\frac{\partial s^*}{\partial p}$ is strictly negative. Having established $s^* < \hat{s}$, the right side of the equation is also strictly negative. Thus $\frac{\partial s^*}{\partial p} > 0$.

Implicitly differentiating the first order condition with respect to $\hat{s}$ and rearranging yields

$$(pb_{ss}(s^*, \hat{s}) - (1 - p)c_{ss}(s^*)) \frac{\partial s^*}{\partial \hat{s}} = -pb_{s\hat{s}}(s^*, \hat{s}).$$

From the above logic, the term multiplying $\frac{\partial s^*}{\partial \hat{s}}$ is strictly negative. By supermodularity of $b(\cdot)$, the right side of the equation is also strictly negative. Hence $\frac{\partial s^*}{\partial \hat{s}} > 0$. ■

Appendix B. Implementation, Filtering, and Sample Demographics

Survey implementation

As described in the main text, all three surveys were implemented in the Qualtrics online survey platform using sample provided by Bovitz/Forthright Access. At the beginning of each survey, subjects provided written informed consent and subjects who did not consent did not provide data and were not compensated. There was no deception used in the surveys. Subjects were compensated by Bovitz at their standard rates. We did not learn the identification of individual respondents and all survey data were further anonymized prior to analysis by removing potentially identifying information like IP addresses.

Demographic covariates were gathered separately by Bovitz and provided to us so that we could merge them with completed survey responses. Details of those coding rules appear in the main text and sample demographics for each study appear below. Additional exact question wording is listed below.

For respondents who took the survey more than once, we used their first response.

Filtering

At the beginning of each survey, respondents were asked to read a short news article describing a bank robbery. The article was adapted from an actual news article. The article was introduced with this text: “First, we’d like to know how you feel about local news coverage. Please read this short article. You will only be able to advance after 10 seconds. On the following page, we will ask you a few questions about your reactions to this article.”

After reading the article, respondents proceeded to a new survey page and were asked four questions: “Do you think this article is typical of local news coverage?” (Yes, Maybe, No); “Do you think there is too much coverage of crime in local newspapers?” (Yes, Maybe, No); “How was

[NAME] identified by police for the crime he allegedly committed?” (A police officer recognized him, From video surveillance, Because he left his ID, He turned himself in, None of the above); and “How much money did [NAME] allegedly steal?” (About \$500, About \$1,500, About \$25,000, About \$1 million dollars, None of the above).

We used responses to the last two questions as screeners, and respondents who selected the correct answer to both items were coded as having passed the attention check. In Study 1, 82.3% of respondents passed both screeners. In Study 2, 60.0% of respondents passed both screeners. In Study 3, 80.6% of respondents passed both screeners. We do not know why passage rates were lower in Study 2 despite a common sample vendor and screener. We report demographics for the screened and unscreened samples below.

Missing Data

In the filtered samples, we report missing data rates for selected items below. As a reminder, in Studies 2 and 3 respondents could enter -99 to skip an item, whereas in Study 1 we requested responses but did not give explicit directions about how to indicate a desire not to answer an item.

Study 1:

Number Guilty Given Charged (out of 100): 0%

Number Innocent Given Charged (out of 100): 0%

Study 2:

Number Guilty Given Charged (out of 100): 5.5%

Number Guilty Given Found Guilty (out of 100): 5.6%

Study 3:

Number Guilty Given Charged Robbery (out of 100): 9.9%

Number Guilty Given Found Guilty Robbery (out of 100): 9.6%

Threshold to Convict Robbery (out of 100): 6.9%

Number Guilty Given Charged Burglary (out of 100): 9.7%

Number Guilty Given Found Guilty Burglary (out of 100): 9.6%

Threshold to Convict Burglary (out of 100): 6.9%

Demographics

Table B.1: Summary Statistics Across Studies

Variable	Study 1		Study 2		Study 3	
	All	Screened	All	Screened	All	Screened
Respondent is White (not others)	0.69 [0.46]	0.70 [0.46]	0.67 [0.47]	0.69 [0.46]	0.68 [0.47]	0.70 [0.46]
Respondent is Black	0.16 [0.36]	0.15 [0.36]	0.17 [0.38]	0.16 [0.36]	0.16 [0.37]	0.16 [0.36]
Respondent is Asian	0.07 [0.25]	0.07 [0.25]	0.06 [0.25]	0.06 [0.24]	0.06 [0.24]	0.06 [0.23]
Respondent is Other/Refused	0.11 [0.31]	0.11 [0.31]	0.12 [0.32]	0.11 [0.32]	0.11 [0.32]	0.11 [0.31]
Respondent is Hispanic (any race)	0.17 [0.38]	0.17 [0.37]	0.19 [0.39]	0.17 [0.37]	0.17 [0.38]	0.17 [0.38]
Age in years	47.32 [17.25]	47.99 [17.21]	46.94 [16.83]	47.83 [16.81]	47.26 [16.94]	48.41 [16.87]
Educ. = some college (baseline=HS or less)	0.39 [0.49]	0.38 [0.49]	0.44 [0.50]	0.43 [0.50]	0.39 [0.49]	0.41 [0.49]
Educ. = college	0.25 [0.43]	0.25 [0.43]	0.21 [0.41]	0.24 [0.43]	0.23 [0.42]	0.24 [0.42]
Educ. = post-BA	0.14 [0.34]	0.14 [0.35]	0.10 [0.31]	0.11 [0.31]	0.12 [0.32]	0.12 [0.32]
Gender = Female	0.48 [0.50]	0.48 [0.50]	0.50 [0.50]	0.51 [0.50]	0.50 [0.50]	0.50 [0.50]
Gender = non-binary	0.03 [0.16]	0.03 [0.17]	0.02 [0.14]	0.03 [0.16]	0.02 [0.14]	0.02 [0.14]
Income (1-11; 6=refused)	3.88 [2.34]	3.91 [2.30]	3.73 [2.29]	3.71 [2.23]	3.80 [2.27]	3.84 [2.25]
Income refused	0.03 [0.18]	0.03 [0.18]	0.03 [0.17]	0.02 [0.15]	0.03 [0.18]	0.03 [0.17]
Observations	854	703	1,734	1,041	2,602	2,096

Standard deviations in brackets.

Detailed Question Wording

Study 1

“Think about people who have been charged with murder in the United States in the last year. If found guilty, the sentence is life in prison.

Out of every 100 people who were charged with murder, how many do you think were actually guilty of murder?”

(Response entered as a number, 0-100)

“Think about people who have been charged with murder in the United States in the last year. If found guilty, the sentence is life in prison.

Out of every 100 people who were charged with murder, how many do you think were actually innocent?”

(Response entered as a number, 0-100)

“Think about 100 people on trial for murder, that is unlawfully killing a person. If found guilty, the sentence is life in prison.

How many of the 100 people would have to be *[guilty/innocent]* rather than *[innocent/guilty]* before you would *[convict/release]* all 100 people? (Response entered as a number, 0-100)

Study 2

“Think about people who have been charged with murder in the United States in the last year.

Out of every 100 people who were charged with murder, how many do you think were actually guilty of murder?

Enter your answer as a whole number from 0 to 100. Please enter -99 if you would like to skip this question.”

“Now think about people who have been charged with murder in the United States in the last

year AND FOUND GUILTY.

Out of every 100 people who were charged with murder and FOUND GUILTY, how many do you think were actually guilty of murder?

Enter your answer as a whole number from 0 to 100. Please enter -99 if you would like to skip this question.”

“Suppose that if a person is convicted of murder, they are sentenced to life in prison. If they are not convicted, they go free.

Now think about 100 people on trial for murder. Suppose some of these 100 people are guilty and the remainder are innocent. You are the juror in this case and are given a choice: either convict all 100 people or let them all go free.

What is the smallest number of people who would have to be GUILTY for you to convict all 100 people?

Enter your answer as a whole number from 0 to 100, where entering 60 means that if at least 60 of the people were guilty, you would convict all 100 of them. Please enter -99 if you would like to skip this question”

“Suppose that if a person is convicted of murder, they are sentenced to life in prison. If they are not convicted, they go free.

Now think about someone on trial for murder. You are the juror in this case and are given a choice: either convict or let them go free.

How certain would you have to be of their guilt to convict them?

Enter your answer as a whole number from 0 to 100, where 0 means you are completely certain they are innocent, 50 means you think they are equally likely to be innocent or guilty, and 100 means you are completely certain they are guilty. Please enter -99 if you would like to skip this question.”

Study 3

“Armed robbery is the taking, or attempted taking, of something of value from someone else by threat of force or violence while carrying a gun or other lethal weapon.

Out of every 100 {*Black/White/[BLANK]*} people charged with armed robbery in the last year, how many do you think are guilty?

(Enter your answer as a whole number from 0 to 100. If you’d like to skip this question, please enter -99.)”

Note: In Study 3, race was randomly assigned at equal rates to {*Black/White/[BLANK]*}. This randomization was held fixed for all the items a respondent evaluated.

“Out of every 100 {*Black/White/[BLANK]*} people convicted of armed robbery in the last year, how many do you think are guilty?

(Enter your answer as a whole number from 0 to 100. If you’d like to skip this question, please enter -99.)”

“Suppose a {*Black/White/[BLANK]*} person is definitely guilty of armed robbery. By law, the maximum sentence is 25 years in prison.

If they have no prior convictions, how many years in prison is the most appropriate punishment? Enter your answer as a whole number from 0 to 25. If you think probation (with no prison term) is the appropriate punishment, please enter 0.

(If you’d like to skip this question, please enter -99.)”

“[NAME], a 24-year old {*Black/White/[BLANK]*} man, is charged with armed robbery. He has no prior criminal convictions.

He is accused of confronting another man from his neighborhood on the street at 8pm on a Tuesday, pointing a gun at him, and stealing about \$2,500 in jewelry and cash. The victim was not injured.

If convicted, [NAME] will receive a sentence of {*Preferred Sentence/Above Preferred Sentence/Below*

Preferred Sentence }.

You are the juror in this case and are given a choice: either convict or let him go free. How certain would you have to be of his guilt to convict him?

Enter your answer as a whole number from 0 to 100, where 100 means you would only convict if you were certain he was guilty (100% chance of guilt), 50 means you would convict if he was equally likely to be guilty or innocent (50% chance of guilt), and 0 means that you would you would convict even if you were certain he was innocent (0% chance of guilt)

(If you'd like to skip this question, please enter -99.)”

Note: For respondents whose preferred sentence was at least 5 years from either end of the scale, Above Preferred Sentence was their preferred sentence plus 5 and Below Preferred Sentence was their preferred sentence minus 5. For respondents closer to either end of the scale, Above Preferred Sentence was their preferred sentence plus X and Below Preferred Sentence was their preferred sentence minus X, where X was the number of units they were from the end of the scale. Finally, for respondents at either end of the preferred sentence scale (0 or 25 years), they could be assigned to see only one direction of deviation. For those who preferred a sentence of 0, Above Preferred Sentence was 1. For those who preferred sentence was at the top of the scale, it was that minus 5 years.

These items were repeated for burglary, where burglary was defined as “Burglary is the act of illegally entering a building with the intention of committing a theft or any other felony.”

Preferred burglary sentences were assessed on a 0-20 scale.

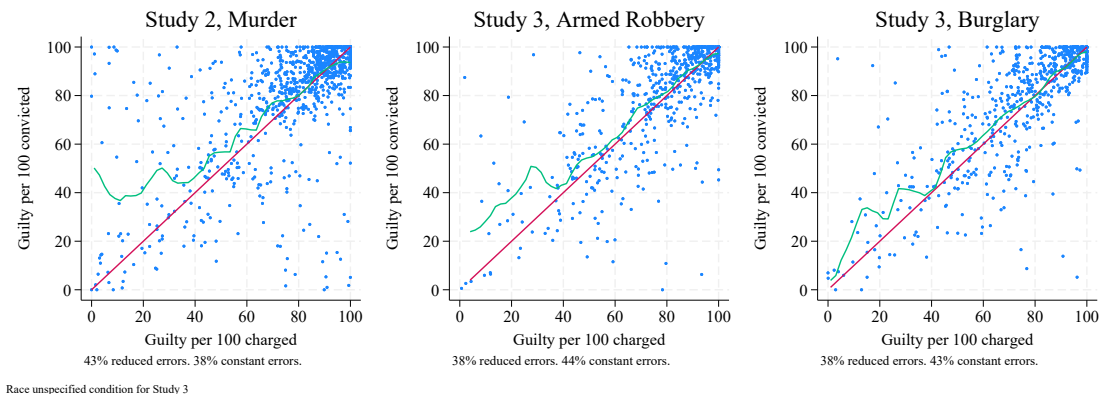
The burglary vignette read: “[NAME], a 22-year old {*Black/White/[BLANK]*} man, is charged with burglary. He has no prior criminal convictions.

He is accused of breaking into an apartment in his neighborhood at 2am while carrying a knife and stealing about \$2,000 in electronics and cash. The resident of the apartment, a 45-year old man, was asleep in the bedroom and did not awaken during the break in.

If convicted, [NAME] will receive a sentence of {*Preferred Sentence/Above Preferred Sentence/Below*

Preferred Sentence}....”

Figure C.1: Comparing Respondent Beliefs about Guilt Among the Charged and the Guilty: Studies 2 and 3



Appendix C. Supplemental Analysis

The way in which the criminal justice system reduces errors can also be seen by comparing estimates, within person, of the number of individuals who are guilty following charging and conviction. We plot these data in Figure C.1, where each point is a respondent's estimate for the proportion guilty following charging (the horizontal axis) and following conviction (the vertical axis). The 45 degree line indicates a trial that does not reduce errors, points above that line indicate a reduction in errors (fewer innocent people among those convicted), and points below the line indicate trials introduce more errors (more innocent people convicted than charged). The green line is a local polynomial smoother showing the relationship between the two estimates.

Figure C.1 demonstrates that most individuals view trials as either reducing or holding constant error rates (most points are at or above the red line, and the green line is generally above the red line for most of the ranges of the data). 43% of respondents believe trials reduce errors in cases of murder, and 38% believe they do so for Robbery and Burglary. About 80% of respondents, across crimes, provide estimates that imply trials do not make errors more common.

Table C.1: Determinants of Preferred Punishments for Murder Given Certain Guilt Using Alternative Outcome Variable Coding

	(1) Death Penalty (1=Yes)	(2) Collapse Life/Death (0-4)
Respondent Black	-0.140 [0.040]***	-0.049 [0.059]
Respondent Hispanic (any race)	-0.051 [0.050]	-0.104 [0.082]
Age in years	0.004 [0.001]***	0.007 [0.001]***
Educ. = some college	0.000 [0.042]	-0.040 [0.048]
Educ. = college	-0.031 [0.050]	-0.125 [0.063]**
Educ. = post-BA	-0.066 [0.065]	-0.110 [0.070]
Gender=Female	-0.084 [0.033]**	0.092 [0.041]**
Gender=non-binary	-0.184 [0.086]**	-0.238 [0.183]
Income (1-11; 6=refused)	0.002 [0.008]	0.008 [0.010]
Income refused	0.032 [0.110]	0.023 [0.128]
Constant	0.208 [0.069]***	3.462 [0.082]***
Observations	859	859
R-squared	0.055	0.064
Race Unspecified Only	Yes	Yes
Sample mean	0.330	3.780
Robust standard errors in brackets		
* significant at 10%; ** significant at 5%; *** significant at 1%		

Table C.2: Regression Results of Our Items vs. Traditional Items: Standardized Means In Lieu of Factor Scores

	(1)	(2)	(3)	(4)	(5)	(6)
	Confidence courts/legal system (0-3)	Confidence police/law enforce. (0-3)	System tough (2=Not tough, 0=too tough)	Prison selling marij. (2=yes, 1=dep., 0=no)	Death Penalty Murder (2=yes, 1=dep., 0=no)	Worse to Convict Innocent (2=yes, 1=not sure, 0=no)
Factor 1: Beliefs about Prob. Guilty	0.069 [0.022]***	0.132 [0.024]***	0.083 [0.020]***	0.047 [0.018]**	0.117 [0.018]***	0.005 [0.021]
Factor 2: Preferred Sentence if Guilty	0.008 [0.020]	0.106 [0.022]***	0.204 [0.017]***	0.114 [0.017]***	0.188 [0.016]***	-0.122 [0.020]***
Factor 3: Conviction Threshold	0.027 [0.023]	0.033 [0.023]	-0.003 [0.020]	-0.026 [0.018]	-0.051 [0.018]***	0.110 [0.021]***
Constant	1.489 [0.019]***	1.674 [0.021]***	1.438 [0.017]***	0.533 [0.017]***	1.175 [0.017]***	1.542 [0.018]***
Observations	1733	1732	1732	1732	1732	1733
R-squared	0.011	0.046	0.094	0.034	0.102	0.046
Sample mean	1.489	1.674	1.439	0.535	1.177	1.540
Robust standard errors in brackets						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Figure C.2: Scree plot following factor analysis of 8 novel survey items, Study 3 data

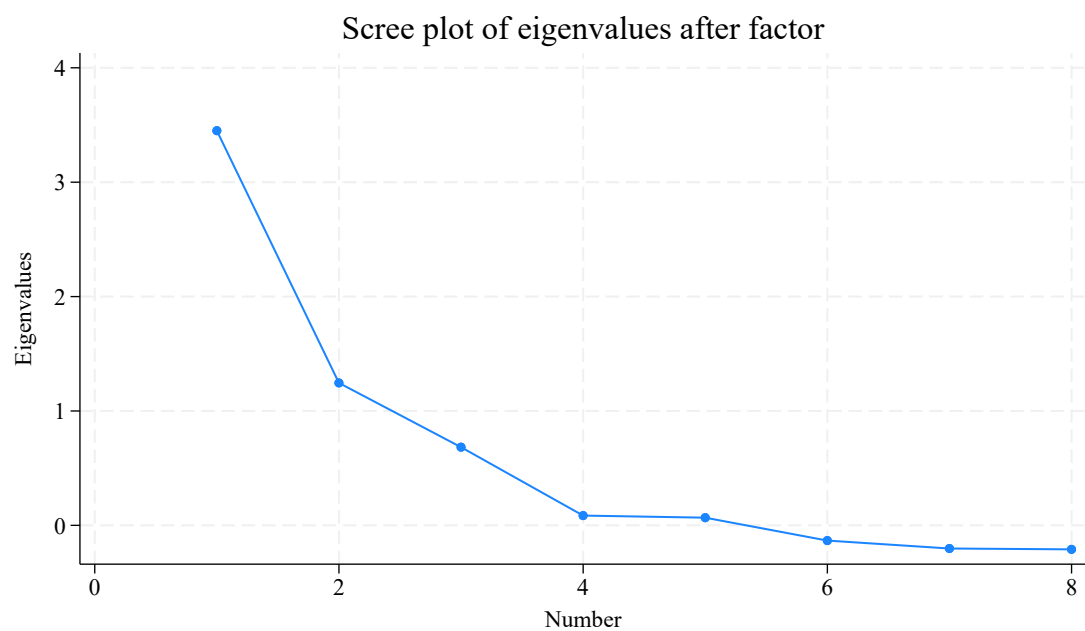


Table C.3: Regression Results of Our Items vs. Traditional Items: Non-Black Respondents Only

	(1) Confidence courts/legal system (0-3)	(2) Confidence police/law enforce. (0-3)	(3) System tough (2=Not tough, 0=too tough)	(4) Prison selling marij. (2=yes, 1=dep., 0=no)	(5) Death Penalty Murder (2=yes, 1=dep., 0=no)	(6) Worse to Convict Innocent (2=yes, 1=not sure, 0=no)
Factor 1: Beliefs about Prob. Guilty	0.069 [0.025]***	0.136 [0.027]***	0.089 [0.022]***	0.039 [0.022]*	0.087 [0.021]***	0.018 [0.024]
Factor 2: Preferred Sentence if Guilty	0.000 [0.027]	0.157 [0.030]***	0.257 [0.023]***	0.142 [0.024]***	0.245 [0.022]***	-0.141 [0.027]***
Factor 3: Conviction Threshold	0.036 [0.032]	0.057 [0.032]*	-0.022 [0.028]	-0.029 [0.027]	-0.050 [0.025]**	0.151 [0.031]***
Constant	1.546 [0.023]***	1.779 [0.025]***	1.498 [0.020]***	0.549 [0.020]***	1.238 [0.020]***	1.543 [0.022]***
Observations	1216	1215	1216	1216	1216	1216
R-squared	0.009	0.049	0.107	0.033	0.098	0.049
Sample mean	1.546	1.779	1.498	0.549	1.238	1.543
Robust standard errors in brackets						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Table C.4: Regression Results of Our Items vs. Traditional Items: No More than One Response of “50”

	(1)	(2)	(3)	(4)	(5)	(6)
	Confidence courts/legal system (0-3)	Confidence police/law enforce. (0-3)	System tough (2=Not tough, 0=too tough)	Prison selling marj. (2=yes, 1=dep., 0=no)	Death Penalty Murder (2=yes, 1=dep., 0=no)	Worse to Convict Innocent (2=yes, 1=not sure, 0=no)
Factor 1: Beliefs about Prob. Guilty	0.078 [0.023]***	0.142 [0.025]***	0.113 [0.022]***	0.048 [0.020]**	0.131 [0.020]***	0.006 [0.022]
Factor 2: Preferred Sentence if Guilty Certain	0.027 [0.026]	0.153 [0.028]***	0.235 [0.021]***	0.140 [0.023]***	0.237 [0.021]***	-0.156 [0.025]***
Factor 3: Conviction Threshold	0.029 [0.031]	0.063 [0.031]**	0.054 [0.028]*	-0.035 [0.025]	-0.046 [0.023]**	0.136 [0.029]***
Constant	1.507 [0.021]***	1.705 [0.023]***	1.429 [0.019]***	0.546 [0.018]***	1.179 [0.018]***	1.553 [0.020]***
Observations	1443	1442	1443	1442	1442	1443
R-squared	0.011	0.049	0.097	0.034	0.109	0.049
Sample mean	1.507	1.705	1.429	0.546	1.179	1.553
Robust standard errors in brackets						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Appendix D. Ethics Compliance

The research contained herein complies fully with APSA's *Principles of Ethics for Human Subjects Research and Guidance*.

Consent: At the beginning of each survey, subjects provided written informed consent and subjects who did not consent did not provide data and were not compensated. There was no deception used in the surveys. Subjects were compensated by Bovitz at their standard rates. We did not learn the identification of individual respondents and all survey data were further anonymized prior to analysis by removing potentially identifying information like IP addresses.