American Economic Review 2023, 113(5): 1255–1293
https://doi.org/10.1257/aer.20210970

Law and Norms: Empirical Evidence[†]

By Tom Lane, Daniele Nosenzo, and Silvia Sonderegger*

A large theoretical literature argues laws exert a causal effect on norms, but empirical evidence remains scant. Using a novel identification strategy, we provide a compelling empirical test of this proposition. We use incentivized vignette experiments to directly measure social norms relating to actions subject to legal thresholds. Our large-scale experiments ($n = 7{,}000$) *run in the United Kingdom, United States, and China show that laws can causally influence social norms. Results are robust across different samples and methods of measuring norms, and are consistent with a model of social image concerns where individuals care about the inferences others make about their underlying prosociality.* (*JEL* C91, C92, D91, K00, K42, P37)

Legal rules play a vital role in the functioning of societies. Across all walks of life, laws regulate and constrain social behaviors, from the taxes individuals pay to governments, to the way they treat employees at work, or to the public health behaviors they are required to take during a pandemic. However, an emerging literature in behavioral economics shows that many behaviors are also influenced by informal rules of conduct that define what society perceives as socially appropriate or inappropriate. Unlike laws, these *social norms* are not formally codified or sustained by extrinsic reinforcements such as material penalties or fines, yet they are commonly recognized within a given society and informally enforced by social sanctions and rewards. Recent research has suggested that norms are an essential determinant of many of the social behaviors that are also regulated by law, such as the untruthful reporting of private information (e.g., Abeler, Nosenzo, and Raymond 2019), tax evasion (e.g., Dwenger et al. 2016), bribery and corruption (e.g., Fisman and Miguel 2007), or the expression of discriminatory behaviors or opinions (e.g., Barr, Lane, and Nosenzo 2018; Bursztyn, Egorov, and Fiorin 2020).

*Lane: School of Economics, University of Nottingham Ningbo (email: Tom.Lane@nottingham.edu.cn); Nosenzo: Department of Economics and Business Economics, Aarhus University (email: daniele.nosenzo@econ.au.dk); Sonderegger: School of Economics and Centre for Decision Research and Experimental Economics (CeDEx), University of Nottingham (email: silvia.sonderegger@nottingham.ac.uk). Stefano DellaVigna was the coeditor for this article. We thank first-round coeditor Roland Bénabou and three anonymous referees for their constructive comments. We also received helpful comments from Andrea Albanese, Antonio Alonso Arechar, Kai Barron, Roberto Galbiati, Klarita Gërxhani, Luise Görges, Benjamin Häusinger, Andrea Mercatanti, Daniel Seidmann, Adriaan Soetevent, Sofie Waltl, Roberto Weber, and participants at several seminars and workshops. This work was supported by the University of Nottingham, the Economic and Social Research Council (grant ES/K002201/1), the Luxembourg Institute of Socio-Economic Research (LISER), and the Aarhus University Research Foundation (AUFF Starting Grant 36835). This study received IRB approval from the University of Nottingham. Preregistrations for the study are available at Lane, Nosenzo, and Sonderegger (2021, 2022).
[†]Go to https://doi.org/10.1257/aer.20210970 to visit the article page for additional materials and author disclosure statements.

What is the relationship between these two institutions—law and norms—that frequently regulate very similar types of behavior? Do they have independent influence on behavior, one through the deterrent power of material incentives, the other by the power of social incentives? Or are there interdependencies in their influence on social behavior? And, more specifically, can lawmakers use the law to affect behavior, not just through the deterring power of incentives, but also through what has been labeled *the expressive function of law* (Sunstein 1996), i.e., by shaping the underlying social norms of a society?

This paper presents compelling empirical evidence on the causal influence of law on social norms. While this question has attracted the interest of many researchers from multiple disciplines in the last two decades, and a plethora of theoretical mechanisms have been proposed to explain how law may shape norms, the empirical evidence remains scant. This is mainly because the identification of the causal effects of law on norms presents substantive challenges to empirical research. First, for many years social scientists have struggled to translate the concept of social norm into a measurable construct that can be used in empirical analysis. Therefore, previous empirical research has mostly been limited to studying the influence of legal rules either on *behavior*—arguing that the observed effects cannot be merely explained by deterrence and thus providing indirect evidence for the influence of law on norms (e.g., Funk 2007)—or on *personal opinions* (e.g., Chen and Yeh 2014; Aksoy et al. 2020)—a construct that is related to but quite distinct from social norms. This paper exploits recent advancements in empirical research on social norms (e.g., Bicchieri and Xiao 2009; Krupka and Weber 2013; Bursztyn, González, and Yanagizawa-Drott 2020) to design a series of vignette experiments that allow us to measure, directly and with incentive compatibility, the social norms pertaining to various social behaviors. Through these measurements, we can directly observe the influence that law exerts on norms.

A second, pervasive challenge faced by empirical research in this area concerns the difficulty in establishing a clear direction of causality between law and social norms. This is because law and norms coevolve: they might influence one another and are often simultaneously codetermined by external factors, such as the availability of information about the harms of certain behaviors.[1] Our study overcomes this identification problem by exploiting a special subclass of laws that regulate behavior by *legal thresholds* defining the cutoff point above (or below) which a certain behavior becomes illegal (e.g., speed limit laws; drink-drive laws; age of consent laws; etc.). If a social norm exists that governs the same behavior also regulated by a legal threshold, it is reasonable to assume this norm, absent the law, would not make sharp distinctions between behaviors arbitrarily near the threshold (e.g., driving with blood alcohol content [BAC] of 0.079 or 0.081 percent, if the limit is 0.08 percent), since these behaviors are virtually identical in all respects except their legality. Thus, if we observe a discrete change in the perceived social appropriateness of behaviors

[1] Several scholars argue that norms often precede the law and lead to its creation (e.g., Posner 1997; Chen and Yeh 2013). Indeed, some authors contend the law's effectiveness in regulating behavior crucially depends on whether it reflects the normative intuition of the governed community (e.g., Acemoglu and Jackson 2017).

just on either side of a legal threshold, we can causally attribute this difference to the influence of law.[2]

We formalize these ideas in Section I, which adapts the theoretical framework proposed by Bénabou and Tirole (2006, 2011) to model a simple mechanism by which the law can induce sharp discontinuities in the perceived social appropriateness of legal and illegal behavior. In our model, social norms are functions that describe the social sanctions ("stigma") and rewards ("esteem") accruing to an individual for engaging in behavior that is observable by others. Individuals care about social norms as they gain utility from esteem and lose utility from stigma. Individuals also care about the negative externalities they impose on others, i.e., they are "prosocial." We show that, under standard assumptions and in absence of a law discriminating between legal and illegal behaviors, actions producing very similar negative externalities (like driving with a BAC of 0.079 or 0.081 percent) attract very similar levels of social esteem.

We then argue laws introduce sharp payoff discontinuities between legal and illegal behaviors, even when these behaviors are virtually identical. These payoff discontinuities occur both because laws assign material penalties to lawbreakers and because criminal offenses are registered in criminal records, which makes illegal behavior visible to a wider audience (e.g., future employers who did not directly observe the individual's actions when he/she committed the crime).[3] As a consequence, the esteem of illegal behavior is reduced as audiences take into account that someone engaging in criminal behavior is willing to do so despite the payoff discontinuities existing between legal and illegal actions. Thus, an individual who "marginally" breaks the law (e.g., by driving with a BAC of 0.081 percent) does not just receive marginally lower esteem than someone acting within the law (e.g., driving with a BAC of 0.079 percent). We interpret this ability of the law to introduce discontinuities in social esteem between legal and illegal behaviors as a manifestation of its "expressive power," which we refer to as the ability of law to shape the social norms prevailing in a society.

Section II reports an experiment designed to provide direct empirical evidence of the existence of discontinuities in our measurements of social norm functions in the presence of legal thresholds. Our vignette experiments asked subjects to evaluate the social appropriateness of various behaviors regulated by legal thresholds. We consider five types of legal thresholds, pertaining to sexual activity with minors, sale of alcohol to minors, undeclared cash imports into a country, drink-driving, and speeding. Across several treatments, we presented subjects with slightly modified versions of the vignettes which described behavior that is either legal or illegal, and

[2] Direction of causality can also be readily established in laboratory experiments, where the researcher tightly controls the decision environment and can introduce exogenous changes in the "rules" governing lab behavior. A number of papers have studied effects of such "lab laws" using experimental games. They show requirements about specific actions mandated by the experimenter (e.g., a minimum contribution level in a public goods game, a minimum admissible wage in a gift-exchange game) can affect behavior even if supported by weak, nondeterrent material sanctions, and that the effect can last even after the requirement has been lifted (e.g., Falk, Fehr, and Zehnder 2006; Galbiati and Vertova 2008, 2014; Barron and Nurminen 2020; Engl, Riedl, and Weber 2021; Govindan 2021). Unlike these studies, our paper focuses on the effects of laws that regulate behavior outside the lab, circumventing the issue of external validity sometimes raised for experiments focusing on how individuals respond to the legal environment.

[3] This second mechanism echoes an argument proposed informally by Posner (1998, 2000, 2002). Recently, Tirole (2021) built on a similar intuition to analyze effects of technologies that allow collection of personal data and can thus widen the number of people able to observe an individual's actions.

either closer or further away from the threshold (for example, driving with a BAC of 0.001, 0.002, 0.003 or 0.004 percentage units above or below the legal limit). In each case, we used incentivized experimental techniques to measure the social norm pertaining to the behavior described in the vignette, and thus elicit a "normative function" expressing the social appropriateness of behavior as a function of age, cash amount imported, BAC, or speed, depending on the type of vignette. We measure the expressive effect of law on each norm by testing for a discontinuity in the corresponding normative function at the legal threshold.

The experiments featured 1,248 UK participants from across three samples: one student sample and two samples representative of the general population. We used two different methods to elicit social norms, using coordination games (Krupka and Weber 2013) and a sequential opinion matching method (e.g., Bicchieri and Xiao 2009; Bursztyn, González, and Yanagizawa-Drott 2020). We report results in Section III. In all samples and across both methods, we find clear evidence of marked discontinuities in the normative functions at the legal thresholds. However, we also observe differences in the expressive power of law across the five types of behavior considered. In particular, we find strong effects of law on norms associated with sexual relations with minors, selling alcohol to minors, and importing undeclared cash amounts. We find instead weaker or no effects for laws regulating drink-driving and speeding behavior. We provide suggestive evidence that these heterogeneous effects are related to differences across the five domains of law in perceptions of the intentionality of illegal behavior and ability of law enforcement to detect it, which is consistent with our model.

To probe these results' robustness, Section IV reports four additional experiments run with another 5,771 subjects, which (i) conducted placebo tests introducing arbitrary thresholds unrelated to the law to measure whether they create similar discontinuities to the legal thresholds (they do not); (ii) collected more evidence on the proposed mechanism underlying the norm discontinuities by testing whether we also observe discontinuities directly in subjects' perceptions of the trustworthiness, honesty, and altruism of someone who engages in legal or illegal behavior (we do); (iii) tested whether the direction of the discontinuities in trustworthiness, honesty, and altruism is reversed when, by violating a law, an individual engages in prosocial behavior (it is); (iv) tested whether our findings generalize to a society (China) with relatively weak rule of law (they do). Finally, we discuss a possible alternative interpretation of our results based on a mechanism where the social sanctions and rewards accruing to an individual depend on how "common" their behavior is (i.e., on "descriptive" norms). We show this alternative mechanism can also explain our data, but with additional assumptions not always supported by existing empirical evidence.

Our paper contributes to an interdisciplinary literature, both theoretical and empirical, on the expressive function of law. In theoretical work, scholars have discussed a number of mechanisms to explain the source of the expressive power of law. A prominent approach proposes the law can act as a public signal containing information citizens use to update beliefs about relevant features of the decision environment, such as the prevailing standards of behavior or the distribution of agents' preferences (e.g., McAdams 2000, 2015; Bénabou and Tirole 2011; van der Weele 2012). Another approach suggests individuals may comply with a norm of legal obedience whereby they feel obliged to follow the law, and therefore automatically consider as appropriate the behaviors that are legal and as inappropriate those that are not (e.g., McAdams and Rasmusen 2007). Our paper builds on the framework introduced by Bénabou and Tirole (2011) to formalize an alternative mechanism through which law can exert expressive power. In our model the law conveys no information about the decision environment.[4] Instead, by drawing a "line in the sand" between legal and illegal behaviors, the law allows individuals to reveal private information *about themselves*, based on how they behave in relation to it. This confers discontinuously different "social meaning" to behaviors that fall within or outside the confines of the law—which is the key intuition driving our empirical strategy.

It is notoriously challenging to design empirical research that establishes a clear direction of causality from laws to norms. This explains why only a few empirical studies exist in this literature. Funk (2007); Wittlin (2011); and Rees-Jones and Rozema (2019) show the law affects behavior beyond what one would expect based on the mere deterrent power of incentives, but cannot establish that these effects are actually mediated by shifts in the underlying social norms since they do not measure them.[5] Three recent papers (Tankard and Paluck 2017; Casoria, Galeotti, and Villeval 2020; Galbiati et al. 2020) measure directly the impact of law on social norms, exploiting changes in existing laws and, in Casoria, Galeotti, and Villeval (2020), using incentivized norm-elicitation techniques similar to those we use (Tankard and Paluck [2017] study the effects of a US Supreme Court ruling in favor of same-sex marriage; Casoria, Galeotti, and Villeval [2020] and Galbiati et al. [2020] exploit changes in COVID-19 public health regulations). As with all designs relying on changes in existing laws for identification, the concern is that such legislative changes may take place simultaneously to other events with the potential to shape norms (e.g., enhanced media debate of the relevant issues), hence casting doubt on whether changes in norms are necessarily caused by the changes in laws that precede them. The novel identification strategy we use in this paper rests on milder assumptions about causality and complements this existing work by providing first-hand empirical evidence that legal rules have causal power to shape normative intuitions about the behaviors that they regulate.

As we discuss in Section V, our paper has broader implications for the body of literature investigating the interactions between formal and informal incentives (e.g., Gneezy and Rustichini 2000; Bénabou and Tirole 2003, 2006, 2011; Fehr and Rockenbach 2003; Birke 2020; for recent reviews see Bowles and Polanía-Reyes 2012; Charness, Cooper, and Reddinger 2020). Like our paper, this literature argues formal incentives and institutions can affect behavior not merely by altering the material payoffs associated with actions, but also by influencing their social meaning. We provide compelling and direct evidence that laws can strongly influence the social meaning of human activities by shaping the perception of what is right and wrong in a society.

[4] We discuss this assumption in Section II and Section IVA. We argue it is appropriate for the types of laws we empirically explore. We also use placebo experiments in Section IVA that suggest the expressive power of law we empirically observe is unlikely to derive from transmitting information about the environment.

[5] Funk (2007) shows abolishing the legal duty to vote in four Swiss cantons had a detrimental effect on voter turnouts, unlikely to be due to (lack of) deterrence as fines for not voting were very low (less than $1 in most cases). Wittlin (2011) shows differences in seat belt use across US states cannot be solely explained by state-level variations in penalties for not wearing one, and that enacting a seat belt law in one state has spillover effects on its neighbors. Rees-Jones and Rozema (2019) show the effects of changes in US cigarette tax law are mediated by the intensity of media coverage, lobbying efforts, and other activities related to the lawmaking process.

I. Theoretical Framework

A. *Model*

To fix ideas and substantiate our empirical analysis, we first sketch a simple model of the expressive power of law. This formal model was introduced after we had collected part of this paper's data.[6] However, the design of our experiments was informed by the ideas formalized in this section, which were discussed informally already in the paper's first version (Lane and Nosenzo 2019). The model borrows from the literature on social image concerns (e.g., Bernheim 1994; Ellingsen and Johannesson 2008; Andreoni and Bernheim 2009) and in particular the "social esteem" framework proposed by Bénabou and Tirole (2006, 2011).[7]

We consider an individual presented with a randomly drawn opportunity for material gain that may however impose a negative externality on others. Opportunities differ in the severity of the negative externality they generate upon being taken: opportunity o creates a negative externality of size o drawn from a distribution with continuous differentiable density $g(\cdot)$ and full support $[o_{min}, o_{max}]$, where $o_{min} > 0$.

In line with our empirical strategy, we focus on laws regulating behavior by *legal thresholds* that define a clear cutoff point between legal and illegal behavior, such as speed laws, age of consent laws, laws against the sale of alcohol to minors, and so on.

For instance, a shopkeeper may face the opportunity to materially profit by selling alcohol to a young customer who, depending on the draw of o, may be above or below the legal drinking age. The shopkeeper has to decide whether to take this opportunity. For concreteness, in the following we model legal thresholds by considering a law that introduces a threshold $\bar{o}$ above which seizing an opportunity is illegal (the case where the law sets a threshold $\underline{o}$ can be accommodated through a simple relabeling exercise).

The individual derives utility from material payoff, but also experiences a psychological cost from imposing negative externalities. The individual also cares about the social esteem he/she obtains from seizing or leaving the opportunity. Utility is given by

$$u_a(o;\theta) = (t - \theta o - pKI_{o>\bar{o}})a + S(o,a). \tag{1}$$

When the individual takes an opportunity o he/she receives a material gain t and, if the opportunity exceeds the legal threshold (i.e., $o > \bar{o}$), also faces a material penalty K with probability $p \in (0,1]$, the probability of being caught. The individual suffers a psychological cost of size θo for imposing externality o on others, where θ is his/her (privately known) type, measuring how much he/she cares about causing negative externalities—this subsumes a host of possible prosocial characteristics, from trustworthiness, to altruism, to honesty. Types are drawn from a distribution with continuous differentiable density $f(\cdot)$ with mean μ_θ and full support $[\theta_{min}, \theta_{max}]$, where $\theta_{min} \geq 0$. A higher type suffers a higher psychological cost for imposing a negative externality.

[6] Specifically, we had data from samples 1 and 2 of the experiments in Section II and experiment 4 of Section IV.

[7] Adriani and Sonderegger (2019); Ali and Bénabou (2020); and Tirole (2021) also employ a similar framework.

The individual also cares about *social norms*: utility depends on the social rewards ("esteem") and sanctions ("stigma") associated with seizing or leaving an opportunity. This is represented by the last term in (1), measuring the inferences other people ("observers") make about the individual's type θ upon observing his/her choice a and the opportunity o the individual is presented with. Formally, the esteem conferred to an individual who seizes opportunity o is

$$S(o,1) \equiv E(\theta \mid o, a = 1),$$

while the esteem conferred to an individual who leaves opportunity o is

$$S(o,0) \equiv E(\theta \mid o, a = 0).$$

To sum up, when an individual of type θ faced with opportunity o decides *not* to take the opportunity, his/her utility is

$$u_{a=0}(o;\theta) = S(o,0),$$

independent of θ. If the individual chooses to take the opportunity, expected utility is

$$u_{a=1}(o;\theta) = \begin{cases} t - \theta o + S(o,1), & \text{if } o \leq \bar{o} \text{ (legal);} \\ t - \theta o - pK + S(o,1), & \text{if } o > \bar{o} \text{ (illegal).} \end{cases}$$

B. *Analysis*

We focus on equilibria where opportunities generating stronger negative externalities are seized by less prosocial types (monotonicity). We also restrict attention to interior solutions. Our equilibrium concept is perfect Bayesian equilibrium.

We consider two cases. First, we analyze the esteem accruing in equilibrium to an individual who takes opportunity o in the benchmark case where no law sets a cutoff point beyond which opportunities are illegal. This analysis characterizes the social norm function governing behavior in the absence of law. Then, we analyze how introducing a law prohibiting opportunities $o > \bar{o}$ affects the shape of the norm function. The proofs of the results presented in this section are in online Appendix A.

1. Esteem in the Absence of Law.—If no law prohibits opportunities $o > \bar{o}$, this is formally equivalent to setting $p = 0$. Consider an individual of type θ. Taking social esteem as given, the *net* utility from seizing the opportunity is given by $t - \theta o + S(o,1) - S(o,0)$, decreasing in θ. For each opportunity o, we can therefore identify the highest θ who takes o. Denote this as $\hat{\theta}_o$. In equilibrium, esteem is

$$S(o,1) = \mathcal{M}^{-}(\hat{\theta}_o) \quad \text{and} \quad S(o,0) = \mathcal{M}^{+}(\hat{\theta}_o), \tag{2}$$

where $\mathcal{M}^-(\theta_o) \equiv E(\theta|\theta < \theta_o), \mathcal{M}^+(\theta_o) \equiv E(\theta|\theta > \theta_o)$. For any o, the highest type who seizes o satisfies the indifference condition:

$$t - \hat{\theta}_o o - \Delta(\hat{\theta}_o) = 0, \tag{3}$$

where $\Delta(\theta_o) \equiv \mathcal{M}^+(\theta_o) - \mathcal{M}^-(\theta_o)$. Under our assumptions the solution of (3) is interior. Furthermore, $\hat{\theta}_o$ is continuously decreasing in o (monotonicity).[8]

PROPOSITION 1: *In the absence of a law prohibiting $o > \bar{o}$, the esteem $S(o,1)$ conferred to an individual who takes opportunity o is continuously decreasing in o.*

The result follows straightforwardly from the fact that $\hat{\theta}_o$ is continuously decreasing in o.[9] The proposition shows that, when no laws set a limit on allowed opportunities, the esteem for seizing an opportunity varies *continuously* with the magnitude of the negative externality imposed. This formalizes the key identifying assumption we will make in the empirical part of the paper, where we argue that, in the contexts we consider, absent the law, norms do not make sharp distinctions between arbitrarily close behaviors.

Formally, the continuity result relies on two features. First, we focus on environments with no discontinuities in externality generation. This means, when $\epsilon \to 0$, the externalities generated by seizing two opportunities $o - \epsilon$ and $o + \epsilon$ have the same magnitude. We think this assumption applies naturally to the decision situations we will consider empirically (e.g., selling alcohol to a customer who is $o - \epsilon$ years old or $o + \epsilon$ years old, for ϵ small). It may not naturally extend, however, to other types of situations where there may exist discontinuous externalities even between marginally similar actions (e.g., causing no harm or some harm to an innocent bystander). Second, the distribution of individual types is continuous with full support. This is a standard assumption ruling out, for instance, any distributions of types containing "holes."

2. *Esteem in the Presence of Law.*—Suppose now that, for some $\bar{o} \in (o_{min}, o_{max})$, an individual taking opportunities $o > \bar{o}$ breaks the law. There are two complementary effects. First, with probability $p \in (0,1]$ the individual is caught and incurs a material penalty $K > 0$. Second, as discussed in the legal literature (e.g., Posner 1998, 2000, 2002), when an individual's criminal offense is registered in his/her criminal record, the individual's behavior becomes known to a *larger* set of observers since even people who cannot directly observe an individual's actions or the opportunities he/she has faced can learn his/her behavior via the criminal record.[10] We consider each effect in turn.

[8] A sufficient condition for monotonicity is that, for all θ_o, $o_{min} \geq -\Delta'(\theta_o)$. To rule out corner solutions, we assume $t - o_{max}\theta_{min} - \mu_\theta + \theta_{min} > 0 > t - o_{min}\theta_{max} + \mu_\theta - \theta_{max}$.

[9] We present our results in terms of $S(o,1)$, omitting $S(o,0)$, because in the empirical analysis we measure esteem associated with *taking* (rather than *leaving*) an opportunity. In online Appendix A we comment on results pertaining to $S(o,0)$.

[10] See also Tirole (2021) for a related approach where an individual's behavior can become known to an audience either via direct observation or via publicly disclosed "social scores" containing information on the individual's behavior in multiple contexts not directly observed by the audience.

Legal Sanctions: The mere presence of a legal sanction $K > 0$ is sufficient to give the law expressive power, because the expected cost associated with seizing $o > \bar{o}$ has a spillover effect on the esteem the individual obtains. To see why, consider two opportunities, $\bar{o} - \epsilon$ and $\bar{o} + \epsilon$, where ϵ is vanishingly small. The expected return from seizing $\bar{o} + \epsilon$ is substantially smaller, due to the additional expected cost pK incurred if the individual is caught. An individual willing to take a marginally illegal opportunity $\bar{o} + \epsilon$ is thus not simply a marginally "worse" type, on average, than an individual who seizes $\bar{o} - \epsilon$. Since the individual is willing to seize an opportunity associated with a substantially smaller return, his/her type must be, on average, substantially lower. Observers recognize and take this into account when forming beliefs about the individual's type. Therefore, seizing $\bar{o} + \epsilon$ carries significantly lower esteem than seizing $\bar{o} - \epsilon$, despite these two opportunities generating very similar externalities. Formally, the equilibrium esteem function is given by

$$S(o,1) = \begin{cases} \mathcal{M}^-(\hat{\theta}_o), & \text{if } o \le \bar{o}; \\ \mathcal{M}^-(\tilde{\theta}_o), & \text{if } o > \bar{o}; \end{cases} \quad \text{and} \tag{4}$$

$$S(o,0) = \begin{cases} \mathcal{M}^+(\hat{\theta}_o), & \text{if } o \le \bar{o}; \\ \mathcal{M}^+(\tilde{\theta}_o), & \text{if } o > \bar{o}; \end{cases}$$

where $\hat{\theta}_o$, the highest type seizing $o \le \bar{o}$, is defined by (3): $t - \hat{\theta}_o o - \Delta(\hat{\theta}_o) = 0$, while $\tilde{\theta}_o$, the highest type seizing $o > \bar{o}$, is defined (when interior) by

$$t - pK - \tilde{\theta}_o o - \Delta(\tilde{\theta}_o) = 0. \tag{5}$$

PROPOSITION 2: *In the presence of a law prohibiting $o > \bar{o}$, the function $S(o,1)$ exhibits a downward discontinuity at $\bar{o}$:*

$$\lim_{\epsilon \to 0}[S(\bar{o} - \epsilon, 1) - S(\bar{o} + \epsilon, 1)] = \mathcal{M}^-(\hat{\theta}_{\bar{o}}) - \mathcal{M}^-(\tilde{\theta}_{\bar{o}}) > 0.$$

The crucial observation here is that $\hat{\theta}_o$ always lies strictly above $\tilde{\theta}_o$. Since $\mathcal{M}^-(\cdot)$ is an increasing function, this implies that the law generates a downward discontinuity at $\bar{o}$ in $S(o,1)$. Proposition 2 contains the key result of our theoretical analysis. While in the absence of law the esteem function decreases continuously in the magnitude of the externality imposed, the introduction of the law—and its associated material penalties—creates a sharp discontinuity in the esteem function at the legal threshold $\bar{o}$. We interpret this as a manifestation of "the expressive power of law": laws shape the social norms that prevail within a society by creating sharp discontinuities in the social rewards and sanctions accruing to individuals for taking legal or illegal actions.

Criminal Records: The result of Proposition 2 is driven by the presence of a material penalty for individuals caught breaking the law. However, by creating a record of an individual's criminal history, legal systems also enhance the visibility of (illegal) behavior within society at large, beyond the circle of people normally able to directly observe an individual's actions and the opportunities he/she faces.

To capture this second complementary effect, we extend the model to allow for the presence of "distant observers," socially distant people who cannot directly observe an individual's behavior, but can observe his/her criminal record, which acts as a public signal on whether the individual has been convicted for seizing an illegal opportunity. Individuals are concerned about the beliefs of both "close observers" (described in the previous subsections) and distant observers.

The full analysis incorporating close and distant observers is in online Appendix A. We assume, if an individual is caught seizing an illegal opportunity, the nature of the infraction (i.e., the size o of the externality generated) becomes known to distant observers, who assign to the individual the same esteem as close observers. On the other hand, if the individual has no criminal conviction (either because he/she has not broken the law or because he/she escaped conviction), distant observers only observe the absence of a criminal record and update beliefs about the individual's type accordingly. The setup closely resembles the framework described previously. Utility is given by

$$u_a(o;\theta) = (t - \theta o - pKI_{o>\bar{o}})a + S(o,a),$$

but now $S(o,a)$ represents *expected total esteem*, obtained from both close and distant observers. The conditions for $\hat{\theta}_o$ and $\tilde{\theta}_o$ must therefore account for this. We can show that, in monotone environments, the following result holds.[11]

PROPOSITION 3: *When both close and distant observers are present, the function $S(o,1)$ experiences a downward discontinuity at $\bar{o}$ also for the case where $K = 0$.*

Now the discontinuity result holds even when $K = 0$, as might occur for symbolic penalties or sentences involving a "public shaming" component as an alternative to prison, such as bumper stickers for DUI offenders (on this point see also Bénabou and Tirole [2011]). This is because, even if a convicted individual does not incur a material penalty, he/she nonetheless incurs a discontinuous reputational cost via the added visibility of behavior to distant observers.

Online Appendix A presents further extensions of our simple model to probe our result's robustness and study its properties. We relax our assumptions on the discrete (binary) action space and show our key result still holds in a more standard, Spence-like model where the individual directly chooses the level of externality. We also allow t, p, K and audience size to vary with o (actions with larger negative externalities may be more profitable, detectable, harshly sanctioned, and visible), which for simplicity we ruled out above. Our results continue to hold if we relax this assumption. Finally, we examine a series of factors that may affect the size of the discontinuity at the legal threshold (likelihood of enforcement, precision with which law enforcement can detect illegal behavior, and intentionality of criminal behavior). These comparative static results will guide our interpretation of the differences in expressive power of law observed in our empirical analysis across situations

[11] We discuss the sufficient conditions for monotonicity in the presence of both close and distant observers in online Appendix A. We talk about *expected* total esteem because distant observers observe illegal behavior only with probability p.

with stronger/weaker perceived tolerance, measurement error and intentionality of behavior.

II. Empirical Strategy: Main Experiment

Our empirical strategy mirrors the analysis of the previous section. Our empirical approach is novel in focusing on a special subset of laws regulating behavior by setting thresholds to distinguish legal and illegal actions, such as laws defining the minimum age for the sale of alcohol, the minimum age for sexual activities, the maximum speed one may drive, etc. This focus enables reliance on reasonably mild assumptions to resolve causal identification problems otherwise pervasive in the empirical literature on the topic. While it is, for example, difficult to defend assuming the enactment of a ban on sale of alcohol to minors is independent from pre-existing normative considerations about the appropriateness of alcoholic consumption by the young, a much less demanding assumption is that such a norm is unlikely to make a priori sharp distinctions between behaviors that are in all respects very similar. For instance, without any pre-existing drinking age limit, it is unlikely that a norm would sharply distinguish between selling alcohol to a customer aged 18 years and 1 month instead of 17 years and 11 months, such that this would inform the lawmaker's decision to position the legal threshold exactly at 18 years. This is because all factors that may matter for appropriateness judgments (e.g., how harmful drinking alcohol is at those ages) do not vary sharply across close age groups. This corresponds to our theoretical assumption in Section I of small differences in the magnitude of the externality implied by similar opportunities. If this assumption is valid, one can consider the existence of a sharp discontinuity in the underlying norm exactly at 18 years as causally determined by the existence of a legal limit at that age.[12]

Our reasoning here is similar to the arguments used to support the local randomization assumption in regression discontinuity designs. As in those, we assume the "outcome" variable—in our case, the esteem function $S(\cdot)$ defined in Section I—is continuous in the vicinity of the legal threshold, absent an expressive power of the law. If so, we can identify the causal effect of the law on norms by testing for a discontinuity in $S(\cdot)$ at the legal threshold.[13]

More precisely, our experiment will use incentivized norm-elicitation procedures, described below, to measure the social appropriateness of a series of actions varying in distance from a legal threshold $\bar{o}$ (for instance, the appropriateness of selling alcohol to a person aged 17 years and 10 months, 17 years and 11 months, 18 years and

[12] A possible threat to identification is that the position of the threshold itself may be decided based on "natural" discontinuities in $S(\cdot)$. For instance, if there is a general principle that "adulthood commences at 18," the function $S(\cdot)$ may be naturally discontinuous at that point, even with no law. The law may then simply embody and formalize this pre-existing discontinuity. Our identifying assumption excludes these cases. We believe this is a mild assumption, and particularly likely to hold in some of this study's vignettes where it is difficult to think of general social principles tied to the position of the threshold (e.g., speed limits, since these vary regularly across and within countries; the amount of cash that can be legally imported into one's country without declaring to customs).

[13] One difference between regression discontinuity designs and ours is that, as explained below, we measure the outcome variable $S(\cdot)$ among individuals *randomly assigned* (by us) to either side of the legal threshold. Thus, we need not worry about potential manipulations of the "assignment" variable on the part of subjects, which is a major concern in regression discontinuity designs.

1 month, 18 years and 2 months, when the law prohibits its sale to those under 18). We use the measurements of appropriateness for legal behavior to estimate its associated esteem, $S(o,1 \mid o \leq \bar{o})$, while we use the measurements of appropriateness for law-violating behavior to estimate $S(o,1 \mid o > \bar{o})$. Under the assumption that, absent the law, the function $S(o,1)$ is continuous in o, we identify the causal effect of the law on the social norm by estimating

$$(6) \qquad \big(S(o,1 \mid o \leq \bar{o}) - S(o,1 \mid o > \bar{o}) \mid o = \bar{o}\big)$$
$$= \lim_{\epsilon \to 0}\big[S(\bar{o} - \epsilon, 1) - S(\bar{o} + \epsilon, 1)\big].$$

Note that we will estimate the "social esteem" accruing to an individual for taking a certain action by measuring whether that action is perceived as "socially appropriate" or "socially inappropriate." This follows the empirical social norm compliance literature, which uses shared appropriateness judgments to measure the social approval and disapproval associated with given behavior (e.g., Krupka and Weber 2013; Görges and Nosenzo 2020). Section IV will report the results of an additional experiment using an alternative approach to elicit social esteem that corresponds more closely to the model of Section I (we measure directly the inferences from actions to prosocial traits).

Also, it is worth pointing out that legal thresholds may exert their expressive power on norms in ways beyond the discontinuity effects we set out to empirically identify. For instance, laws may affect the whole shape of the $S(\cdot)$ function, by changing how actions further away from the legal threshold are evaluated. For example, in many countries people consider it acceptable to speed up to a certain distance from a speeding threshold—they might consider speeds up to 74 miles per hour (mph) appropriate if the limit is 70 mph, but if the limit was reduced to 65mph they may then consider 74 mph inappropriate. Our strategy is designed to measure the effect of law exactly at the legal threshold and therefore does not capture these additional expressive effects that may nevertheless be empirically relevant.

A. *Experimental Design*

To directly measure the effect an action's legality has on the social norm pertaining to it, our experiment used *vignettes*. We presented subjects with a series of hypothetical scenarios describing a fictitious person's behavior, and in each vignette elicited subjects' evaluations of its social appropriateness. We used five vignettes describing situations where the legality of some behavior is determined by the side of a legal threshold it falls. We considered five types of legal threshold, concerning (i) the age of consent, (ii) the legal drinking age, (iii) the maximum amount of cash which can be imported into one's country without declaring to customs, (iv) the BAC drink-driving limit, and (v) the driving speed limit on a motorway.

The five vignettes each described a situation involving one of these legal thresholds. The age of consent vignette described an adult having sex with a younger person he had met at a party. The alcohol to youth vignette described a shopkeeper selling alcohol to a youth known to be a local vandal. In the cash at customs vignette, a person was returning from abroad with a cash amount that he did not declare at customs. In the drink driving vignette, a woman drove home after drinking on a night out. Finally, the speeding vignette described a woman driving on a motorway. The vignettes are reproduced in online Appendix B.

We designed these five vignettes to achieve variation in the illegal behavior's severity as well as the extent to which behavior, even if legal, would be deemed socially inappropriate. For instance, in the alcohol to youth vignette, we made the customer a local vandal to reduce the appropriateness of selling him alcohol even when he was legally allowed to buy it.[14] We also chose situations differing in relevant features of law enforcement, such as the ability to monitor or accurately detect whether a behavior exceeds the legal threshold, which we will exploit to shed light on the possible mechanisms underlying our observed effects.

We always made it clear the person in the vignette knew the legal threshold and could verify which side their own behavior would fall. For example, in the age of consent vignette, the adult checks the younger person's ID card to verify whether she is above the age of consent. We deemed this important for two reasons. First, we wanted to subtly remind (or inform) our subjects about the existing relevant legal rules. Second, we wanted subjects to evaluate the behavior of a person knowingly following or breaking the law, to remove any ambiguity about potential "ignorance of the law," which may have affected judgments of appropriateness.

For each situation, we designed eight (or four, depending on the sample—see below) different versions of the vignette, differing only in the described behavior's side of the legal threshold and distance from it. This included only just legal or only just illegal behaviors, so as to measure the appropriateness of actions virtually identical in all respects except their legal status. For instance, for the age of consent situation, we designed versions of the vignette with the younger person one, two, three, or four months above the age of consent, and versions with her one, two, three, or four months below it.

The different versions of the vignettes were administered in a between-subject design: each subject evaluated only one behavior per situation. For example, some subjects were (only) described the vignette with the younger person one month above the age of consent, others were (only) described the vignette with her two months above, etc. These between-subject measurements of appropriateness allow us to obtain, for each situation, an estimate of the norm function $S(\cdot)$ regulating behavior in a neighborhood around the relevant legal threshold. Our identification strategy consists of testing, for each of the five vignettes, whether $S(\cdot)$ is discontinuous at the legal threshold.

Our experiments also included ten additional filler vignettes, which, along with the five legal threshold vignettes that are our focus, were presented in random order (except that all subjects' first three vignettes were fillers, for reasons explained in footnote 17). These were included to avoid it becoming salient that we were interested in the evaluation of behaviors regulated by legal thresholds. Thus, the filler vignettes featured various behaviors either unregulated by law (e.g., refusing to give a beggar money) or regulated by law but not via legal thresholds (e.g., leaving a

[14] Moreover, in this specific vignette we focused on the evaluation of the shopkeeper's behavior, rather than the customer's, as this better captures the notion of taking opportunities for material gain that may be harmful to others.

restaurant without paying). The filler vignettes were not manipulated (i.e., we did not prepare different versions of them), so each one was identical for every subject.

B. *Incentivization*

Subjects received monetary incentives to evaluate the social appropriateness of the behavior in the vignettes. We used two different incentive schemes, corresponding to the two most popular existing methods used to elicit social norms. One procedure was proposed by Krupka and Weber (2013), which we refer to as the "Krupka-Weber method." The other procedure has been used in different guises by multiple authors, e.g., Bicchieri and Xiao (2009); d'Adda et al. (2020); Bursztyn, González, and Yanagizawa-Drott (2020). For reasons that will become clear, we refer to it as the "opinion matching method."

In the Krupka-Weber method, subjects indicated for each vignette how socially appropriate they thought the described behavior was by selecting one option on a four-point ordered scale: "very socially appropriate," "somewhat socially appropriate," "somewhat socially inappropriate," or "very socially inappropriate." They were paid a bonus in addition to a participation fee only if their evaluation of the behavior in a vignette was the same as that chosen by the most other subjects in the same version of the experiment; otherwise, they were only paid the participation fee.

In the opinion matching method, subjects were randomly assigned to one of two different conditions (between-subject design). In one condition, which was run first, subjects had to report their personal belief of how appropriate the behavior described in the vignette was. Responses, not incentivized, were indicated on a four-point scale, as above, but which used the terms "appropriate/inappropriate" rather than "socially appropriate/inappropriate."[15] In the second condition, subjects had to guess the most common appropriateness judgment among the first group. These respondents were probabilistically paid a bonus on top of their participation fee if their guess was correct, and were only paid the participation fee otherwise.

Both methods reward subjects for accurately reporting their perception of how appropriate a particular behavior is *commonly* regarded (i.e., second-order beliefs of appropriateness), rather than their own personal evaluation of the behavior.[16] This is important since social norms reflect opinions about what is *collectively approved or disapproved of within a society*, rather than personal opinions about appropriateness

[15] In each case, we provided subjects with lengthy explanations on how to understand these terms. See online Appendix C for full details. For social appropriateness, the explanation began: "By socially appropriate, we mean behaviour that you think most people would agree is the 'right' thing to do. Another way to think about what we mean is that if someone were to behave in a socially inappropriate way, then other people might be angry at them." For appropriateness, the wording was similar but dropped "most people would agree" and replaced "other people" with "you."

[16] The opinion matching method also delivers a measure of personal norms (i.e., first-order beliefs of appropriateness) from the unincentivized participants. A potential criticism of the Krupka-Weber method is that, conceptually, it is unclear whether it measures second-order or higher-order beliefs of appropriateness. Subjects are asked to report their perception of the social norm (i.e., second-order beliefs about what most others think is appropriate). A subject wanting to coordinate may however form *third-order* beliefs of appropriateness, i.e., beliefs about others' second-order beliefs (this logic can be iterated further, resulting in even higher order beliefs). However, a counterargument is that second-order beliefs are the most salient high-order belief in the task because the instructions are heavily framed in the language of social norms, and, if subjects use salient focal points to coordinate (Schelling 1960), they should indeed report second-order beliefs.

(for a discussion of the difference between personal opinions and social norms, see Bicchieri 2006; Krupka and Weber 2013).

The Krupka-Weber method achieves this by transforming the task into a coordination game where subjects are incentivized to rate behavior in the same way as other simultaneously participating subjects. The rationale is that, if a norm exists regarding the behavior being evaluated, this constitutes a particularly salient focal point subjects can use to successfully coordinate. As such, subjects' evaluations indirectly reveal the underlying social norm pertaining to the behavior in the vignette. However, one concern is that in principle subjects may instead use any other focal points to coordinate. Because our vignettes explicitly mention the law, subjects may use legality itself as a focal principle for coordination, and rate legal actions as "appropriate" and illegal actions as "inappropriate," regardless of whether the social norm truly prescribes this. This alternate coordination strategy would also give us a discontinuity at the threshold—but for the wrong reason. To minimize this concern, our experiment was designed to emphasize the distinction between the concepts of "social appropriateness" and "legality" and to increase the relative salience of the former.[17]

The opinion matching method sidesteps this concern altogether, since subjects were incentivized to guess the most prevalent opinion among another group who had no strategic incentive to coordinate. However, a possible concern here is that subjects are incentivized to match the *unincentivized* responses of others. If these first-step responses are vulnerable to noise or responding biases, second-step responses may reproduce the same effects (see, e.g., Aycinena, Bogliacino, and Kimbrough 2022).

As each method has advantages and disadvantages, we report norm elicitations based on both. Subjects were randomly assigned to one procedure only (between-subject design). Moreover, if assigned to the opinion matching method, they either participated in the nonincentivized or incentivized condition.

C. *Samples and Procedures*

Our main experiment was run between September 2017 and March 2021 with a total of 1,248 participants separately recruited from three different UK samples: one student sample and two from the general population. Table 1 summarizes the samples used.

The student sample comprised 197 British students at the University of Nottingham. Recruited in September 2017, they completed the experiment using the Krupka-Weber method. Subjects were told, to earn the bonus from the vignette

[17] We included two design features to achieve this. First, the instructions told subjects what constitutes appropriate behavior " … may not necessarily be made explicit or supported by laws, nor enforced by the threat of legal sanctions. So an action may be 'appropriate' even if it is not legal; or 'inappropriate' even if it is not illegal." This passage aims to reduce the incentive to use legality as a norm-unrelated coordination device, by breaking any cycle of beliefs supporting it as a successful coordination strategy (subjects should now doubt others may use legality to coordinate as they are explicitly told not to). Second, the first three vignettes subjects evaluated were always fillers explicitly designed to train them to think of social appropriateness as a concept distinct from legality. These vignettes described behavior unlikely to be considered very inappropriate, but which in one case was regulated by law and legal (a person deciding not to illegally download a movie), in another regulated by law and illegal (a person driving very slowly and safely without wearing a seatbelt), and in the third case unregulated by law (a person choosing between booking a holiday and giving money to charity).

Table 1—Samples Used in the Main Experiment

	Observations	Nationality	Year	Subject pool type	Method
Sample 1	197	British	2017	Students	Krupka-Weber
Sample 2	375	British	2019	General population[a]	Krupka-Weber
Sample 3	676	British	2021	General population[b]	Opinion matching

Notes:
[a] Representative in terms of gender, age, and yearly income.
[b] Representative in terms of gender, age, and ethnicity.

task, they had to match the responses of other participants of their own sample (i.e., other University of Nottingham students). For each vignette with a legal threshold, subjects were randomly assigned to one of four possible versions of the vignette. Thus, our estimates of the norm function $S(\cdot)$ rely on 4 distinct measurements (2 legal and 2 illegal) per vignette, with approximately 50 subjects in each case. Students completed the experiment online in around 10 minutes, and one-fifth were selected for payment. The selected subjects were paid a £10 participation fee, plus a £30 bonus if they had successfully coordinated in one of the 15 vignettes they evaluated (5 target vignettes + 10 fillers), randomly selected after the experiment.

To probe the generalizability of our findings, we repeated the experiment using two samples of the UK general population. The first, recruited in March 2019, consisted of 375 British participants recruited by the online panel survey company Qualtrics. We set recruitment quotas to obtain a sample representative of the general population along three dimensions: gender (51 percent female), age (11 percent aged 18–24; 21 percent aged 25–34; 23 percent aged 35–44; 24 percent aged 45–54; 21 percent aged 55 and older), and yearly income (23 percent less than £20,000; 42 percent £20,000–£40,000; 20 percent £40,000–£60,000; 15 percent more than £60,000).

The experiment again used the Krupka-Weber method with subjects told they had to coordinate within their own sample (other British individuals recruited through Qualtrics). Again, subjects were randomly assigned to one version of each of the five vignettes with legal thresholds. This time, however, we designed eight different versions of each vignette (four legal and four illegal), to increase the precision of our estimate of the norm function $S(\cdot)$. All subjects received a base incentive of approximately £0.40 for participating online. In addition, we randomly selected one-fifth of participants and paid them (through Qualtrics) according to the same rules used for the UK student sample (£30 for successful coordination on one randomly selected vignette).

The second UK general population sample was recruited using a different online sample provider, Prolific. The experiment was preregistered (Lane, Nosenzo, and Sonderegger 2021). We recruited 676 subjects in March 2021 (but, to keep comparability with the earlier samples, subjects were asked to evaluate behavior in the vignettes in a prepandemic world). We again set recruitment quotas to obtain a sample representative of the general population in respect to gender (51 percent female), age (9 percent aged 18–24; 17 percent aged 25–34; 19 percent aged 35–44; 17 percent aged 45–54; 38 percent aged 55 and older), and ethnicity (roughly 81 percent white; 7 percent Asian; 5 percent black; 4 percent mixed; 3 percent other; however, we are missing ethnicity data for approximately 4 percent of the subjects).

The experiment used the opinion matching method: 342 subjects were assigned to the unincentivized condition and 334 to the incentivized condition. The first group was asked their personal opinions about the appropriateness of the described behavior (first-order beliefs), while the second group had to guess, for each vignette, the most frequent appropriateness judgment of the first group (second-order beliefs). We again used eight different versions of each vignette (four legal and four illegal). All subjects received a base incentive of £1.88 for participating online. In addition, we randomly selected one-fifth of the subjects in the incentivized condition and paid them (through Prolific) £30 if they matched the first group's most common response in one randomly selected vignette.

III. Results

Figure 1 plots the norm functions—showing the average (social) appropriateness of the behaviors subjects evaluated—elicited in the five legal threshold situations in our three samples (see online Appendix D for the full distributions of appropriateness ratings). Following convention in the social norms literature, we assign evenly-spaced values of $+1$ to the rating "very (socially) appropriate," $+0.33$ to "somewhat (socially) appropriate," -0.33 to "somewhat (socially) inappropriate," and -1 to "very (socially) inappropriate." Thus, the norm functions $S(\cdot)$ assume positive values for actions evaluated on average as appropriate, and negative values for inappropriate actions. Blue circles show the function values for the student sample, while red squares show them for the general population samples. Filled squares indicate the responses of the 2019 sample (elicited with the Krupka-Weber method). The 2021 sample's responses were elicited with the opinion matching method, delivering both first-order and second-order beliefs of appropriateness; Figure 1 plots both (dotted squares for first-order beliefs; empty squares for second-order beliefs).

In each panel, the dashed black line indicates the legal threshold. Actions left of it are legal, while those to the right are illegal. The first three panels reveal, in all samples, the legal threshold exerts a very strong influence on the norm function, with sharp drops in (social) appropriateness values as we move from the legal to illegal side. For the age of consent vignette, there is a drop of between 0.96 units (general population 2019) and 0.74 units (students) as the age of the young person changes from 16 years and 1 month (legal) to 15 years and 11 months (illegal). For the vignette where a shopkeeper sells alcohol to a youth, the drop is of between 1.10 units (general population 2021, second-order beliefs) and 0.86 units (general population 2019) as the customer's age changes from 18 years and 1 month (legal) to 17 years and 11 months (illegal). Finally, in the cash at customs vignette, there is a drop of between 0.97 units (general population 2021, second-order beliefs) and 0.86 units (general population 2021, first-order beliefs) as the person imports an amount of cash €100 above rather than below the legal threshold. In contrast, in all cases the small increments in the running variables (age and cash amount) are clearly inconsequential for behaviors that are both on the legal side of the threshold, or both on the illegal side.

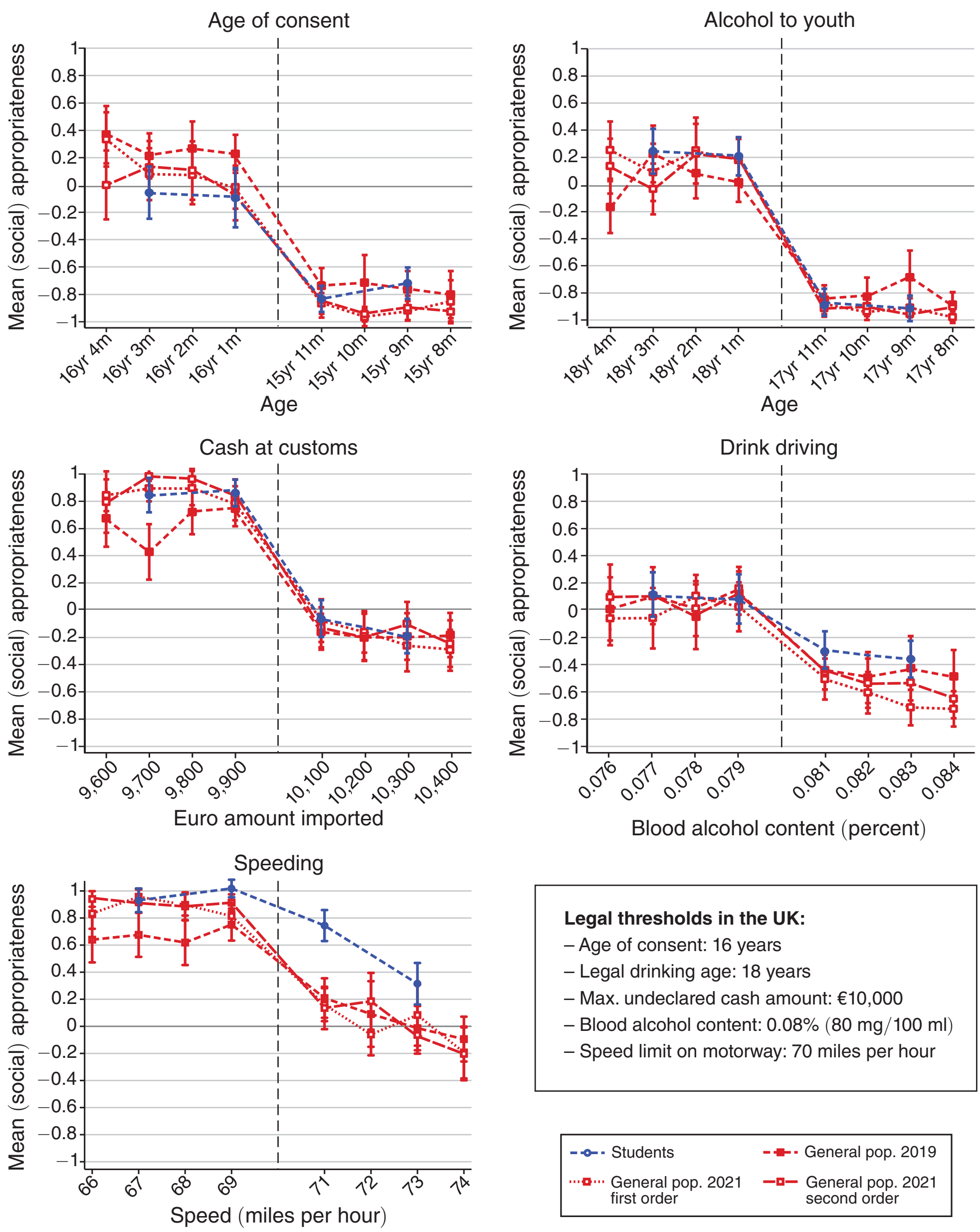

Figure 1. Norms in the Five Legal Threshold Situations, UK Samples

Notes: Each panel plots the average (social) appropriateness of actions at various distance from a legal threshold (1 = very [socially] appropriate, −1 = very [socially] inappropriate). The dashed black line indicates the position of the legal threshold in each situation (values of the thresholds are reported in the bottom-right box). Actions to the left of the threshold are legal, actions to the right are illegal. Bars are 95 percent confidence intervals.

The drop in appropriateness at the threshold is, however, smaller in the drink driving and speeding vignettes, for both the student and general population samples. Here, the functions tend to decrease over the range of behavior measured, but without such sharp threshold discontinuities. For the drink driving vignette, the appropriateness drop varies between 0.61 units (general population 2021, second-order beliefs) and 0.38 units (students) as the BAC changes from 0.079 percent (legal) to 0.081 percent (illegal). For the speeding vignette, appropriateness drops by between 0.78 units (general population 2021, second-order beliefs) and 0.27 units (students) as speed changes from 69 mph (legal) to 71 mph (illegal).

Based on the identification strategy sketched in equation (6), we formally examine these patterns by estimating the following regression model for each vignette:

$$(7) \qquad s(o_i) \;=\; \alpha + \beta_1(T - o_i) + \beta_2 Illegal_i + \beta_3(T - o_i) \times Illegal_i + \epsilon_i,$$

where $s(o_i)$ is subject i's evaluation of appropriateness of behavior in the vignette describing opportunity o_i, $(T - o_i)$ measures the distance between the legal threshold and opportunity o_i, $Illegal_i$ is a dummy equal to one if subject i evaluated a version of the vignette containing illegal behavior and zero otherwise, and ϵ_i is the error term. This model allows the slope of the relationship between appropriateness and distance from the threshold to differ between legal and illegal opportunities. The coefficient β_1 measures the relationship for legal opportunities, i.e., the slope of the esteem function $S(\cdot)$ below the legal threshold.[18] The coefficient β_3 measures how this slope changes for illegal, rather than legal, opportunities; i.e., it allows us to derive the slope of $S(\cdot)$ for opportunities exceeding the threshold. The coefficient of most interest is β_2, which measures the difference between the estimates of the norm function for opportunities just above or just below the legal threshold T, thus capturing the discontinuity of the norm at T, i.e., the causal effect of law on normative considerations.

We estimated equation (7) separately for each sample and vignette, using ordinary least squares (OLS) regressions with heteroskedasticity-robust standard errors. Table 2 shows the results, in panel A for the student sample, panel B for the 2019 general population sample, and panels C (first-order beliefs) and D (second-order beliefs) for the 2021 general population samples.[19]

Starting with the students, the estimate of the coefficient β_2 is negative and highly significant in models A1, A2, and A3 (the age of consent, alcohol to youth, and cash at customs vignettes), indicating strong discontinuities at the legal thresholds for these situations (the magnitude of β_2 ranges from -0.778 to -1.035 across the three vignettes). In contrast, the estimates of β_2 are much smaller in models A4 and A5 (the drink driving and speeding vignettes). The coefficient is in fact not significant for the speeding vignette, and only significant at the 10 percent level for the drink driving vignette ($p \;=\; 0.068$).

Similar patterns emerge in the general population samples (panels B, C, and D). Here we also find strong discontinuities in the norm functions for the age of consent, alcohol to youth, and cash at customs vignettes (coefficients ranging from -0.803

[18] In two of our five vignettes (age of consent and alcohol to youth) opportunities below the threshold are illegal, while in the other three opportunities in excess of the threshold are illegal. To ease interpretation, we code $(T - o_i)$ so that positive values are always assigned to legal opportunities and negative values to illegal ones. In other words, the variable is actually defined as $(o_i - T)$ for the age of consent and alcohol to youth vignettes, while it is defined as $(T - o_i)$ for the other three vignettes.

[19] For the general population samples we also have data on participants' age, gender and income, which we use as controls in the regressions (not shown in Table 2). We did not collect any sociodemographic data from the students.

Table 2—OLS Regressions, UK Samples

Panel A. Students	Age of consent (A1)	Alcohol to youth (A2)	Cash at customs (A3)	Drink-driving (A4)	Speeding (A5)
$(T - o_i)$	0.019	0.019	−0.007	0.016	−0.044
	(0.072)	(0.054)	(0.039)	(0.061)	(0.027)
Illegal	−0.778	−1.035	−0.866	−0.326	−0.103
	(0.184)	(0.138)	(0.132)	(0.178)	(0.107)
$(T - o_i) \times$ *Illegal*	−0.078	0.004	0.072	0.017	0.258
	(0.081)	(0.064)	(0.061)	(0.077)	(0.055)
Constant	−0.058	0.246	0.869	0.067	0.977
	(0.167)	(0.114)	(0.079)	(0.141)	(0.053)
Controls	No	No	No	No	No
R^2	0.293	0.613	0.567	0.139	0.319
Observations	197	197	197	197	197

Panel B. General population, 2019	Age of consent (B1)	Alcohol to youth (B2)	Cash at customs (B3)	Drink-driving (B4)	Speeding (B5)
$(T - o_i)$	0.026	−0.029	−0.055	−0.014	−0.039
	(0.038)	(0.039)	(0.038)	(0.043)	(0.033)
Illegal	−0.890	−0.920	−0.948	−0.522	−0.461
	(0.127)	(0.118)	(0.124)	(0.143)	(0.127)
$(T - o_i) \times$ *Illegal*	−0.001	0.034	0.058	0.024	0.145
	(0.051)	(0.045)	(0.051)	(0.058)	(0.049)
Constant	0.215	0.264	0.596	−0.040	0.623
	(0.140)	(0.119)	(0.145)	(0.153)	(0.126)
Controls	Yes	Yes	Yes	Yes	Yes
R^2	0.467	0.405	0.373	0.160	0.263
Observations	375	375	375	375	375

(*continued*)

to -1.137), but weaker effects in the drink driving and speeding vignettes, where the coefficients are roughly half the magnitude of those of the other three vignettes (ranging from -0.461 to -0.592), although always strongly significant.

In all samples, Chow tests confirm there are no significant differences between the coefficients of the *Illegal* variable in the first three vignettes, $\beta_2^{consent}$, $\beta_2^{alcohol}$, and β_2^{cash} (all $p \geq 0.136$), or between the estimates of $\beta_2^{drink-drive}$ and $\beta_2^{speeding}$ (all $p \geq 0.347$).[20] We instead find statistically significant differences between the estimates of the first and second group of coefficients. Specifically, among students, we find significant differences in all such comparisons (all $p \leq 0.028$) except between $\beta_2^{drink-drive}$ and $\beta_2^{consent}$ ($p = 0.124$); in the 2019 general population sample, we find significant differences in all such comparisons (all $p \leq 0.088$); in the 2021 general population sample we find significant differences in first-order beliefs between $\beta_2^{alcohol}$ and both $\beta_2^{drink-drive}$ and $\beta_2^{speeding}$ (both $p \leq 0.009$), and in

[20] The Chow test p-values we report have been adjusted to take into account the multiple comparison problem, using the Benjamini-Hochberg false discovery rate method (Benjamini and Hochberg 1995; Simes 1986).

Table 2—OLS Regressions, UK Samples (*continued*)

Panel C. General population, 2021 first order	Age of consent (C1)	Alcohol to youth (C2)	Cash at customs (C3)	Drink-driving (C4)	Speeding (C5)
$(T - o_i)$	0.096	0.003	0.016	−0.041	0.010
	(0.039)	(0.041)	(0.029)	(0.042)	(0.023)
Illegal	−0.803	−1.047	−0.821	−0.487	−0.592
	(0.119)	(0.116)	(0.129)	(0.143)	(0.097)
$(T - o_i) \times$ *Illegal*	−0.107	0.016	0.055	0.132	0.094
	(0.047)	(0.044)	(0.046)	(0.053)	(0.040)
Constant	−0.323	0.299	0.750	0.204	1.063
	(0.141)	(0.145)	(0.129)	(0.149)	(0.101)
Controls	Yes	Yes	Yes	Yes	Yes
R^2	0.509	0.610	0.554	0.281	0.525
Observations	332	332	332	332	332
Panel D. General population, 2021 second order	Age of consent (D1)	Alcohol to youth (D2)	Cash at customs (D3)	Drink-driving (D4)	Speeding (D5)
$(T - o_i)$	0.041	−0.042	0.011	−0.019	0.013
	(0.047)	(0.040)	(0.035)	(0.044)	(0.015)
Illegal	−0.813	−1.137	−0.971	−0.542	−0.577
	(0.138)	(0.113)	(0.127)	(0.139)	(0.109)
$(T - o_i) \times$ *Illegal*	−0.030	0.044	0.015	0.078	0.108
	(0.053)	(0.043)	(0.050)	(0.054)	(0.041)
Constant	−0.350	0.224	0.879	0.161	0.979
	(0.158)	(0.125)	(0.120)	(0.145)	(0.094)
Controls	Yes	Yes	Yes	Yes	Yes
R^2	0.478	0.595	0.568	0.253	0.549
Observations	334	334	334	334	334

Notes: Dependent variable is the evaluation of appropriateness of the behavior described in a vignette. Robust standard errors in parentheses. Regressions with bootstrapped standard errors yield very similar results. Controls (age, gender, and income) are included in the regressions of panels B, C, and D, but not reported in the table. In panel C, we have 332 observations (instead of 342) because 10 subjects have missing values for some of the control variables.

second-order beliefs in all comparisons (all $p \leq 0.056$), except those involving $\beta_2^{consent}$ (both $p = 0.256$).[21]

It is reassuring that the results of the student sample are successfully replicated using representative samples of the broader population, and—especially given the methodological concerns discussed above—that our data show remarkable similarities across the two methods used to measure norms, regardless of whether or not players had material incentives to coordinate with others (also see Bicchieri et al. [2022], who report consistent results across these two different methods). Overall, the results show the law can have a strong influence in shaping the norms governing

[21] We also conducted a robustness (placebo) analysis where we re-estimated regression equation (7) replacing the *Illegal* dummy with a "placebo" dummy defined in relation to fictitious thresholds, different from the actual legal threshold. We report results in online Appendix E. The analysis confirms the systematic nature of the discontinuities at the legal threshold observed in the experiments as compared to these placebo discontinuities.

behaviors targeted by the law. However, they also show the expressive power of law does not hold uniformly across all situations. In particular, our data show that, in the United Kingdom, laws related to driving behaviors seem to hold weak power on the underlying norms. In online Appendix E we explore potential explanations for this variability in the expressive power of law. We show, using data from follow-up questions included at the end of our general population experiments, that illegal behavior in the speeding and drink-driving vignettes is perceived to be relatively difficult for law enforcement to accurately measure, and relatively likely to occur unintentionally. We also show the estimated effects of laws on norms are generally weaker among subjects who believe illegal behavior may be unintentional or difficult to measure. Overall, this analysis is consistent with our model that, as shown in online Appendix A, predicts the perceived intentionality and measurability of behavior can be moderators of the expressive power of law, and thus explain some of the between-vignette variability in our estimated discontinuities. We also find some support for police tolerance toward illegal behavior being a moderator—which is also consistent with our model—though here the evidence is rather weaker.

IV. Alternative Mechanisms and Robustness Analysis

In this section we consider alternative explanations for our results, and report four additional experiments designed to further probe their robustness. Table 3 presents an overview of the experiments.[22] In experiment 1 we conducted *placebo tests* to rule out that the discontinuities observed at the legal thresholds in the main experiment are driven by the use of legality as a norm-unrelated focal point, or by an information transmission mechanism whereby individuals use legal thresholds to learn about community standards. In experiment 2 we employed an alternative design which, instead of measuring the esteem function $S(\cdot)$ through perceptions of actions' social appropriateness, directly elicited subjects' inferences about a person's *prosocial traits*, upon observing his/her actions. Experiment 3 elicited inferences about prosocial traits under a *"bad" law* which, if followed, would harm a third party. We use this experiment to assess a meta-norm explanation of our results, whereby law breaking is inappropriate because it violates a meta-norm of legal obedience. Experiment 4 probed our results' generalizability by studying the expressive power of law in a country (China) where the *rule of law* is relatively weak compared to the United Kingdom and United States, where our other experiments were run. Finally, Section IVE discusses an alternative *conformity-based mechanism*, which we are unable to empirically rule out but argue is a weaker candidate explanation than the mechanism we propose.

A. *Experiment 1: Placebo Thresholds*

One concern with the results in Section III is that the downward discontinuities at the legal threshold in the norm functions may be caused by two alternative mechanisms distinct from our preferred interpretation (that the law produces discontinuous

[22] We preregistered experiments 1, 2, and 3 (see Lane, Nosenzo, and Sonderegger 2021, 2022).

TABLE 3—OVERVIEW OF THE ADDITIONAL EXPERIMENTS

	Observations	Nationality	Year	Subject pool type	Method
Experiment 1 (placebo thresholds)	1,554	British	2021	General population	Opinion matching and Krupka-Weber
Experiment 2 (prosocial traits)	2,767	British	2021	General population	Opinion matching
Experiment 3 ("bad" law)	1,202	American	2022	General population	Opinion matching
Experiment 4 (weaker rule of law)	248	Chinese	2017	Students	Krupka-Weber

social esteem at the threshold through its impact on the inferences drawn about a perpetrator's "type"). The first is especially relevant for the Krupka-Weber method and relates to the aforementioned concern that the legal threshold may give subjects a salient focal point to coordinate their responses. Despite the precautions adopted in our design (see Section IIB) and the encouraging results reported at the end of the previous section, one may still question whether subjects coordinated by rating illegal actions as "inappropriate" and legal actions as "appropriate," irrespective of whether this truly reflected their perception of the social norm. Such a coordination strategy would produce a downward discontinuity at the threshold, even if the law had no expressive power.

The second mechanism could arise if subjects believe the exact position of the legal threshold chosen by a government reflects information, privy to the government, about the existence of sharp variations at the threshold in some relevant aspects of the decision situation. For instance, the government may set the legal drinking age limit at 18 years based on information that the harms of alcohol are discontinuously higher below that age. As explained in Section II, our identification strategy rules out this possibility by assumption. We believe the assumption is justified: although considerations about the harms of alcohol by age are undoubtedly factored into the chosen legal drinking age limit, it is very unlikely that this information sharply discriminates between points very near the threshold (e.g., 18 years and 1 month versus 17 years and 11 months). Nevertheless, subjects may *believe* this is the case and thus react to the threshold's position *as if* it was indeed carrying information about some sharp variation in the function $S(\cdot)$ at that exact point, that would exist even without a law. If so, the expressive power of law identified in the main experiments could be at least partly driven by a similar information transmission mechanism to those discussed in McAdams (2000, 2015); Bénabou and Tirole (2011); and Bursztyn, Egorov, and Fiorin (2020).

To probe the robustness of our results against these alternative explanations, we designed an experiment introducing a *placebo threshold* in each of our five vignettes. The placebo threshold was always positioned at a close distance from the actual legal threshold (five or six "units" above or below the legal threshold).[23]

[23] We set the placebo thresholds at 75 mph for the speeding vignette (legal threshold: 70 mph), 0.075 percent BAC for the drink driving vignette (legal threshold: 0.08 percent) and €10,500 for the customs vignette (legal threshold: €10,000). For thresholds based on age, it felt more natural to place the placebo threshold half a year from the actual legal threshold. We therefore placed them 6 months below (age of consent) or above (legal drinking age) the UK legal threshold (16 and 18 years, respectively).

We introduced the placebo thresholds in the vignettes using narratives of a fictitious group of people advocating an alternative limit to the described behavior, either above (i.e., more permissive) or below the actual legal threshold. For instance, in the speeding vignette, the described person recalls hearing of “a petition to raise speed limits on motorway to 75 mph.” Across vignettes, we changed the narratives to create variation in the extent subjects could interpret the placebo thresholds as conveying information about the decision situation. We reasoned subjects may be likelier to believe the placebo threshold has informational value if the alternative limit in the hypothetical decision situation is advocated by a group more representative of society and/or with more expertise. Therefore, the “high informational content” narratives described the placebo thresholds as proposed by experts and/or relatively large lobbying groups, such as “a panel of scientists” (drink driving vignette), a public “petition” (speeding vignette), or a “campaign group” (alcohol to youth vignette).[24] The “low informational content” narratives revolved instead around the opinions of a smaller number of unqualified people, such as a “group of friends” (age of consent vignette) and “custom officials working in the airport” (cash at custom vignette).

We always explicitly mentioned that the people advocating the alternative limit believed the placebo threshold neatly separated appropriate from inappropriate behavior. For example, in the speeding vignette we said the fictitious petition argued “it is appropriate to drive at speeds up to 75 mph, and inappropriate at higher speeds.” We did this to maximize the placebo threshold’s salience as a focal coordination point. Thus, the vignette explicitly spelled out a strategy subjects could use to coordinate, by rating behavior as appropriate if below the placebo threshold, and inappropriate otherwise.

Our test consists of measuring whether the placebo thresholds produce discontinuities in the norm function analogous to those produced by the legal thresholds. If similar-sized discontinuities systematically appeared at the placebo thresholds, this would question our interpretation of the main experiment’s results. The mechanism we propose is that discontinuities in the social norm function are driven by payoff discontinuities at the legal threshold that separate sharply between individuals who take marginally legal and illegal actions. These payoff discontinuities do not arise at the placebo thresholds (exceeding the placebo limit does not trigger a criminal record and/or material sanction). Thus, observing systematic discontinuities in the $S(\cdot)$ function at the placebo threshold would suggest at least part of the effect identified in Section III is driven by alternative mechanisms. We can quantify the observed influence of these alternative mechanisms by comparing the magnitude of the discontinuities at the legal and placebo thresholds (if any), and by comparing discontinuities in vignettes with high versus low informational content placebo thresholds.

We recruited 1,554 subjects from the UK general population via the online platform Prolific in May to June 2021. We elicited appropriateness judgments using both the Krupka-Weber method (653 subjects) and the opinion matching method (901 subjects; 260 assigned to the nonincentivized condition and 641 to the

[24] We avoided directly using the government or parliamentary committees as advocates of the alternative limits because we feared this may signal an imminent change of law, introducing a different reason to respond to the placebo threshold.

incentivized condition).[25] Other than including the placebo thresholds in the five target vignettes, the experiment was in all important respects identical to that described in Section II.[26] The incentives were also similar to the earlier experiments' (£1.88 participation fee plus, for a randomly-selected fifth of those doing the incentivized tasks, a £30 bonus for matching the most common response in a randomly selected vignette).

Figure 2 shows the main results, both from the Krupka-Weber method and the opinion matching method (we present only the second-order beliefs data).[27] The new experiments reproduce the patterns observed in Section III: strong discontinuities at the legal thresholds for the age of consent, alcohol to youth, and cash at customs vignettes, under both Krupka-Weber and opinion matching methods, and more modest discontinuities in the drink driving and speeding vignettes.

The placebo thresholds nearly always have a markedly smaller influence on the norm functions, regardless of the elicitation method. For the age of consent, alcohol to youth, and cash at customs vignettes, Figure 2 shows hardly any discontinuity at the placebo threshold. For the speeding vignette, the discontinuity measured with the opinion matching method goes in the opposite direction than one would expect (exceeding the placebo threshold increases appropriateness). Only in the drink driving vignette do we observe a discontinuity at the placebo threshold roughly of the same (small) magnitude as observed at the legal threshold.

We formally analyze these effects using similar regression models to those presented earlier in equation (7), which also include a dummy variable for the placebo threshold and corresponding interaction with the distance from threshold variable. We report more details on the model and the full regression estimates in online Appendix F.

The regressions reproduce our earlier results on the effects of the legal thresholds on the norm functions. Under the Krupka-Weber method, we find the placebo threshold produces a marginally significant discontinuity only in the cash at customs vignette ($p = 0.054$). The size of the placebo discontinuity is 0.22 compared to 0.88 at the legal threshold, and the latter is significantly larger than the former ($p = 0.000$). Under the opinion matching method, we find only two cases of a significant discontinuity at the placebo threshold. The speeding vignette sees an increase in appropriateness of 0.22 after the placebo threshold, significant at the 5 percent level ($p = 0.042$). The magnitude of the discontinuity at the corresponding legal threshold (0.61) is significantly larger than that at the placebo ($p = 0.026$). The drink driving vignette has a statistically significant discontinuity at the placebo ($p = 0.000$), roughly of the same size as that observed at the legal threshold ($p = 0.878$).

[25] We did not recruit a representative sample for any of the robustness experiments because this constraint would have made it impossible to recruit enough subjects, given our budget and the platform pool of volunteers. Also for budget reasons, we recruited a smaller sample in the nonincentivized condition, only with the purpose of using it to incentivize the elicitation of second-order beliefs.

[26] The complete wordings of the placebo vignettes are in online Appendix B, while details of minor changes to the preliminary screens of the experimental instructions, relative to our main experiment, are available in online Appendix C.

[27] Although they are not of primary interest, the complete response distributions in the first-order beliefs condition are also reported in online Appendix D. This broadly demonstrates patterns of responses to the two conditions of the opinion matching method were similar (i.e., second-order beliefs elicited in the incentivized condition tended to be well calibrated)

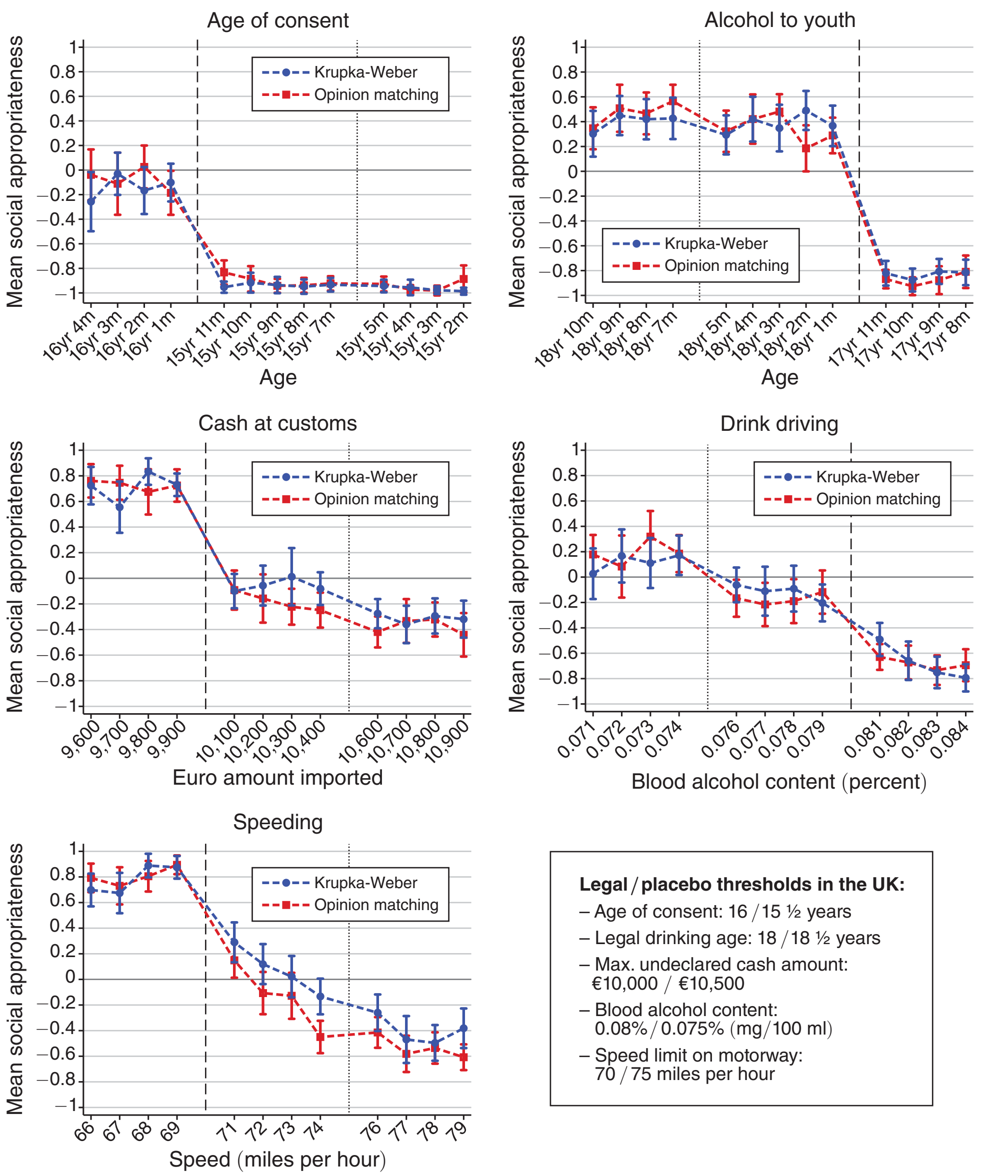

FIGURE 2. LEGAL AND PLACEBO THRESHOLDS

Notes: Each panel plots the average social appropriateness of actions at various distance from a legal and a placebo threshold (1 = very socially appropriate, −1 = very socially inappropriate). The dashed black line indicates the position of the legal threshold in each situation; the dotted line indicates the position of the placebo threshold (values of the legal and placebo thresholds are reported in the bottom-right box). Bars are 95 percent confidence intervals.

Taken together, the results from the placebo experiments and our main results provide a coherent set of evidence that corroborates our preferred interpretation, and provide little support for the notion that subjects exploit focal points alternative to the norm to coordinate responses in the Krupka-Weber task. This result suggests the Krupka-Weber method—an influential tool to elicit norms—is robust to the presence of norm-unrelated focal points (also see Krupka, Leider, and Jiang 2016; Fallucchi and Nosenzo 2022, for further evidence on this point). We also lack much evidence of the effects being driven by a mechanism whereby laws are perceived to transmit information about relevant features of the decision environment. Note that we fail to observe meaningful differences between the vignettes with low versus high informational content; there is one significant discontinuity out of four cases when the placebo thresholds have low informational content, and two discontinuities out of six cases when they have high informational content (one of which, in the speeding vignette, actually goes against the informational content of the placebo).

Taken together, our results suggest the discontinuities observed in our experiments are unlikely to be driven by an information transmission mechanism. Nevertheless, an important caveat when interpreting these findings is that the perceived informational weight of the placebos may not be comparable to that of what has been passed into law (following extensive studies, debates, expert reports, etc.). This is especially true in our experiments, where the groups advocating the placebo thresholds were fictitious and hypothetical. Thus, it may still be possible that, when the perceived informational weight is very high as in the case of the law, part of the effects are due to a perception that the law transmits information about the decision environment.

B. *Experiment 2: Prosocial Traits*

To further probe our interpretation of the main result, we designed another experiment where, instead of measuring the social appropriateness of the behaviors in the vignettes, we asked subjects to report their perception of the prosocial traits of individuals engaging in those behaviors. This may be seen as a more direct test of the mechanism proposed in the model of Section I, where the social esteem $S(\cdot)$ that accrues for engaging in a certain behavior is determined by the inferences observers make about the doer's "type," defined in terms of their prosociality (i.e., how much they care about affecting others' payoffs).

After reading a vignette, subjects in this alternative experiment had to report the likelihood that the person in the vignette would engage in three different types of prosocial behaviors, involving *trustworthiness*, *honesty*, and *altruism*. Trustworthiness was captured by eliciting the perceived likelihood that the person would keep a promise made to a friend. Honesty was measured by eliciting the likelihood that the person would spontaneously return excess change accidentally given by a cashier. Altruism was measured by asking the likelihood that the person would volunteer for a charity. In each case, subjects responded using a four-point ordered scale ("very likely," "somewhat likely," "somewhat unlikely," "very unlikely") that mirrored the scale used in our other experiments.

Subjects also reported the likelihood of three additional behaviors included in the experiment as fillers to distract subjects from the study's true objective. We chose behaviors that are socially desirable (as the target behaviors arguably are), but unrelated to prosociality. The behaviors were exercising regularly to keep fit, keeping a healthy diet, and reading at least two books per month.

To make the task more manageable for subjects and to further reduce the scope for experimenter demand, each subject was presented with only one target vignette plus three filler vignettes.[28] The three filler vignettes were always presented first, in random order, with the target vignette presented last. We used eight different versions of each target vignette, varying whether the described behavior was legal or illegal and its distance from the threshold (four versions legal and four illegal). Each subject was randomly assigned to only one version of each target vignette.

To incentivize responses, we used the opinion matching method. We recruited a total of 2,767 subjects and randomly assigned 783 of them to a nonincentivized condition which simply asked them to report their personal opinion of the likelihood that the described person would engage in each of the six behaviors. The remaining 1,984 subjects were asked to guess the most common response to each question among the first group. The experiments were run in June 2021 with a sample of the UK general population recruited on Prolific. All subjects were paid a £0.94 participation fee, and we randomly selected one-tenth of those in the incentivized condition and paid them £30 if their response to one randomly selected question matched the most common response among the nonincentivized group.

Our test consists of measuring whether, in each of the five target vignettes, we detect a discontinuity at the legal threshold in the inferences subjects make about a person's trustworthiness, honesty and altruism. Figure 3 reports the main results, based on the responses in the incentivized condition.[29] To construct the figure, we assigned evenly-spaced values of $+1$ to the response that a person is "very likely" to engage in a certain prosocial behavior, $+0.33$ to the response "somewhat likely," -0.33 to "somewhat unlikely," and -1 to "very unlikely." Thus, the figure uses the same scale as our previous appropriateness figures. Positive values indicate that, on average, a person is evaluated as likely to engage in prosocial behavior, while negative values indicate the opposite. Each panel shows separate functions for the likelihood the person engages in trustworthy (blue circles), honest (red squares), and altruistic behavior (green triangles).

Several interesting results emerge. First, in all three vignettes where our main experiment found the strongest expressive power of law (age of consent, alcohol to youth, and cash at customs), we observe marked discontinuities in perceived trustworthiness, honesty, and (to a lesser extent) altruism. Across the three vignettes, the size of the discontinuities—estimated by regression analysis, which mirrors the analysis of the main experiment and is reported in full in online Appendix G—ranges between 0.63 and 0.45 for trustworthiness, between 0.81 and 0.51 for honesty, and between 0.42 and 0.26 for altruism. The discontinuities are statistically significant in all cases (all $p \leq 0.009$).[30]

Second, we observe somewhat smaller discontinuities in the two vignettes that found weaker expressive power of law (speeding and drink driving). For speeding, the discontinuity sizes are 0.37 for honesty, 0.24 for trustworthiness, and 0.20 for

[28] The three filler vignettes were the three vignettes subjects saw at the beginning of all our experiments, which we use to emphasize the difference between legality and appropriateness; for details see footnote 17.

[29] As in the placebo experiment, our budget only allowed us to recruit a smaller number of subjects in the nonincentivized condition with the purpose of using their data to incentivize the second group. Online Appendix D also reports the complete distributions of responses to the nonincentivized condition. This shows, as in our other experiments, second-order beliefs were generally well calibrated.

[30] In online Appendix G we also report a robustness (placebo) analysis analogous to that performed for our main experiment. This shows the discontinuities at the legal threshold are larger and more systematic than the placebo discontinuities.

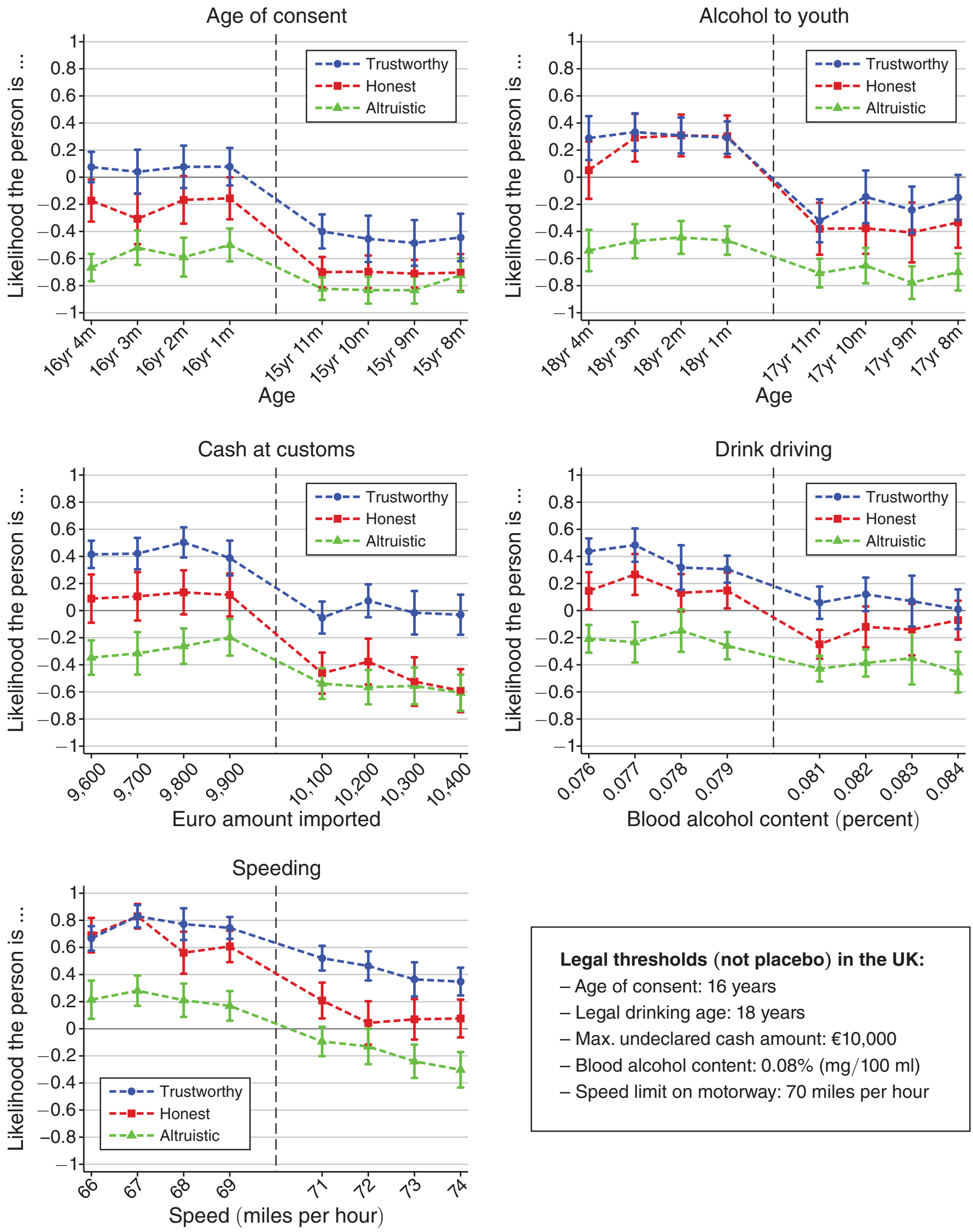

FIGURE 3. LEGAL THRESHOLDS AND PERCEIVED TRUSTWORTHINESS, HONESTY, AND ALTRUISM

Notes: Each panel plots the average perception that the person in the vignette engages in trustworthy, honest, and altruistic behavior based on opportunities at various distance from the legal threshold ($1 = $ very likely to be trustworthy/honest/altruistic, $-1 = $ very unlikely). The dashed black line indicates the position of the legal threshold in each situation (values of the legal thresholds are reported in the bottom-right box). Bars are 95 percent confidence intervals.

altruism; the altruism discontinuity is significant at the 5 percent level ($p = 0.045$), while the other two are significant at the 1 percent level ($p \leq 0.002$). For drink driving, the discontinuities are 0.42 for honesty, 0.16 for altruism, and 0.15 for trustworthiness. The effect is insignificant for trustworthiness ($p = 0.130$), marginally significant for altruism ($p = 0.079$), and significant at the 1 percent level for honesty ($p = 0.000$).

These results mirror our previous findings and corroborate the interpretation proposed earlier: laws that set limits for age of consent, drinking age, and cash imported at customs have strong and discontinuous effects on the inferences observers make about a person's prosocial "type." In contrast, the influence of law is somewhat weaker for speeding and especially drink driving. In either case, the fact that learning about a person's (il)legal behavior allows participants to make (discontinuous) inferences about the person's *future* prosocial behavior in *other* domains speaks in favor of our model of Section I, as it is a strong indication that the effects observed in our experiments are related to a notion of "prosocial type" that is applied consistently and meaningfully across decision situations. In addition, these findings shed further light on the *specific dimensions of prosociality* these inferences revolve around. It appears observers strongly update beliefs about a person's trustworthiness and honesty, and only to a lesser extent about their altruism (which in the experiment was measured in terms of charitable volunteering).

C. *Experiment 3: "Bad" Law*

As an alternative explanation for the expressive power of law, some legal scholars have argued it stems from individuals feeling obliged to obey the law, i.e., a meta-norm of legal obedience (e.g., McAdams and Rasmusen 2007). Thus, (il)legal behavior is judged as (in)appropriate because it complies with (violates) this meta-norm. In online Appendix H we formalize this alternative mechanism in a model where people suffer an individual-specific psychological disutility for breaking rules (including laws). To mirror our model of Section I, we also let utility depend on the inferences others make about the size of the individual's disutility from rule breaking (individuals who suffer more from rule breaking receive higher esteem). We show this model predicts a downward discontinuity in social appropriateness at the legal threshold in the vignettes of the previous sections. Moreover, the model may also explain the observed discontinuities in perceptions of prosociality between those seizing legal and illegal opportunities (Section IVB). Trustworthiness and honesty, in particular, may be viewed as proxies for the psychological cost of rule breaking since breaking promises and lying are themselves violations of well-established normative principles. That we observe weaker discontinuities for altruism can also be consistent with the model: the normative principle guiding altruistic behavior may not be as well defined and "dichotomous" as the rules "I should keep promises" and "I should not lie," so altruism may be a weaker proxy for the cost of rule breaking.

Thus, the data presented in the previous sections may be both interpreted as supportive of the model of Section I and of a meta-norm explanation of the expressive power of law, since the two theories have very similar behavioral implications for the laws we studied so far. However, as online Appendix H shows, the two theories can make distinct predictions for laws that, intentionally or accidentally, prescribe behaviors imposing negative externalities on others (or proscribe behaviors generating positive externalities). This is because an audience may update positively their beliefs about a person's prosociality after observing a violation of a "bad" law (breaking the law *avoids* a negative externality), but a violation always triggers a negative update of the audience's beliefs about the extent the person suffers from rule breaking. In a model where observers value rule following and not prosociality, compliance with bad laws is predicted to attract esteem, which seems doubtful.

Bad laws are uncommon, but not unprecedented. One case that has received recent attention is a set of US laws known as criminal activity nuisance ordinances (CANOs), which set penalties against property owners if repeated incidents of criminal activity occur at their properties in a given time period. The rationale for CANOs is to set incentives against wasting police time and encourage individuals (particularly landlords) to take action to forestall criminal activities at their properties. However, these laws have been heavily criticized because they increase the cost of access to emergency services for victims, especially of domestic violence.[31] Thus, in some circumstances, CANOs can be examples of bad laws, in that sometimes violating the ordinance is the most prosocial (i.e., positive externality generating) action an individual can take.

We designed a new vignette based on a CANO and ran a new experiment, following similar procedures as our previous "prosocial traits" experiments.[32] The vignette described a landlord witnessing an episode of criminal behavior at a property of his (an apartment he rented to a single woman, who is assaulted by her ex-boyfriend). The landlord calls 911 to report the incident. Across four different versions of the vignette, we varied whether the landlord's report triggers a violation of the local CANO, depending on the number of previous reports of criminal activity at the same property in the previous 30 days. In two "illegal" versions of the vignette, by reporting the incident the landlord triggers a violation of the ordinance, while in two "legal" versions the report does not trigger this. Each subject was randomly assigned to only one version of the vignette.

In each case, we asked subjects to indicate the likelihood that the landlord would engage in different types of behavior, using the opinion matching method. We recruited a total of 1,202 subjects and randomly assigned 400 to a nonincentivized condition and the remaining 802 to an incentivized condition where they had to guess the first group's most common response. To some subjects ($N = 599$), as in our experiments of Section IVB, we asked the likelihood that the landlord is (i) trustworthy, (ii) honest, and (iii) altruistic (plus three other filler behaviors). To another group ($N = 603$) we asked the likelihood that the landlord, in general, complies with rules (plus the same three filler behaviors). We added this question about rule compliance to help us separate between the two explanations (see online Appendix H).

The logic of our test is as follows. In a model where individuals care about being seen as rule followers and where trustworthiness, honesty and possibly altruism are used as a proxy for one's preference for rule following, the perceptions of these traits should exhibit a downward discontinuity (if any) as we move from behavior in compliance with the ordinance to behavior in violation of it. However, if

[31] For evidence on the negative consequences of CANOs, see Golestani (2022) and Desmond and Valdez (2012).

[32] We designed the vignette based on the CANO in force at the time of writing in Washington County, Oregon (https://library.municode.com/or/Washington_County/codes/code_of_ordinances?nodeId=TIT8HESA_CH8.44CHNUPR). We ran a pilot to fine-tune the design of the vignette. See our preregistration for more details (Lane, Nosenzo, and Sonderegger 2022).

trustworthiness, honesty, and altruism are *not* used as a proxy for rule following but as a proxy for prosociality, we may instead observe an *upward* discontinuity across legal/illegal versions of the vignette, since by violating the ordinance the individual incurs a (discontinuous) cost to take a prosocial action (calling 911 to stop an assault). Thus, if we observe upward discontinuities in the perceptions of the landlord's trustworthiness, honesty or altruism, these cannot be explained by a model of rule following, but are compatible with the model of Section I. Moreover, even if we do not observe upward discontinuities in trustworthiness, honesty, and altruism, we can compare whether people update differently across the traits of rule compliance and trustworthiness/honesty/altruism, as an additional way to disentangle the two models.

The experiments were run in March 2022 using a sample of the US general population recruited on Prolific. We used the same incentive scheme as our experiments of Section IVB (£0.94 participation fee plus, for the incentivized group, a 1/10 chance of receiving £30 for matching the most common response among the nonincentivized group).

Figure 4 shows the results (for second-order beliefs). Formal analyses are reported in online Appendix I. We observe upward discontinuities in trustworthiness (magnitude: +0.11), honesty (+0.36), and altruism (+0.17), albeit only honesty is statistically significant ($p = 0.002$). We observe a negative discontinuity very close to zero for the rule compliance trait (magnitude: −0.04). The discontinuities in the prosocial traits are small compared to those found in the five vignettes used in the other experiments (especially age of consent, drinking age, and cash imported at customs). This is not completely surprising: although we used a CANO as a bad law, some subjects may not have agreed with this interpretation, since these ordinances can be socially beneficial in some circumstances (e.g., by reducing police costs and deterring crime). Thus, violating the ordinance may be taken as a mixed signal of someone's prosociality. Nevertheless, that we observe upward discontinuities in all prosocial traits—and significantly so for honesty—suggests, on balance, violating the CANO is viewed as a positive signal of prosociality.

Overall, these results cast doubt on the meta-norm explanation of our findings. The upward discontinuity in honesty in the CANO vignette suggests people update beliefs about this trait based on the prosociality of the behavior they observe, and not on whether it conforms to a legal rule. The directionally similar but statistically insignificant results for trustworthiness and altruism provide analogous suggestive evidence for these traits too. Taken together, these patterns suggest beliefs about these traits are not merely proxies for beliefs about rule-following tendencies, but instead relate to inferences about a person's prosociality, as proposed in the model of Section I.

D. *Experiment 4: Weaker Rule of Law*

The final robustness experiment we report was a replication of the UK student experiment, conducted at the same time but with subjects instead drawn from China. The purpose was to probe the extent to which our results would generalize to a very different legislative environment, with weaker rule of law. We relegate the exposition of this experiment to online Appendix J. There we show, despite some differences

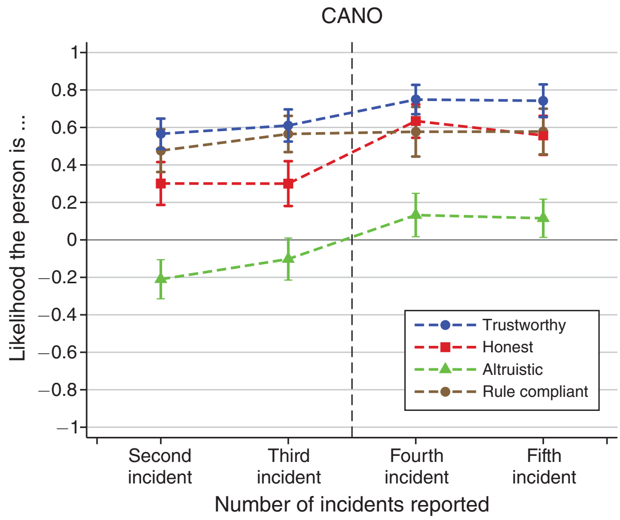

FIGURE 4. PERCEIVED TRUSTWORTHINESS, HONESTY, ALTRUISM, AND RULE COMPLIANCE IN THE CANO VIGNETTE

Notes: The figure plots the average perception that the landlord engages in trustworthy, honest, altruistic, and rule compliant behavior based on opportunities at various distance from the legal threshold (1 = very likely to be trustworthy/honest/altruistic/rule compliant, −1 = very unlikely). The dashed black line indicates the position of the legal threshold (the CANO we used in the experiment specified that a property is declared a "chronic nuisance" if four or more separate incidents are reported within a 30-day period). Actions to the left of the threshold are in compliance with the CANO; actions to the right violate the ordinance. Bars are 95 percent confidence intervals.

in the results between the United Kingdom and China, the general finding that laws often, but not always, exert strong effects on norms remains true among the Chinese sample. This demonstrates that strong rule of law is not necessary for the expressive power of law to take hold.

E. *Alternative Mechanism: Conformity and Descriptive Norms*

In this final subsection we discuss one more alternative mechanism that could explain our data. Our mechanism proposed in Section I relies on individuals caring about the social sanctions and rewards they receive for taking actions. We defined these sanctions/rewards as a function of the inferences observers make about an individual's prosociality. In an alternative model, the sanctions/rewards could be defined as a function of the distance between the individual's behavior and the average/most common behavior in their society (e.g., Kandel and Lazear 1992; Sliwka 2007; Grout, Mitraille, and Sonderegger 2015). The distinction between the two models echoes the difference in the literature between "injunctive" and "descriptive" norms (e.g., Cialdini, Reno, and Kallgren 1990; Bicchieri 2006).[33]

[33] Note however that our model of Section I encompasses both an "injunctive" and a "descriptive" mechanism. The injunctive component lies in that we are assuming an injunctive norm of prosociality (i.e., prosociality is a desirable trait). The descriptive component operates through the esteem for taking a certain action being related to the share of people taking it, since this statistic contains information about the highest "type" willing to take the action and hence about the distribution of types above and below this cutoff.

In online Appendix K, we sketch a formal model capturing this intuition. We have two versions of the model. In the first, individuals derive utility from material gain and conformity with the most common behavior in their society. Under the assumption that judgments of social appropriateness are informed by common or normal behavior in society, we show such a model also predicts discontinuities at the legal threshold in the norm functions of our main experiment. Intuitively, the discontinuous material incentives between legal and illegal behavior induce a discontinuity in the share of individuals seizing legal/illegal opportunities, which—if the extent to which an action is judged "socially appropriate" reflects the share of people taking it—translates in an analogous discontinuity in norm judgments. However, such a simple model does not seem well equipped to explain the results of the experiment reported in Section IVB. There we observe discontinuities in the perceptions of trustworthiness, honesty, and altruism of subjects seizing legal or illegal opportunities. The model cannot explain this without further assuming a positive correlation between one's degree of conformity-seeking and prosociality.

In a second version, we therefore augment the basic model by assuming that, in addition to monetary gains and conformity, individuals also care (heterogeneously) about prosociality (i.e., they suffer a disutility when imposing a negative externality, and the extent of this disutility varies across individuals). This second model mirrors the theory in Section I, except that we replace the payoff for social esteem/stigma with one for conforming with the most common behavior. Under the assumption that social appropriateness is defined in terms of common behavior, this version of the model can also explain the results of Section III, for the same reason as before. Additionally, the model predicts a discontinuity in the degree of prosociality between individuals seizing legal and illegal opportunities, which can explain the results of Section IVB.

Although this second version of the conformity model can explain our experimental data, it presents disadvantages compared to our preferred explanation of Section I. First, under some circumstances, it generates counterintuitive predictions. In online Appendix K we show it can support equilibria where a "non-prosocial" action (i.e., one generating a negative externality) is taken by the least prosocial individuals and yet judged socially appropriate, while a "prosocial" action (generating no negative externality) is taken by the most prosocial individuals but judged inappropriate. These counterintuitive results are due to the conformity model featuring a disconnect between prosociality and social appropriateness, since the latter is defined in terms of what most people do, and not of the underlying type of those choosing a certain action.

Second, the conformity model seems less parsimonious than our model of social esteem. While the model of esteem relies on interconnected behavioral motives (prosociality and the desire to be seen as prosocial), the conformity model needs two disconnected behavioral motivations to explain the data (prosociality and conformity). Moreover, to explain our data, the model requires the assumption that the behaviors people find appropriate are those taken more frequently, which our preferred model does not need.

Finally, while it may be plausible that perceptions of social appropriateness ("injunctive" norms) are partly influenced by the prevalence of behavior ("descriptive" norms), in the model of conformity the distinction between these two types of norm is blurred. This is not completely in line with many theoretical accounts of social norms that often make a clear distinction between injunctive and descriptive norms and allow the two to conflict with one another (e.g., Cialdini, Reno, and Kallgren 1990; Bicchieri 2006). In fact, evidence from a variety of contexts suggests beliefs about what is socially appropriate are disconnected from what people expect others to do (e.g., Klimm 2019; d'Adda et al. 2020; Krysowski and Tremewan 2021; Kölle and Quercia 2021; Bicchieri et al. 2022).[34] Moreover, there is also evidence that judgments of appropriateness are unresponsive to information about others' behavior (Krupka, Leider, and Jiang 2016). Taken together, the empirical evidence casts doubt on the key assumption of the conformity model about the close association between descriptive and injunctive norms.

V. Conclusions

For some years scholars across the social sciences have asserted legal rules carry expressive power, i.e., the ability to shape the social norms within a society. However, because societal laws and norms typically coevolve, it has been difficult to design empirical strategies to establish a clear causal effect from laws to norms, which explains the paucity of empirical work on this topic. This paper has employed a novel empirical strategy to identify the causal influence of laws on norms. Our design takes advantage both of recent advances in methods to estimate norms, and vignettes with laws characterized by thresholds. Our empirical strategy rests on the crucial assumption that norms do not differentiate between arbitrarily close actions, and therefore that actions that are close but on opposite sides of a legal threshold would obtain similar normative status in absence of the law. Although this assumption may not hold in general (some marginally different actions may produce substantially different effects and therefore be normatively evaluated very differently), we argue it is likely to apply to the situations studied in our vignettes. Our analysis allows us to conclude an action's legal status does causally influence its normative appropriateness.

Our results have important implications for the effectiveness of laws and formal institutions more generally. They imply the impact of formal rules on behavior is greater than their mere deterrent effect, the standard mechanism through which economists have traditionally argued laws and institutions take effect (e.g., Becker 1968). Instead, our findings show laws can also affect behavior by strengthening the social disapproval towards illegal actions. The effects on behavior of this expressive power may be substantial: in some vignettes we find the effect of law on norms is not just statistically significant, but of a quantitatively large magnitude.

As discussed above, different mechanisms have been proposed for why laws would have expressive power on norms. Theories that laws transmit information about "community standards" (e.g., Bénabou and Tirole 2011), though empirically

[34] Klimm (2019) finds people expect most others to cheat and yet view cheating as socially inappropriate. D'Adda et al. (2020) and Krysowski and Tremewan (2021) study variants of the dictator game where people expect a large fraction of others to be selfish and yet view this as inappropriate behavior. Kölle and Quercia (2021) study public goods games where subjects view full contribution as the most appropriate action and yet expect most others to contribute only half of the endowment. Bicchieri et al. (2022) study a donation game where the most appropriate action is to give to charity but where most subjects predict most others do not give.

possible, cannot be tested in our design because our empirical strategy rules out by assumption the possibility that legal thresholds carry informational content (an assumption corroborated in our placebo experiments). Alternative theories, such as the meta-norm explanations that obeying the law is itself a norm, do not seem able to explain all our data. We argue that a mechanism to explain our large set of findings can be provided by the signaling theory sketched in the paper, which uses a "social image" framework (Bénabou and Tirole 2006, 2011) to formalize the arguments proposed by Posner (1998, 2000, 2002). In this model, illegality can make behavior less appropriate because of the signal it sends about the person committing it. Our experiments corroborate this intuition by showing people update their beliefs about someone's prosociality upon observing their legal compliance.

More broadly, our findings speak to the theoretical and empirical literature on the relation between formal and informal incentives regulating behavior in social settings (Bénabou and Tirole 2003, 2006, 2011; Bowles and Polanía-Reyes 2012). This literature has often highlighted how formal incentives can crowd out social incentives to engage in prosocial behavior, suggesting the two types of incentives can act as *substitutes*. Our paper provides evidence of a mechanism that produces instead a *complementarity* between formal and social incentives: by shaping social norms, laws can harness the power of social incentives to reinforce the deterrent effect of formal incentives. An interesting question for further research would be to explore the interplay between these substitution and complementarity effects and investigate whether they might systematically relate to the nature of the institutions setting the formal incentives, e.g., whether they are governments (as in our experiment) or private organizations (as in much of the crowding-out literature).

Our results also raise interesting new questions—both for theory and empirical work—about the scope for laws and formal institutions to initiate societal change. Take, for instance, gender gaps in social and economic outcomes, which are believed to be partly driven by gender norms perpetuating socioeconomic inequality between men and women (e.g., Akerlof and Kranton 2000; Bertrand, Kamenica, and Pan 2015). What is the scope for law to influence and shape norms so as to correct these gender inequalities? The answer to this question is far from obvious, as it depends whether laws have the *same* expressive power across heterogeneous subgroups of the population, such as gender or racial groups. The mechanism sketched in our model suggests this need not be the case: the effect of law on norms may systematically differ across men and women, or across races. For instance, if there are systematic differences in law enforcement between different groups (e.g., more prevalent wrongful convictions among certain groups), a person's criminal record will be a more or less noisy signal of his/her type for individuals belonging to different subgroups of the population, leading to smaller or larger discontinuities in the norm function. In follow-up work, we are exploring these conjectures theoretically and empirically (Görges et al. 2023).

REFERENCES

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond. 2019. "Preferences for Truth-Telling." *Econometrica* 87 (4): 1115–53.

Acemoglu, Daron, and Matthew O. Jackson. 2017. "Social Norms and the Enforcement of Laws." *Journal of the European Economic Association* 15 (2): 245–95.

Adriani, Fabrizio, and Silvia Sonderegger. 2019. "A Theory of Esteem Based Peer Pressure." *Games and Economic Behavior* 115: 314–35.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–53.

Aksoy, Cevat G., Christopher S. Carpenter, Ralph De Haas, and Kevin D. Tran. 2020. "Do Laws Shape Attitudes? Evidence from Same-Sex Relationship Recognition Policies in Europe." *European Economic Review* 124: 103399.

Ali, S. Nageeb, and Roland Bénabou. 2020. "Image versus Information: Changing Societal Norms and Optimal Privacy." *American Economic Journal: Microeconomics* 12 (3): 116–64.

Andreoni, James, and Douglas Bernheim. 2009. "Social Image and the Fifty–Fifty Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica* 77 (5): 1607–36.

Aycinena, Diego, Francesco Bogliacino, and Eric Kimbrough. 2022. "Measuring Norms: Assessing the Threat of Social Desirability Bias to the Bicchieri and Xiao Elicitation Method." SSRN 4029493.

Barr, Abigail, Tom Lane, and Daniele Nosenzo. 2018. "On the Social Inappropriateness of Discrimination." *Journal of Public Economics* 164: 153–64.

Barron, Kai, and Tuomas Nurminen. 2020. "Nudging Cooperation in Public Goods Provision." *Journal of Behavioral and Experimental Economics* 88: 101542.

Becker, Gary. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (2): 169–217.

Bénabou, Roland, and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies* 70 (3): 489–520.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78.

Bénabou, Roland, and Jean Tirole. 2011. "Laws and Norms." NBER Working Paper 17579.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B* (*Methodological*) 57 (1): 289–300.

Bernheim, Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy* 102 (5): 841–77.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households." *Quarterly Journal of Economics* 130 (2): 571–614.

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms.* Cambridge, UK: Cambridge University Press.

Bicchieri, Cristina, Eugen Dimant, Simon Gächter, and Daniele Nosenzo. 2022. "Social Proximity and the Erosion of Norm Compliance." *Games and Economic Behavior* 132: 59–72.

Bicchieri, Cristina, and Erte Xiao. 2009. "Do the Right Thing: But Only If Others Do So." *Journal of Behavioral Decision Making* 22 (2): 191–208.

Birke, David J. 2020. "Anti-bunching: A New Test for Signaling Motives in Prosocial Behavior." Unpublished.

Bowles, Samuel, and Sandra Polanía-Reyes. 2012. "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature* 50 (2): 368–425.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. 2020. "From Extreme to Mainstream: The Erosion of Social Norms." *American Economic Review* 110 (11): 3522–48.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott. 2020. "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia." *American Economic Review* 110 (10): 2997–3029.

Casoria, Fortuna, Fabio Galeotti, and Marie Claire Villeval. 2020. "Perceived Social Norm and Behavior Quickly Adjusted to Legal Changes during the COVID-19 Pandemic in France." SSRN 3670895.

Charness, Gary, Michael Cooper, and J. Lucas Reddinger. 2020. "Wage Policies, Incentive Schemes, and Motivation." In *Handbook of Labor, Human Resources and Population Economics*, edited by Klaus F. Zimmermann. New York: Springer, Cham. https://doi.org/10.1007/978-3-319-57365-6_125-1.

Chen, Daniel L., and Susan Yeh. 2013. "Distinguishing between Custom and Law: Empirical Examples of Endogeneity in Property and First Amendment Precedents." *William & Mary Bill of Rights Journal* 21 (4): 1081.

Chen, Daniel L., and Susan Yeh. 2014. "The Construction of Morals." *Journal of Economic Behavior and Organization* 104: 84–105.

Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren. 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology* 58 (6): 1015–26.

d'Adda, Giovanna, Martin Dufwenberg, Francesco Passarelli, and Guido Tabellini. 2020. "Social Norms with Private Values: Theory and Experiments." *Games and Economic Behavior* 124: 288–304.

Desmond, Matthew, and Nicol Valdez. 2012. "Unpolicing the Urban Poor: Consequences of Third-Party Policing for Inner-City Women." *American Sociological Review* 78 (1): 117–41.

Dwenger, Nadja, Henrik Kleven, Imran Rasul, and Johannes Rincke. 2016. "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany." *American Economic Journal: Economic Policy* 8 (3): 203–32.

Ellingsen, Tore, and Magnus Johannesson. 2008. "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review* 98 (3): 990–1008.

Engl, Florian, Arno Riedl, and Roberto A. Weber. 2021. "Spillover Effects of Institutions on Cooperative Behavior, Preferences, and Beliefs." *American Economic Journal: Microeconomics* 13 (4): 261–99.

Falk, Armin, Ernst Fehr, and Christian Zehnder. 2006. "Fairness Perceptions and Reservation Wages—the Behavioral Effects of Minimum Wage Laws." *Quarterly Journal of Economics* 121 (4): 1347–81.

Fallucchi, Francesco, and Daniele Nosenzo. 2022. "The Coordinating Power of Social Norms." *Experimental Economics* 25 (1): 1–25.

Fehr, Ernst, and Bettina Rockenbach. 2003. "Detrimental Effects of Sanctions on Human Altruism." *Nature* 422 (6928): 137–40.

Fisman, Raymond, and Edward Miguel. 2007. "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets." *Journal of Political Economy* 115 (6): 1020–48.

Funk, Patricia. 2007. "Is There an Expressive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines." *American Law and Economics Review* 9 (1): 135–59.

Galbiati, Roberto, Emeric Henry, Nicolas Jacquemet, and Max Lobeck. 2020. "How Laws Affect the Perception of Norms: Empirical Evidence from the Lockdown." SSRN 3684710.

Galbiati, Roberto, and Pietro Vertova. 2008. "Obligations and Cooperative Behaviour in Public Good Games." *Games and Economic Behavior* 64 (1): 146–70.

Galbiati, Roberto, and Pietro Vertova. 2014. "How Laws Affect Behavior: Obligations, Incentives and Cooperative Behavior." *International Review of Law and Economics* 38: 48–57.

Gneezy, Uri, and Aldo Rustichini. 2000. "A Fine Is a Price." *Journal of Legal Studies* 29 (1): 1–17.

Golestani, Aria. 2022. "Silenced: Consequences of the Nuisance Property Ordinances." Unpublished.

Görges, Luise, Tom Lane, Daniele Nosenzo, and Silvia Sonderegger. 2023. "Equal before the (Expressive Power of) Law?" Unpublished.

Görges, Luise, and Daniele Nosenzo. 2020. "Measuring Social Norms in Economics: Why It Is Important and How It Is Done." *Analyse and Kritik* 42 (2): 285–311.

Govindan, Pavitra. 2021. "Effect of Moderate and Radical Rules on Behavior and Caste Norms: Lab-in-the-Field Experiment in India." Unpublished.

Grout, Paul A., Sebastien Mitraille, and Silvia Sonderegger. 2015. "The Costs and Benefits of Coordinating with a Different Group." *Journal of Economic Theory* 160: 517–35.

Kandel, Eugene, and Edward P. Lazear. 1992. "Peer Pressure and Partnerships." *Journal of Political Economy* 100 (4): 801–17.

Klimm, Felix. 2019. "Suspicious Success Cheating, Inequality Acceptance, and Political Preferences." *European Economic Review* 117: 36–55.

Kölle, Felix, and Simone Quercia. 2021. "The Influence of Empirical and Normative Expectations on Cooperation." *Journal of Economic Behavior Organization* 190: 691–703.

Krupka, Erin L., Stephen Leider, and Ming Jiang. 2016. "A Meeting of the Minds: Informal Agreements and Social Norms." *Management Science* 63 (6): 1708–29.

Krupka, Erin L., and Roberto A. Weber. 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association* 11 (3): 495–524.

Krysowski, Eryk, and James Tremewan. 2021. "Why Does Anonymity Make Us Misbehave: Different Norms or Less Compliance?" *Economic Inquiry* 59 (2): 776–89.

Lane, Tom, and Daniele Nosenzo. 2019. "Law and Norms: Empirical Evidence." Centre for Research and Experimental Economics Discussion Paper 2019-08.

Lane, Tom, Daniele Nosenzo, and Silvia Sonderegger. 2021. "Pre-registration #66794 for Law and Norms: Empirical Evidence." *Aspredicted.org*. https://aspredicted.org/blind.php?x=UCQ_QBH (accessed May 25, 2021).

Lane, Tom, Daniele Nosenzo, and Silvia Sonderegger. 2022. "Pre-registration #90860 for Law and Norms: Empirical Evidence." *Aspredicted.org*. https://aspredicted.org/blind.php?x=LBY_75F (accessed March 14, 2022).

Lane, Tom, Daniele Nosenzo, and Silvia Sonderegger. 2023. "Replication Data for: Law and Norms: Empirical Evidence." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.388/E182995V1.

McAdams, Richard. 2000. "An Attitudinal Theory of Expressive Law." *Oregon Law Review* 79: 339–90.

McAdams, Richard. 2015. *The Expressive Powers of Law.* Cambridge, MA: Harvard University Press.

McAdams, Richard, and Eric Rasmusen. 2007. "Norms and the Law." In *Handbook of Law and Economics*, edited by A. Mitchell Polinsky and Steven Shavell, 1573–1618. Amsterdam: Elsevier.

Posner, Richard A. 1997. "Social Norms and the Law: An Economic Approach." *American Economic Review* 87 (2): 365–69.

Posner, Eric A. 1998. "Symbols, Signals, and Social Norms in Politics and the Law." *Journal of Legal Studies* 27 (2): 765–97.

Posner, Eric A. 2000. "Law and Social Norms: The Case of Tax Compliance." *Virginia Law Review* 86 (8): 1781–1819.

Posner, Eric A. 2002. *Law and Social Norms.* Cambridge, MA: Harvard University Press.

Rees-Jones, Alex, and Kyle Rozema. 2019. "Price Isn't Everything: Behavioral Response around Changes in Sin Taxes." NBER Working Paper 25958.

Schelling, Thomas. 1960. *The Strategy of Conflict.* Cambridge, MA: Harvard University Press.

Simes, Robert J. 1986. "An Improved Bonferroni Procedure for Multiple Tests of Significance." *Biometrika* 73 (3): 751–54.

Sliwka, Dirk. 2007. "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *American Economic Review* 97 (3): 999–1012.

Sunstein, Cass R. 1996. "On the Expressive Function of Law." *University of Pennsylvania Law Review* 144 (5): 2021–53.

Tankard, Margareth, and Elizabeth Paluck. 2017. "The Effect of a Supreme Court Decision Regarding Gay Marriage on Social Norms and Personal Attitudes." *Psychological Science* 28 (9): 1334–44.

Tirole, Jean. 2021. "Digital Dystopia." *American Economic Review* 111 (6): 2007–48.

van der Weele, Joel. 2012. "The Signaling Power of Sanctions in Social Dilemmas." *Journal of Law, Economics, and Organization* 28 (1): 103–26.

Wittlin, Maggie. 2011. "Buckling under Pressure: An Empirical Test of the Expressive Effects of Law." *Yale Journal on Regulation* 28 (2): 419–69.